

Universidad de Costa Rica

Escuela de Matemática



Serie: Cabécar

Teoría de Estadística

Matemática



Editorial CIMPA



Prof. Jorge Poltronieri

2006

..**

P772e Poltronieri Vargas, Jorge, 1952–

Teoría de Estadística Matemática / Jorge Poltronieri.–

[San José, C.R.: CIMPA], 2006.

231p. : il. ; 27cm (Serie cabécar)

A la cabeza de la port: Universidad de Costa Rica,

Escuela de Matemática.

ISBN ****_****_*.*

1. ESTADÍSTICA MATEMÁTICA, I. Título.

CIP/1423

CC/SIBDI.UCR

Teoría de Estadística Matemática

© Jorge Poltronieri Vargas, Catedrático

Escuela de Matemática, Universidad de Costa Rica.

Diagramación en \LaTeX realizada por el autor

Diseño y concepto de carátula realizada por el autor

Impreso en Costa Rica

San José, 2006

Contenidos

1	La ley normal	1
1.1	La distribución multinormal	1
1.2	Distribuciones condicionadas	3
1.3	Independencia entre vectores gaussianos (normales)	3
1.4	Distribuciones de formas cuadráticas	4
1.4.1	Teorema general sobre formas cuadráticas	8
1.5	Independencia de formas cuadráticas	9
1.6	Independencia de formas lineales y cuadráticas	11
1.7	Ejercicios	11
2	Inferencia estadística	23
2.1	Introducción a la estadística inferencial	23
2.2	Formalismo matemático	24
2.2.1	Definición de estructura estadística	24
2.3	Estructura dominada	25
2.4	Estadísticas–Estadísticas Integrables	26
2.4.1	Estructura inducida	27
2.5	Estructuras completas	28
2.6	Producto de estructuras	29
2.6.1	Muestra empírica o muestra aleatoria	30

2.7	Problemas de decisión	31
2.7.1	Pruebas de hipótesis	31
2.7.2	Estimación	33
2.8	Noción de intervalo de confianza	35
2.9	Noción de suficiencia estadística	36
2.9.1	Teorema de factorización	37
2.9.2	Aplicaciones al reconocimiento de estadísticas suficientes	42
2.10	Proyección de una estadística. Teorema de Rao–Blackwell	44
2.10.1	Aplicación a la estimación	45
2.11	Método de máxima verosimilitud: una técnica de estimación	46
2.12	Noción de libertad	47
2.13	Ejercicios	48
3	Pruebas de hipótesis	57
3.1	Generalidades	57
3.2	Relación de preorden sobre las pruebas de hipótesis	59
3.3	Pruebas óptimas	60
3.4	Determinación de pruebas óptimas	61
3.4.1	Caso de dos hipótesis simples	61
3.4.2	\mathfrak{P}_0 simple, \mathfrak{P}_1 compuesta	66
3.5	Métodos en ausencia de pruebas óptimas	68
3.5.1	Restricción de la clase de pruebas consideradas	68
3.5.2	Prueba del cociente de verosimilitud	68
3.6	Indicaciones sobre la búsqueda de una prueba U.M.P.I.	75
3.6.1	Indicaciones en el caso \mathfrak{P}_0 compuesta	77
3.6.2	Pruebas libres deducidas de pruebas condicionadas (Estructuras de Neyman)	77
3.7	Indicaciones sobre los problemas de estimación puntual	78
3.7.1	Desigualdad de Cramer–Rao	78

3.8	Estudio de muestras empíricas de leyes gaussianas	82
3.8.1	Estudio de la estadística (\bar{X}, S^2)	82
3.8.2	Aplicación al intervalo de confianza	83
3.8.3	Pruebas usuales para muestras normales	83
3.9	La prueba χ^2 (chi-cuadrado)	87
3.9.1	La prueba χ^2 como una prueba de ajuste	89
3.9.2	Prueba de homogeneidad	92
3.9.3	Comparación de k proporciones observadas	93
3.9.4	Caso general	94
3.9.5	Prueba de independencia de caracteres	94
3.9.6	Método donde se minimiza aproximadamente el χ^2 de una muestra	96
3.9.7	Utilización de la estadística χ^2	98
3.10	Ejercicios	99
4	Estimación paramétrica	113
4.1	Estimación paramétrica en \mathbb{R}^r	113
4.1.1	Caso $\theta \in \mathbb{R}$	114
4.1.2	Caso $\theta \in \mathbb{R}^r$	115
4.2	Estimación puntual unidimensional	117
4.2.1	Caso de una muestra aleatoria	117
4.2.2	Valor mínimo de la varianza de un estimador sesgado	118
4.2.3	Estimación puntual para grandes muestras	118
4.2.4	Intervalos de confianza para muestras de distribución discreta	126
4.2.5	Estimación por intervalo para grandes muestras	127
4.2.6	Estimación por intervalos para funciones de estimación generales	127
4.2.7	Intervalo de confianza asintótico más pequeño	128
4.3	Estimación puntual multidimensional	129
4.3.1	Estimación multidimensional para grandes muestras	132

4.3.2	Eficiencia asintótica de los estimadores de máxima verosimilitud	134
4.3.3	Regiones de confianza multidimensionales	134
4.3.4	Regiones de confianza más pequeñas, para grandes muestras	136
4.4	Ejercicios	140
5	Complementos de la teoría de pruebas de hipótesis	145
5.1	Distribución asintótica del cociente de verosimilitud para grandes muestras	145
5.1.1	Consistencia de la prueba del cociente de verosimilitud	148
5.1.2	Potencia asintótica de la prueba del cociente de verosimilitud	148
5.2	Prueba del cociente de verosimilitud de una hipótesis simple	152
5.3	Prueba del cociente de verosimilitud de una hipótesis compuesta	154
5.4	Ejercicios	157
A	Resumen de Teoría de la media e integración	161
A.1	Propiedades y definiciones	161
A.2	Propiedades de funciones medibles numéricas	163
A.3	Medidas positivas	164
A.4	Integración	165
A.5	Medidas definidas por densidades	168
A.6	Medida imagen	168
A.7	Modos de convergencia. Definiciones	170
A.8	Relaciones entre los tipos de convergencia	171
A.9	Ley débil de grandes números	171
A.10	Convergencia en ley. Criterios	172
A.11	Estudio de la convergencia casi segura (c.s.)	173
A.12	Propiedades de la convergencia en ley y en probabilidad	176
B	Anexos	179
B.1	Notas sobre las funciones eulerianas	180

B.1.1	La función Γ	180
B.1.2	La función β	182
B.2	Sobre las funciones de densidad	184
B.3	Matrices	185
B.3.1	Teoremas	185
B.4	Formas cuadráticas	185
B.5	Determinantes	186
B.6	Derivadas de matrices y vectores	187
B.7	Matrices idempotentes	187
B.8	Máximos, mínimos, Jacobianos	188
B.9	Combinación lineales de variables aleatorias	191
B.10	Características de las variables aleatorias conocidas	192
B.11	Teorema de Rao-Cramer	194
B.12	Convergencia de un proceso aleatorio	195
B.13	Tablas estadísticas	198

Capítulo 1

La ley normal

1.1 La distribución multinormal

Recordemos que un vector aleatorio X es multinormal o gaussiana¹, si su función de densidad es de la forma $f(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})}$, con matriz de covarianzas Σ y media $\boldsymbol{\mu} \in \mathbb{R}^n$, y lo denotamos $X \sim N_n(\boldsymbol{\mu}, \Sigma)$.

La función característica es $\Phi_X(t) = e^{it'\boldsymbol{\mu} - \frac{1}{2}t'\Sigma t}$, $t \in \mathbb{R}^n$.

- Si $X \sim N(\boldsymbol{\mu}, \Sigma)$, $A_{k \times n}$ y $Y = AX$, entonces $Y \sim N_k(A\boldsymbol{\mu}, A\Sigma A')$.

- Sea $X \sim N(\boldsymbol{\mu}, \Sigma)$, con $\Sigma = \begin{pmatrix} \sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_n^2 \end{pmatrix}$, entonces las componentes X_1, \dots, X_n de X son independientes. Si $\sigma_i \neq 0$, las componentes $X_i \sim N(\mu_i, \sigma_i^2)$ y si $\sigma_i = 0$, $X_i = \mu_i$ c.p.d.

¹**Carl Friedrich Gauss (1777-1855)** Nace de 30 de abril 1777 en Brunswick, (hoy Alemania). Muere el 23 de febrero 1855 en Göttingen, Hanover (hoy Alemania). Hijo de un humilde albañil, fue un niño prodigio en matemáticas y continuó siéndolo toda su vida. Su inteligencia llamó la atención del duque de Brunswick, quien decidió costearle todos sus estudios, entrando en 1795 en la Universidad de Göttingen. El día 30 de marzo de 1796 hizo un brillante descubrimiento. Desde hacía más de 2000 años, se sabía como construir con regla y compás el triángulo equilátero, el cuadrado y el pentágono regular (así como algunos otros polígonos regulares cuyos números de lados son múltiplos de dos, de tres o de cinco), pero ningún otro polígono regular con un número primo de lados. Demostró que sólo ciertos polígonos equiláteros se podían construir con ayuda de regla y compás. Hizo una labor importante en la Teoría de Números, sintetizada en su obra "Disquisitiones arithmeticae". También construyó una geometría no euclídea, basada en axiomas distintos a los de Euclides, pero se negó a publicarla. Lobachevski y Bolyai ostentan el honor de su descubrimiento al publicarla algo más tarde. En 1799 Gauss demostró el teorema fundamental del álgebra. El 1801 demostró el teorema fundamental de la aritmética: todo número natural se puede representar como el producto de números primos de una y solamente una forma. Fuera del dominio de las matemáticas puras, Gauss ganó gran fama por su labor sobre el planetoido Ceres, del que calculó su órbita, siendo nombrado director del observatorio de Göttingen en 1807.

Lema 1.1.1 Sean M, Σ dos matrices simétricas, M de orden n y sea Σ definida no negativa, $\text{rang } \Sigma = r$, entonces existe una matriz $A_{n \times r}$ tal que $\Sigma = AA'$, $A'MA = K$, donde K es una matriz diagonal cuyos elementos no nulos son los mismos que los valores propios no nulos de $M\Sigma$.

Prueba Sea C una matriz ortogonal tal que $C\Sigma C'$ sea diagonal, es decir:

$$C\Sigma C' = \begin{pmatrix} \lambda_1 & \cdots & 0 & \\ \vdots & \ddots & \vdots & 0 \\ 0 & \cdots & \lambda_r & \\ & & 0 & 0 \end{pmatrix},$$

donde los λ_i son los valores propios no nulos de Σ .

Sea K_0 la matriz diagonal de valores $\sqrt{\lambda_i}$, $i = 1, \dots, k$ y sea C_0 la matriz formada por las r primeras líneas de C . Si C_1 es una matriz ortogonal de orden r , la matriz $A = C'_0 K_0 C_1$ es tal que $\Sigma = AA'$. En efecto, $AA' = C'_0 K_0^2 C_0$ y $\Sigma = C' \begin{pmatrix} K_0^2 & 0 \\ 0 & 0 \end{pmatrix} C = C'_0 K_0^2 C_0$.

Consideremos la matriz $K_0 C_0 M C'_0 K_0$, es simétrica y existe una matriz ortogonal P de orden r tal que la matriz $K = P' K_0 C_0 M C'_0 K_0 P$ es diagonal.

Si $C_1 = P$ la matriz $A = C'_0 K_0 P$ satisface las condiciones, pues $AA' = C'_0 K_0 P P' K_0 C_0 = C'_0 K_0^2 C_0 = \Sigma$ y $A'MA = P' K_0 C_0 M C'_0 K_0 P = K$.

Queda por demostrar que $K_0 C_0 M C'_0 K_0$ y $M\Sigma$ tienen los mismos valores propios no nulos. Esto es

equivalente a demostrar que si B es simétrica, las dos matrices $B \begin{pmatrix} K_0^2 & 0 \\ 0 & 0 \end{pmatrix}$ y $(K_0, 0) B \begin{pmatrix} K_0 \\ 0 \end{pmatrix}$ tienen los mismos valores propios no nulos y aplicarlo a la matriz $B = C M C'$. Observemos que

$$\det(I_r - \lambda K) = \det(I_n - \lambda \Sigma M) = \det(I_n - \lambda M \Sigma).$$

Sea $X \sim N(\mu, \Sigma)$ y sea $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$, con $X'_1 = (X_1, \dots, X_p)$, entonces $X_1 \sim N_p(\beta, \Sigma_{11})$, donde

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \text{ y } \beta' = (\mu_1, \dots, \mu_p).$$

La demostración es inmediata aplicando la función característica.

Es importante hacer notar que $f(\mathbf{x}_1) = \frac{1}{(2\pi)^{\frac{1}{2}p} |R_{11 \cdot 2}|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{x}_1 - \beta)' R_{11 \cdot 2}^{-1} (\mathbf{x}_1 - \beta)}$, donde $R_{11 \cdot 2} = R_{11} - R_{12} R_{22}^{-1} R_{21}$

$$\text{y } R = \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix} = \Sigma^{-1}.$$

Sea $X \sim N(\boldsymbol{\mu}, \Sigma)$ su función de densidad es de la forma $K e^{-\frac{1}{2}Q(\mathbf{x})}$, donde $Q(\mathbf{x})$ es la forma cuadrática asociada a Σ^{-1} ; entonces el vector $\boldsymbol{\mu}$ de medias es solución del sistema $\frac{\partial Q}{\partial \mathbf{x}} = 0$.

1.2 Distribuciones condicionadas

Teorema 1.2.1 Sea $X \sim N(\boldsymbol{\mu}, \Sigma)$, $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$, $\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}$, $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$, entonces la densidad condicional $X_1/X_2 = \mathbf{x}_2 \sim N(\boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$.

Demostración En efecto:

$$f(\mathbf{x}_1/\mathbf{x}_2) = \frac{(2\pi)^{-\frac{1}{2}n}|\Sigma|^{-\frac{1}{2}}e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})}}{(2\pi)^{-\frac{1}{2}m}|\Sigma_{22}|^{-\frac{1}{2}}e^{-\frac{1}{2}(\mathbf{x}_2-\boldsymbol{\mu}_2)'\Sigma_{22}^{-1}(\mathbf{x}_2-\boldsymbol{\mu}_2)}} = \frac{e^{-\frac{1}{2}(\mathbf{x}_1-\boldsymbol{\beta}_1)'\Sigma_{11}^{-1}(\mathbf{x}_1-\boldsymbol{\beta}_1)}}{(2\pi)^{\frac{1}{2}(n-m)}|\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}|^{\frac{1}{2}}},$$

con $\boldsymbol{\beta}_1 = \boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}(\mathbf{x}_2 - \boldsymbol{\mu}_2)$.

Corolario 1.2.1 $E(X_1/X_2 = \mathbf{x}_2) = \boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}(\mathbf{x}_2 - \boldsymbol{\mu}_2)$ y $\text{var}(X_1/X_2 = \mathbf{x}_2) = \Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ no depende de \mathbf{x}_2 .

1.3 Independencia entre vectores gaussianos (normales)

Sean $X \sim N_p(\boldsymbol{\mu}_1, \Sigma_{11})$ y $Y \sim N_q(\boldsymbol{\mu}_2, \Sigma_{22})$ y sea $Z = \begin{pmatrix} X \\ Y \end{pmatrix}$ vector de \mathbb{R}^{p+q} . Sea Σ matriz de covarianza de Z , Σ_{12} matriz de covarianza entre X y Y , es decir:

$$\Sigma_{12} = E((X - \boldsymbol{\mu}_1)(Y - \boldsymbol{\mu}_2)') = E(XY') - \boldsymbol{\mu}_1\boldsymbol{\mu}_2'.$$

Así tenemos que:

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{pmatrix}.$$

Teorema 1.3.1 Si X, Y son independientes y gaussianos, entonces Z es gaussiano y $\Sigma_{12} = 0$. Recíprocamente si Z es gaussiano y si $\Sigma_{12} = 0$, X y Y son independientes y gaussianos.

Demostración Sea $t = \begin{pmatrix} u \\ v \end{pmatrix} \in \mathbb{R}^{p+q}$, $t'Z = u'X + v'Y$, entonces

$$\Phi_Z(t) = \Phi_X(u)\Phi_Y(v) = e^{iu'\boldsymbol{\mu}_1 - \frac{1}{2}t'\Sigma t} = e^{iu'\boldsymbol{\mu}_1 + iv'\boldsymbol{\mu}_2 - \frac{1}{2}u'\Sigma_{11}u - \frac{1}{2}v'\Sigma_{22}v - u'\Sigma_{12}v},$$

es decir $u' \Sigma_{12} v = 0 \iff \Sigma_{12} = 0$.

Corolario 1.3.1 Sea X_1, \dots, X_n variables aleatorias no correlacionadas, es decir $\Sigma_{ij} = 0$, $i \neq j$. Si el vector $X = (X_1, \dots, X_n)$ es gaussiano, las variables X_1, \dots, X_n son gaussianas e independientes.

1.4 Distribuciones de formas cuadráticas

Definición 1.4.1 Se denota $\Gamma(a, \lambda)$ donde a, λ son reales positivos, la ley sobre $(\mathbb{R}^+, \mathfrak{B}_{\mathbb{R}^+})$ de densidad $f(x) = \frac{\lambda^a}{\Gamma(a)} e^{-\lambda x} x^{a-1}$.

Sabemos que su función característica es $\Phi_{\Gamma}(t) = (1 - \frac{it}{\lambda})^{-a}$.

- Si $\lambda = \frac{1}{2}$ y $a = \frac{1}{2}n$, con $n \in \mathbb{N}$, tenemos una χ_n^2 .
- Si $X \sim \Gamma(a, \lambda)$, entonces $Y = 2\lambda X \sim \Gamma(a, \frac{1}{2})$.

Definición 1.4.2 Se denota $\beta(a, b)$, con a, b reales positivos, a la ley beta sobre $(\mathbb{R}^+, \mathfrak{B}_{\mathbb{R}^+})$ de densidad $f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{x^{a-1}}{(1+x)^{a+b}}$.

Definición 1.4.3 Leyes gama y beta descentradas Sean a, b, γ, λ parámetros reales positivos, se define respectivamente las leyes gama y beta descentradas por:

$$\Gamma(a, \lambda, \gamma) = \sum_{m=0}^{\infty} \frac{e^{-\gamma} \gamma^m}{m!} \Gamma(a+m, \lambda), \quad \beta(a, b, \gamma) = \sum_{m=0}^{\infty} \frac{e^{-\gamma} \gamma^m}{m!} \Gamma(a+m, b),$$

donde γ es el parámetro de descentrage.

La función característica de la ley $\Gamma(a, \lambda, \gamma)$ es $(1 - \frac{it}{\lambda})^{-a} e^{it\gamma/(\lambda-it)}$.

Proposición 1.4.1 Sean X y Y variables aleatorias independientes de leyes $\Gamma(a, \lambda, \gamma)$ y $\Gamma(b, \lambda)$ respectivamente, entonces las variables aleatorias $U = \frac{X}{Y}$ y $V = X+Y$ tienen leyes respectivas $\beta(a, b, \gamma)$ y $\Gamma(a+b, \lambda, \gamma)$.

Si $\gamma = 0$, U y V son independientes.

Casos particulares

- Si $X \sim N(\mu, \sigma^2)$, $X^2 \sim \Gamma(\frac{1}{2}, \frac{1}{2\sigma^2}, \frac{\mu^2}{2\sigma^2})$.

- Se llama χ_n^2 de parámetro σ , a la ley $\Gamma\left(\frac{1}{2}n, \frac{1}{2\sigma^2}\right)$.
- Si la variable aleatoria $U \sim \beta(\frac{1}{2}n_1, \frac{1}{2}n_2)$ diremos que $\frac{n_1}{n_2}U \sim$ ley de Fisher²–Snedecor con n_1 y n_2 grados de libertad (g.d.l.).
- Si la variable aleatoria $U \sim \beta(\frac{1}{2}n_1, \frac{1}{2}n_2, \gamma)$ diremos que $\frac{n_1}{n_2}U \sim$ ley de Tang con n_1 y n_2 g.d.l.

Proposición 1.4.2 Sean X_1, \dots, X_n , n variables aleatorias independientes de igual ley $N(0, \sigma^2)$, entonces la variable aleatoria $Y = X_1^2 + \dots + X_n^2 \sim \chi_n^2$ de parámetro σ , es decir de densidad $\frac{1}{\sigma^n 2^{\frac{1}{2}} \Gamma(\frac{1}{2}n)} y^{\frac{1}{2}n-1} e^{-y/2\sigma^2}$ y la función característica $(1 - 2it\sigma^2)^{-\frac{1}{2}n}$.

Proposición 1.4.3 Sean X, Y variables aleatorias independientes de leyes respectivas χ_n^2 y χ_m^2 de igual parámetro σ , la variable aleatoria $Z = X/Y \sim \beta(\frac{1}{2}n, \frac{1}{2}m)$ de densidad:

$$\frac{\Gamma(\frac{1}{2}(n+m))}{\Gamma(\frac{1}{2}n)\Gamma(\frac{1}{2}m)} \frac{z^{\frac{1}{2}n-1}}{(1+z)^{\frac{1}{2}(n+m)}}.$$

²**Ronald Aylmer Fisher (1890-1962)** Nace el 17 de febrero de 1890 en East Finchley, cerca de Londres, Inglaterra. Muere el 29 de julio de 1962 en Adelaida, Australia. Ronald Fisher y realiza sus estudios de matemática en el Harrow College y en el Caius College de Cambridge y recibe un B.A. en Astronomía 1912. Estudió la Teoría de errores bajo la dirección de Stratton, usando el manual de Airy en la Teoría de Errores. Fisher estaba interesado en la Teoría de errores de las observaciones astronómicas y eventualmente lo llevó a investigar problemas estadísticos. Tenía una larga disputa con Pearson y al dársele un puesto bajo su dirección, escoge ir a Rothamsted. Fisher fue profesor de matemática en 1919 en la Estación Experimental Agrícola de Rothamsted, donde trabajó como biólogo e hizo muchas contribuciones en estadística y genética, hasta 1933. Allí estudió el Diseño de experimentos, introdujo el concepto de aleatorización y el análisis de varianza, procedimientos ahora usados en todo el mundo. En 1921 introdujo el concepto de verosimilitud. La verosimilitud de un parámetro es proporcional a la probabilidad de los datos y da una función que normalmente tiene un solo máximo, que llamó máximo de verosimilitud. En 1922 dio una nueva definición de estadística. Su propósito era la reducción de datos e identifica tres problemas fundamentales: (i) especificación de la clase de población que provienen los datos (ii) estimación y (iii) distribución. En 1933 fue nombrado profesor de Genética en el University College y de 1943 a 1957 ocupa la cátedra de Genética Arthur Balfour de Cambridge.

Las contribuciones que Fisher hizo incluyen el desarrollo de métodos satisfactorios para muestras pequeñas, como los de Gosset, el descubrimiento de las distribuciones muestrales de varias estadísticas y la invención del análisis de varianza. Introdujo el término de máxima verosimilitud y estudia las Pruebas de Hipótesis. Fisher se considera uno de los fundadores de la Estadística moderna debido a sus muchas importantes contribuciones. Fue electo Miembro de la Real Sociedad en 1929, se le otorgó la Medalla Real de la Sociedad en 1938 y se le otorgó Medalla Darwin de la Sociedad en 1948: ... en reconocimiento de sus contribuciones distinguidas a la teoría de selección natural, el concepto de su gen complejo y la evolución de dominancia. En 1955, se le otorgó la Medalla Copley de la Sociedad Real: ...en reconocimiento de sus contribuciones numerosas y distinguidas al desarrollo de la teoría y aplicación de la estadística para hacer cuantitativo el vasto campo de la biología. Una vez retirado, realiza trabajos investigación en la División de Estadística Matemática de la “Commonwealth Scientific and Industrial Research Organization”, en Adelaida, Australia, donde muere en 1962.

- Se llama ley de Student³ con n g.d.l. la ley de la variable aleatoria $\sqrt{n} \frac{X}{\sqrt{S}}$, donde X y S son variables aleatorias independientes de distribuciones respectivas $N(0, 1)$ y $\Gamma(\frac{1}{2}n, \frac{1}{2})$.

- Sean X_1, \dots, X_n , n variables aleatorias independientes $X_i \sim N(m_i, \sigma^2)$, $i = 1, \dots, n$, la variable aleatoria $Y = X_1^2 + \dots + X_n^2 \sim \chi_n^2$ descentrada $\Gamma(\frac{1}{2}n, \frac{1}{2\sigma^2}, \frac{1}{2\sigma^2}m)$, con $m = \sum_{i=1}^n m_i^2$, de densidad $f(y) =$

$$\sigma^{-n} 2^{-\frac{1}{2}n} e^{-\frac{1}{2\sigma^2}(m+y)} y^{\frac{1}{2}n-1} \sum_{j=0}^{\infty} \frac{\left(\frac{my}{4\sigma^4}\right)^j}{j! \Gamma(\frac{1}{2}(n+j))} \text{ y de función característica } (1 - 2it\sigma^2)^{-\frac{1}{2}n} e^{itm/(1-2it\sigma^2)}.$$

- Sean $X_1, \dots, X_n; Y_1, \dots, Y_p$ variables aleatorias normales independientes de igual varianza σ^2 . Consideremos que las variables Y_i , $i = 1, \dots, p$ están centradas y definimos $m = \sum_{i=1}^n (E(X_i))^2$, entonces la

variable aleatoria $T = \frac{X_1^2 + \dots + X_n^2}{Y_1^2 + \dots + Y_p^2} \sim \beta(\frac{1}{2}n, \frac{1}{2}p, \frac{m}{2\sigma^2})$, de densidad

$$f(t) = \frac{e^{-m/2\sigma^2}}{\Gamma(\frac{1}{2}p)} \frac{t^{\frac{1}{2}n-1}}{(1+t)^{\frac{1}{2}(n+p)}} \sum_{j=0}^{\infty} \frac{\Gamma(\frac{1}{2}(n+p) + j) \left(\frac{mt}{2\sigma^2}\right)^j}{\Gamma(\frac{1}{2}n + j) j! (1+t)^j}.$$

- Sean W_1, \dots, W_n variables aleatorias independientes, con $W_i \sim \chi_{p_i}^2 = \Gamma(\frac{1}{2}p_i, \frac{1}{2\sigma^2}, \frac{m_i^2}{2\sigma^2})$, $i = 1, \dots, n$, entonces $W = \sum_{i=1}^n W_i \sim \chi_{\sum_{i=1}^n p_i}^2 = \Gamma(\frac{1}{2}\sum_{i=1}^n p_i, \frac{1}{2\sigma^2}, \frac{1}{2\sigma^2}\sum_{i=1}^n m_i^2)$.

- Si $Y \sim N(\mu, \sigma^2 I)$, entonces $Y'Y/\sigma^2 \sim \chi_n^2 = \Gamma(\frac{1}{2}n, \frac{1}{2}, \frac{\mu'\mu}{2\sigma^2})$ y $Y'Y \sim \chi_n^2 = \Gamma(\frac{1}{2}n, \frac{1}{2\sigma^2}, \frac{\mu'\mu}{2\sigma^2})$.

³**William Sealy Gosset (1876-1937)** Nace el 13 de junio de 1876 en Canterbury, Inglaterra y muere el 16 de octubre de 1937 en Beaconsfield, Inglaterra. Era el mayor de los hijos de Agnes Sealy Vidal y el coronel Frederic Gosset que vino de Watlington a Oxfordshire. Se educó en Winchester y estudia matemática y química en el New College de Oxford, bajo la dirección de Airy. Lo obtuvo un primer grado de la clase en ambos temas, siendo concedido su grado de las matemáticas en 1897 y su grado de la química dos años más tarde. En 1899 trabaja como químico para la Compañía cervecera Guinness en Dublin, e hizo un trabajo importante sobre estadística. Esta sociedad, que favorecía la investigación, ponía sus laboratorios a disposición de sus químicos y en 1900, abre el "Guinness Research Laboratory" dirigido por el químico más grande de la época, Horace Brown. Los trabajos desarrollados por este laboratorio, se realizaban sobre la calidad y costos de diversas variedades de sorgo y cebada. Es en este ambiente que Gosset se interesa en la estadística.

En 1905 entró en contacto con Karl Pearson y fue a Londres a estudiar en el laboratorio de Pearson y el laboratorio de Galton en el período 1906-07. Aquí trabajó en el muestreo y el límite de Poisson de la distribución binomial, de la media, de la desviación estándar y del coeficiente de correlación. Estando en Oxford, a Gosset se le solicita ayuda sobre ciertos problemas que tenían sus colegas de Guinness. Es ahí donde trabaja los problemas del error en muestras pequeñas. Más adelante publicó tres importantes artículos en los años que estuvo en laboratorio de Pearson. En 1907, Gosset fue nombrado director Experimental de Guinness y utiliza la tabla de Student que había definido para determinar la mejor variedad de sorgo. Guinness autoriza la publicación de sus trabajos bajo el seudónimo que podría ser "Pupil" o "Student". Gosset escoge el segundo.

Gosset no trabajó ciertamente aisladamente. Estableció correspondencia con una gran cantidad de estadísticos y visitaba a menudo a su padre en Watlington, Inglaterra lo que le permitía visitar la Universidad de Londres y la Estación Agrícola Experimental de Rothamsted. Discutió problemas estadísticos con el Fisher, Neyman y Pearson. Para mucha gente es familiar el nombre Student, pero no el de Gosset.

- Sea $Y \sim N(\boldsymbol{\mu}, D)$, con D matriz diagonal, entonces $Y'D^{-1}Y \sim \chi_n^2 = \Gamma(\frac{1}{2}n, \frac{1}{2}, \frac{1}{2}\boldsymbol{\mu}'D^{-1}\boldsymbol{\mu})$.

Demostración Existe B no ortogonal tal que $B'DB = I$. Sea $Z = B'Y$ es decir $Z \sim N(B'\boldsymbol{\mu}, I)$, por lo tanto $\Phi_{Z|Z}(t) = (1 - 2it)^{-\frac{1}{2}n} e^{(it\boldsymbol{\mu}'BB'\boldsymbol{\mu})/(1-2it)}$, o sea $Y'D^{-1}Y$ tiene la distribución buscada.

Teorema 1.4.1 Sea la variable aleatoria $Y \sim N(0, I)$ y sea A una matriz simétrica, entonces la forma cuadrática $Y'AY \sim \chi_k^2 = \Gamma(\frac{1}{2}k, \frac{1}{2}) \iff A$ es idempotente, $\text{rang}(A) = k$.

Prueba

(\Leftarrow) Si A es idempotente de rango k , existe P matriz ortogonal tal que $P'AP = \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix}$. Sea

$$Z = P'Y, \text{ entonces } Y'AY = Z'P'APZ = Z_1'Z_1, \text{ donde } Z = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}.$$

Como $Z \sim N(0, I)$, $Z_1'Z_1 = \sum_{i=1}^k Z_i^2$, es decir $Y'AY \sim \chi_k^2 = \Gamma(\frac{1}{2}k, \frac{1}{2})$.

(\Rightarrow) Si A es simétrica existe C ortogonal tal que $C'AC = D$ es diagonal. Sea $W = C'Y$, entonces $Y'AY = W'C'ACW = W'DW = \sum_{i=1}^n W_i^2 d_{ii}$, pero $W \sim N(0, I)$ pues C es ortogonal, es decir $\Phi_{Y'AY}(t) = \prod_{i=1}^n (1 - 2itd_{ii})^{-\frac{1}{2}} = (1 - 2it)^{-\frac{1}{2}k}$.

Es claro que hay k valores $d_{ii} = 1$ y otros $n - k$ valores $d_{ii} = 0$, por lo que la matriz A es idempotente.

- Sea $Y \sim N(\boldsymbol{\mu}, I)$ y A una matriz simétrica, $\text{rang}(A) = k$, $Y'AY \sim \chi_k^2 = \Gamma(\frac{1}{2}k, \frac{1}{2}, \frac{1}{2}\boldsymbol{\mu}'A\boldsymbol{\mu}) \iff A$ es idempotente.

- Sea $Y \sim N(\boldsymbol{\mu}, \sigma^2 I)$ y sea A una matriz simétrica, $\text{rang}(A) = k$, entonces $Y'AY/\sigma^2 \sim \chi_k^2 = \Gamma(\frac{1}{2}k, \frac{1}{2}, \frac{1}{2\sigma^2}\boldsymbol{\mu}'A\boldsymbol{\mu}) \iff A$ es idempotente.

Además $Y'AY \sim \chi_k^2 = \Gamma(\frac{1}{2}k, \frac{1}{2\sigma^2}, \frac{1}{2\sigma^2}\boldsymbol{\mu}'A\boldsymbol{\mu}) \iff A$ es idempotente.

- Sea $Y \sim N(0, \Sigma)$ y sea B una matriz simétrica, $\text{rang}(B) = k$, entonces $Y'BY \sim \chi_k^2 = \Gamma(\frac{1}{2}k, \frac{1}{2}) \iff B\Sigma$ es idempotente.

Demostración Existe C matriz tal que $C'\Sigma C = I$. Sea $Z = C'Y$, entonces $Z \sim N(0, I)$, así $Y'BY = Z'C^{-1}BC'^{-1}Z \sim \chi_k^2 = \Gamma(\frac{1}{2}k, \frac{1}{2}) \iff C^{-1}BC'^{-1}$ es idempotente, o sea $B\Sigma$ es idempotente.

- Sea $Y \sim N(\boldsymbol{\mu}, \Sigma)$ y sea B una matriz simétrica de rango k , entonces $Y'BY \sim \chi_k^2 = \Gamma(\frac{1}{2}k, \frac{1}{2}, \frac{1}{2}\boldsymbol{\mu}'B\boldsymbol{\mu}) \iff$

$B\Sigma$ es idempotente.

Demostración Sea C tal que $C'\Sigma C = I$ y sea $Z = C'Y \sim N(C'\mu, I)$, entonces $Y'BY = Z'(C^{-1}BC'^{-1})Z \sim \chi_k^2 = \Gamma(\frac{1}{2}k, \frac{1}{2}, \frac{1}{2}\mu' B\mu) \iff C^{-1}BC'^{-1}$ es idempotente, es decir $B\Sigma$ es idempotente.

1.4.1 Teorema general sobre formas cuadráticas

Sea M una matriz simétrica de orden n , $X \sim N(\mu, \Sigma)$, la variable aleatoria $Z = X'MX$ tiene por función característica $\Phi_Z(t) = |I_n - 2itM\Sigma|^{-\frac{1}{2}} e^{it\mu'(I_n - 2itM\Sigma)^{-1}M\mu}$.

Si los valores propios no nulos de $M\Sigma$ son iguales a s y si $\mu'M\Sigma M\Sigma M\mu = \mu'M\Sigma M\mu$, entonces la variable aleatoria

$$Z - \mu' \left(M - \frac{M\Sigma M}{s} \right) \mu \sim \Gamma \left(\frac{1}{2}q, \frac{1}{2s}, \frac{1}{2s^2} \mu' M\Sigma M\mu \right).$$

Demostración Sea $A_{n \times r}$ la matriz tal que $\Sigma = AA'$, $A'MA = K$ del Lema 1.1.1. Sea Y un vector aleatorio de distribución $N(0, I_r)$, con $X = \mu + AY$ y sea $\mu' = (u_1, \dots, u_r) = \mu'MA$, entonces $Z = \mu'M\mu + 2\mu'Y + Y'KY = \mu'M\mu + \sum_{j=1}^r (s_j Y_j^2 + 2u_j Y_j)$. Sabemos que si $W \sim N(0, 1)$, entonces la variable aleatoria $sW^2 + 2uW$ tiene por función característica $(1 - 2its)^{-\frac{1}{2}} e^{-2u^2 t^2 / (1 - 2its)}$. Así la función característica de Z es igual a:

$$\Phi_Z(t) = \prod_{j=1}^r (1 - 2its_j)^{-\frac{1}{2}} e^{it\mu'M\mu - 2t^2 \sum_{j=1}^r u_j^2 / (1 - 2its_j)}.$$

Como $M\Sigma$ y K tienen los mismos valores propios, tenemos $\prod_{j=1}^r (1 - 2its_j) = |I_r - 2itK| = |I_n - 2itM\Sigma|$.

Por otro lado $\sum_{j=1}^r \frac{u_j^2}{1 - 2its_j} = \mathbf{u}'(I_r - 2itK)^{-1} \mathbf{u} = \mu'MA(I_r - 2itK)^{-1} A'M\mu$ y por lo tanto tenemos $\Phi_Z(t) = |I_n - 2itM\Sigma|^{-\frac{1}{2}} e^{it\mu'BM\mu}$, con B igual a $I_n + 2itMA(I_r - 2itK)^{-1}A'$.

Queda por demostrar que $B(I_n - 2itM\Sigma) = I_n$, o equivalentemente

$$-M\Sigma + MA(I_r - 2itK)^{-1}A'(I_n - 2itM\Sigma) = 0,$$

es decir $A' = (I_r - 2itK)^{-1}A'(I_n - 2itM\Sigma)$ i.e. $(I_r - 2itK)A' = A'(I_n - 2itM\Sigma)$ que se verifica de inmediato.

Si la matriz $M\Sigma$ posee q valores propios no nulos iguales a s , se verifica fácilmente que $(I_r - 2itK)^{-1} =$

$I_r + \frac{2itK}{1-2its}$ y se deduce que:

$$\sum_{j=1}^r \frac{u_j^2}{1-2its_j} = \boldsymbol{\mu}' M \Sigma M \boldsymbol{\mu} + \frac{2it}{1-2its} \boldsymbol{\mu}' M \Sigma M \Sigma M \boldsymbol{\mu}$$

y teniendo en cuenta las condiciones suplementarias de las hipótesis

$$\sum_{j=1}^r \frac{u_j^2}{1-2its_j} = \frac{\boldsymbol{\mu}' M \Sigma M \boldsymbol{\mu}}{1-2its},$$

con lo cual se concluye el teorema.

- Si Σ es regular $X' \Sigma^{-1} X \sim \Gamma\left(\frac{1}{2}n, \frac{1}{2}, \frac{1}{2} \boldsymbol{\mu}' \Sigma^{-1} \boldsymbol{\mu}\right)$.

- $E(Z) = \boldsymbol{\mu}' M \boldsymbol{\mu} + \text{tr}(\Sigma M)$.

En efecto $E(Z) = \boldsymbol{\mu}' M \boldsymbol{\mu} + \sum_{i=1}^r s_j = \boldsymbol{\mu}' M \boldsymbol{\mu} + \text{tr}(\Sigma M)$.

- Si $M \Sigma$ es idempotente, Σ regular, entonces $M \Sigma M \Sigma = M \Sigma$, es decir $M \Sigma M = M$ y la variable aleatoria

$Z = X' M X \sim \Gamma\left(\frac{1}{2}q, \frac{1}{2}, \frac{1}{2} \boldsymbol{\mu}' M \boldsymbol{\mu}\right)$, donde q es el rango de M .

1.5 Independencia de formas cuadráticas

Teorema 1.5.1 Sea $Y \sim N(\boldsymbol{\mu}, I)$, A, B matrices simétricas, entonces las formas cuadráticas $Y' A Y$ y $Y' B Y$ son independientes $\iff AB = 0$.

Demostración

(\Leftarrow) Supongamos que $AB = 0$, entonces $AB = BA = 0$, por lo que existe P matriz ortogonal tal que

$P' A P = D_1^*, P' B P = D_2^*$, con D_i^* matriz diagonal $i = 1, 2$. Pero como $AB = 0$, $D_1^* D_2^* = 0$, es decir

$$D_1^* = \begin{pmatrix} D_1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, D_2^* = \begin{pmatrix} 0 & 0 & 0 \\ 0 & D_2 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Sea $Z = P' Y$, entonces $Z \sim N(P' \boldsymbol{\mu}, I)$, es decir $Y' A Y = Z' D_1^* Z = \sum_{j=1}^{p_1} d_{1j} Z_j^2$, $Y' B Y = Z' D_2^* Z =$

$\sum_{j=p_1+1}^{p_1+p_2} d_{2j} Z_j^2$ y como los (Z_i) son independientes, $Y' A Y$ y $Y' B Y$ son independientes.

(\Rightarrow) Se deja de ejercicio.

- Si $Y \sim N(\boldsymbol{\mu}, I)$ y B_1, \dots, B_k son matrices simétricas, $Y' B_1 Y, \dots, Y' B_k Y$ son independientes \iff

$$B_i B_j = 0, i \neq j.$$

- Si $Y \sim N(\boldsymbol{\mu}, I)$, A_1, \dots, A_k son matrices simétricas, $Y' A_i Y$ formas cuadráticas tales que $\text{rang}(A_i) = n_i$, son independientes y $Y' A_i Y \sim \chi_{n_i}^2 = \Gamma(\frac{1}{2}n_i, \frac{1}{2}, \frac{1}{2}\boldsymbol{\mu}' A_i \boldsymbol{\mu}) \iff$ dos de las tres condiciones siguientes se verifican:

- 1) A_i idempotente, $i = 1, \dots, k$
- 2) $\sum_{i=1}^k A_i$ es idempotente
- 3) $A_i A_j = 0, i \neq j.$

- Si $Y \sim N(\boldsymbol{\mu}, I)$ y $Y' Y = \sum_{i=1}^k Y' A_i Y$, donde A_1, \dots, A_k son matrices simétricas, entonces $Y' A_i Y \sim \chi_{n_i}^2 = \Gamma(\frac{1}{2}n_i, \frac{1}{2}, \frac{1}{2}\boldsymbol{\mu}' A_i \boldsymbol{\mu})$ y son independientes \iff cualquiera de las dos condiciones se cumplen:

- 1) A_1, \dots, A_k son idempotentes
- 2) $A_i A_j = 0, i \neq j.$

- Si $Y \sim N(\boldsymbol{\mu}, I)$ y si $Y' Y = \sum_{i=1}^k Y' A_i Y$, donde A_1, \dots, A_k son matrices simétricas, entonces $Y' A_i Y \sim \chi_{n_i}^2 = \Gamma(\frac{1}{2}n_i, \frac{1}{2}, \frac{1}{2}\boldsymbol{\mu}' A_i \boldsymbol{\mu})$ y son independientes $\iff \sum_{i=1}^k \text{rang}(A_i) = \text{rang}(\sum_{i=1}^k A_i) = \text{rang}(I).$

Si $\boldsymbol{\mu} = 0$, el resultado se conoce como *Teorema de Cochran*.

- Si $Y \sim N(\boldsymbol{\mu}, \sigma^2 I)$ y si A, B simétricas, entonces si $\text{tr}(AB) = 0$ implica que $Y' A Y, Y' B Y$ son independientes, cuando A, B son semi-definidas positivas.

Prueba Como A y B son simétricas semidefinidas positivas, existen matrices G y H tales que $A = GG'$ y $B = HH'$. Como $\text{tr}(AB) = \text{tr}(GG'HH') = \text{tr}(G'H(G'H')) = 0$, se tiene $G'H = 0$, es decir $AB = 0$, con lo que se concluye la prueba.

- Sea $Y \sim N(\boldsymbol{\mu}, \sigma^2 I)$ y sea $Y' Y = \sum_{i=1}^k Y' A_i Y$, donde A_i es simétrica y de rango n_i , entonces:

- 1) $Y' A_i Y / \sigma^2 \sim \chi_{n_i}^2 = \Gamma(\frac{1}{2}n_i, \frac{1}{2}, \frac{1}{2\sigma^2}\boldsymbol{\mu}' A_i \boldsymbol{\mu})$
- 2) $Y' A_i Y, Y' A_j Y$ son independientes para $i \neq j \iff$ cualquiera de las tres condiciones siguientes se satisfacen:
 - i) A_i es idempotente, $i = 1, \dots, k$

- ii) $A_i A_j = 0, i \neq j$
 iii) $n_1 + \dots + n_k = n.$

1.6 Independencia de formas lineales y cuadráticas

- Sea una matriz $B_{q \times n}, A_{n \times n}$ una matriz simétrica y sea $Y \sim N(\mu, \sigma^2 I)$, entonces si $BA = 0$ se tiene que BY y $Y'AY$ son independientes.

- Si $Y \sim N(\mu, D)$, con D diagonal, $Y'AY \sim \chi_{n-1}^2 = \Gamma(\frac{1}{2}(n-1), \frac{1}{2}, \frac{1}{2}\mu' A \mu)$, si $A = D^{-1} - (D^{-1} J D^{-1}) / \mathbb{1}' D^{-1} \mathbb{1}$, donde $\mathbb{1}' = (1, \dots, 1)$ y $J = \mathbb{1} \mathbb{1}'$. Además si $\mu = \mathbb{1} \mu, \lambda = \frac{1}{2} \mu' A \mu = 0$.

- Sea $Y \sim N(\mu, \Sigma)$ y sean A y B matrices simétricas, entonces $Y'AY$ y $Y'BY$ son independientes si y sólo si $A \Sigma B = 0$.

- Sea $Y \sim N(0, \sigma^2 I)$, entonces si A es simétrica $E(Y'AY) = \sigma^2 \text{tr}(A)$.

Si A es idempotente de rango k , $E(Y'AY) = k \sigma^2$.

- Si $U \sim F(p, q, \lambda)$ entonces $V = U/k \sim F(r, q)$, donde $k = \frac{p+2\lambda}{p}, r = \frac{(p+2\lambda)^2}{(p+4\lambda)}$.

1.7 Ejercicios

1. Ley beta

a) Sabemos que $\int_0^1 x^{p-1} (1-x)^{q-1} dx = \beta(p, q)$ i.e. $f(x) = \frac{1}{\beta(p, q)} x^{p-1} (1-x)^{q-1} I_{[0,1]}(x)$ es la densidad de una variable aleatoria que sigue una distribución beta $\beta(p, q)$. A esta ley la llamaremos ley beta de 1ª especie.

Si una variable aleatoria X tiene por densidad $f(x) = \frac{1}{\beta(p, q)} \frac{x^{p-1}}{(1+x)^{p+q}} I_{[0,+\infty[}(x)$, diremos que X se distribuye como una beta de 2ª especie.

Demostrar que $\int_0^1 x^{p-1} (1-x)^{q-1} dx = \int_0^1 (1-x)^{p-1} x^{q-1} dx$.

b) Demostrar que $\int_0^1 x^{p-1} (1-x)^{q-1} dx = \int_0^\infty \frac{x^{p-1}}{(1+x)^{p+q}} dx = \int_0^\infty \frac{x^{q-1}}{(1+x)^{p+q}} dx$.

c) Demostrar que si $X \sim \Gamma(p, a)$ y $Y \sim \Gamma(q, a)$ y son independientes entonces:

- i) $\frac{X}{X+Y} \sim \beta(p, q)$ es de 1ª especie.
 ii) $\frac{X}{Y} \sim \beta(p, q)$ es de 2ª especie.

Observe que si $Z \sim \beta(p, q)$ de 2ª especie entonces, $T = \frac{1}{1+Z} \sim \beta(p, q)$ de 1ª especie.

2. **Ley geométrica** Sea $(E_i)_{i \in \mathbb{N}^*}$ una sucesión de eventos independientes definidos sobre el mismo espacio de probabilidad de igual probabilidad no nula p .

Se designa por X la variable aleatoria igual al índice del 1º evento que se realiza.

- Determinar la ley de probabilidad de la variable aleatoria X , su media y su varianza.
- Calcular la ley de probabilidad condicional de que $X = 2$, sabiendo que X es par. Generalice el resultado.
- Demostrar que si n y k son enteros positivos se tiene que $P(X = k + n | X > k) = P(X = n)$.

Enunciar el recíproco.

- Sea la variable aleatoria igual al índice i para el cual, por la primera vez, E_{i-1} y E_i se realizan.

Determinar la función generatriz de Y y deducir la ley de Y .

3. **Ley binomial negativa** Sea E un evento de probabilidad $p \neq 0$, y se efectúa una sucesión numerable de pruebas independientes. Sea r un entero positivo y se designa por Z_r la variable aleatoria igual al rango de la prueba en que se realiza E la r ésima vez.

- Determinar la ley de probabilidad de Z_r , su función generatriz su media y su varianza.

- Expresar sin cálculos la función de repartición de Z_r , con ayuda de una $B(n, p)$.

4. **Ley multinomial** Sean $A_j, j = 1, \dots, r$, eventos de probabilidad respectiva p_j tales que uno y sólo uno de estos eventos se pueda realizar. Se efectúan n pruebas independientes y se llama X el vector aleatorio de \mathbb{R}^r cuyas coordenadas X_j son iguales al n º de veces que se produjo el evento A_j .

- Dar su ley de probabilidad y su generatriz.

- Dar la esperanza de X y la matriz de covarianza.

- ¿Cuáles son las leyes de X_1 , de (X_1, X_2) , de la suma $X_1 + X_2$ y de (X_1, X_2) condicionada a la suma $X_1 + X_2$?

- Calcular la probabilidad de que las m ($m < r$) primeras componentes de X sean nulas. En el caso en que los p_j sean todos iguales, calcular la probabilidad de que las s primeras componentes de

$X(s \leq n)$ sea estrictamente positivas.

5. Ley uniforme

- a) Estudiar la ley uniforme sobre $[0, 1]$.
- b) Sean X_1, \dots, X_n , n variables aleatorias independientes de ley uniforme en $[0, 1]$. Encontrar las leyes de las variables aleatorias $W = \sup X_i$, $Z = \inf X_i$, $R = W - Z$, $Q = Z/W$, así como el coeficiente de correlación del par (Z, W) .

6. Ley gama

- a) Calcular la función característica, la media, varianza y los momentos de la ley $\Gamma(p, a)$.
- b) Demostrar que el cuadrado de una variable aleatoria normal y reducida sigue una ley gama.
- c) Sean X, Y variables aleatorias independientes de leyes respectivas $\Gamma(p, a)$ y $\Gamma(q, a)$. Determinar las leyes de las variables aleatorias $U = X + Y$, $Z = \frac{X}{Y}$, $T = \frac{X}{X + Y}$.

Calcular los momentos de las variables aleatorias Z y T existen.

- d) Sean X_1, \dots, X_n , n variables aleatorias independientes de leyes respectivas $\Gamma(p_i, a)$, $i = 1, \dots, n$.

i) Demostrar que las variables aleatorias $U = X_1 + X_2$, $T = \frac{X_1}{X_1 + X_2}$ son independientes.

ii) Generalizar el resultado anterior demostrando la independencia de las variables aleatorias Y_i , $i = 1, \dots, n$ definidas por:

$$Y_1 = \frac{X_1}{X_1 + X_2}, \quad Y_2 = \frac{X_1 + X_2}{X_1 + X_2 + X_3}, \dots, \quad Y_{n-1} = \frac{X_1 + \dots + X_{n-1}}{X_1 + \dots + X_n},$$

$$Y_n = X_1 + \dots + X_n.$$

7. **Ley de Weibull** Sea una variable aleatoria U de ley de probabilidad exponencial $\Gamma(1, a)$, sea $\alpha > 0$ y sea $X = U^{\frac{1}{\alpha}}$.

- a) Determinar la ley de X . Calcular sus momentos.
- b) Si n variables aleatorias independientes X_i , $i = 1, \dots, n$ siguen la ley de X , demostrar que la variable aleatoria $Y = \inf X_i$ sigue una ley del mismo tipo.
- c) Retomar las dos primeras preguntas del caso $\alpha < 0$.
- d) Dada una ley de probabilidad densidad $f(x)$ y de función de repartición $F(x)$ se llama taza de

sobrevivencia de esta ley cociente $q(x) = \frac{f(x)}{1 - F(x)}$. ¿Cuáles son las leyes de probabilidad cuyas tazas son constantes?

Caracterizar por la tasa de supervivencia las leyes obtenidas en la pregunta a).

8. Sea X una variable aleatoria de ley P_λ dependiendo de un parámetro, que es una variable aleatoria de la cual se da la ley. Se quiere dar la ley de X .
 - a) Estudiar el caso donde X y λ son variables aleatorias discretas tomando un número finito o numerable de valores; los valores tomados por X no dependen de λ .
 - b) Estudiar el caso en que λ es una variable aleatoria discreta y la ley P_λ de X es continua y admite una densidad f_λ . Calcular, en particular la función de repartición $F(x)$ y la densidad de X .
 - c) Se supone que λ es una variable aleatoria continua, con función de repartición $G(\lambda)$. Extender a este caso las fórmulas encontradas en a) y b). Verificar en todos los casos que son en realidad leyes de probabilidad es decir que su suma o integral es uno.
 - d) Calcular la función característica de X a partir de las funciones características de los P_λ .
9. Sean X_1, \dots, X_n, \dots una sucesión de variables aleatorias independientes de igual ley que X . Sea N una variable aleatoria entera no negativa independiente de la variable aleatoria X .
 - a) Calcular la función característica de la variable aleatoria $Y = X_1 + \dots + X_N$. Calcular la función generatriz de Y .
 - b) Aplicar el resultado de a) al caso en que las X_n son de Bernoulli y N sigue una distribución binomial o de Poisson.
10. Determinar la ley de una variable aleatoria X que es $\Gamma(\lambda, a)$ y cuyo parámetro λ sigue una ley geométrica de parámetro p .
11. **Leyes gama y beta descentradas**
 - a) Una variable aleatoria X sigue una ley $\Gamma(p + N, a)$, donde N es un parámetro aleatorio que sigue una ley de Poisson $P(\lambda)$. Calcular en forma de serie que no se trata de calcular, la densidad de la variable aleatoria X . La ley de X se llama gama descentrada y se denota $\Gamma(p, a, \lambda)$.

Calcular la función característica, la media y la varianza de la ley $\Gamma(p, a, \lambda)$.

b) Calcular la función característica del cuadrado de una variable aleatoria X que es $N(\mu, \sigma^2)$. Deducir que si n variables aleatorias independientes X_j son $N(m_j, \sigma^2)$, la variable aleatoria $U = \sum_{j=1}^n X_j^2$ es una ley gama que se precisará.

Demostrar que esta ley se determina también si $n-1$ variables $X_j \sim N(0, \sigma^2)$ y $X_n \sim N(\mu, \sigma^2)$, donde μ se precisará.

c) Determinar en forma de serie que no se intenta sumar la densidad de la variable aleatoria $Z = X/Y$, donde X, Y son independientes siguiendo respectivamente las leyes $\Gamma(p, a, \lambda)$ y $\Gamma(q, a)$. A esta ley se le llama beta descentrada $\beta(p, q, \lambda)$.

Se obtendrá el mismo resultado escribiendo $Z \sim \beta(p + N, q), N \sim P(\lambda)$.

12. Sean U_1, \dots, U_n variables aleatorias independientes de ley uniforme en $[0, 1]$ y sean las variables aleatorias $X = \sup U_i, Y = \inf U_i$.

Dar la ley de (X, Y) y determinar las leyes condicionales de X a Y y de Y a X . Calcular las esperanzas condicionadas y trace las curvas de regresión.

13. Sean X, Y, Z variables aleatorias tales que $X \sim U(0, 1)$, la ley de Y condicionada a $X = x$ tiene densidad:

$$f_x(y) = \begin{cases} (y-x)e^{-(y-x)} & 0 \leq x \leq 1, y > x \\ 0 & \text{si no} \end{cases}$$

y la ley de Z condicionada a $X = x, Y = y$ tiene densidad:

$$f_{XY}(z) = (y-x)e^{-z(y-x)} \quad \text{si } 0 \leq x \leq 1, y > x, z > 0.$$

a) Determinar la densidad de (X, Y, Z) .

b) Determinar la ley de Y . Trace la curva de regresión de Y relativa a X .

c) Determinar la ley de Z . Calcular sus momentos.

d) Determinar la superficie de regresión de Z relativa a X y Y .

e) Considerar las variables aleatorias definidas por $U = Y - X, V = Z(Y - X)$. Determinar la ley de

(X, U, V) . ¿Qué se concluye?

14. Sea X_1, \dots, X_n , n variables aleatorias independientes normales centradas y reducidas $X_i \sim N(0, 1)$.

Determinar la ley de probabilidad de la sucesión de variables aleatorias:

$$Y_j = \left(\frac{1}{j+1} \sum_{i=1}^{j+1} X_i - X_{j+1} \right) \left(1 + \frac{1}{j} \right)^{\frac{1}{2}}, \quad j = 1, \dots, n-1$$

$$Y_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i.$$

¿Qué característica tiene la matriz A tal que $Y = AX$? Determinar la ley de Y .

15. Sea X un vector aleatorio gaussiano cuyas $n + 1$ componentes Z, X_1, \dots, X_n son variables aleatorias $N(0, 1)$, donde:

$$\rho(Z, X_i) = \alpha, \quad i = 1, \dots, n$$

$$\rho(X_i, X_j) = \alpha^2, \quad 1 \leq i < j \leq n,$$

donde α es una constante dada.

Determinar la ley del vector X y deducir las leyes de $S = \frac{1}{n} \sum_{i=1}^n X_i$ y de $T = \frac{S}{Z}$.

Sugerencia Considerar el vector (Z, U_1, \dots, U_n) , donde las componentes U_i se definen por $X_i = \alpha Z + \sqrt{1 - \alpha^2} U_i$.

16. Sea X un vector aleatorio de ley $N(\mu, \Sigma)$ y sean A, B dos matrices que definen los vectores $Y = AX$, $Z = BX$. Demostrar que Y y Z son independientes $\iff A\Sigma B' = 0$.
17. Sean $X \sim N(\mu, \Sigma)$ y A, B matrices cuadradas simétricas definiendo las variables aleatorias $Y = X'AX$, $Z = X'BX$. Demostrar que Y y Z son independientes:

a) cuando $\mu = 0$, si $A\Sigma B\Sigma = 0$,

b) cuando $\mu \neq 0$, si $A\Sigma B = 0$.

Se puede verificar, que si M es una matriz cuadrada del mismo orden que Σ , $t \in \mathbb{R}$, $(I - 2itM\Sigma)^{-1} M = M(I - 2it\Sigma M)^{-1}$.

18. Sea $X \sim N(\mu, \Sigma)$ y sea M una matriz simétrica, demostrar que $E(X'MX) = \mu'M\mu + \text{tr}(M\Sigma)$. Extender el resultado al caso de M no simétrica.

Sugerencia Utilice el hecho de que si $Z = X'MX$ la función característica de Z es $\Phi_Z(t) = |I -$

$$2itM\Sigma|^{-\frac{1}{2}}.$$

19. Sean X, Y dos vectores normales centrados.
- Calcular $E(X'MY)$ donde M es una matriz de orden conveniente.
 - Demostrar que existe una matriz A y sólo una a precisar tal que $E((X - AY)^2)$ es mínima.
20. Sea X un vector gaussiano de ley $N(\mu, \sigma^2 I)$.
- Determinar la ley de $Y = \|X\|^2$ y deducir la ley de $Z = \|X\|$.
 - Efectuar un cambio a coordenadas polares para calcular la constante c , tal que la distribución uniforme es la esfera de centro μ y de radio c tenga los mismos momentos de orden 2 que la ley de X .
21. Sean X, Y dos vectores gaussianos independientes de igual ley $N(0, \Sigma)$.
- Demostrar que $X + Y$ y $X - Y$ son independientes. Deducir la función característica de la variable aleatoria $W = \langle X, Y \rangle$.
- Volver a encontrar este resultado considerando a W como forma cuadrática del vector $Z = (X, Y)$.
- Denotando $Z = \|X\|^2$, determinar la función característica de (W, Z) .
22. Sean Y_1, \dots, Y_n, n variables aleatorias independientes, con $Y_i \sim N(0, \sigma_i^2), i = 1, \dots, n$.
- Demostrar que las relaciones $Y_1 = X_1, Y_2 = X_2 - a_1 X_1, \dots, Y_n = X_n - a_{n-1} X_{n-1}$, donde a_1, \dots, a_{n-1} son reales fijos, definen un vector aleatorio (X_1, \dots, X_n) de ley normal multidimensional, cuya densidad se determinará.
 - Demostrar que para todo $k = 1, \dots, n - 1$, los dos vectores $U_k = (X_1, \dots, X_k)$ y $V_k = (Y_{k+1}, \dots, Y_n)$ son independientes.
 - Expresar los coeficientes $a_k (k = 1, \dots, n - 1)$ en función de σ_k, σ_{k+1} y de los coeficientes de correlación ρ_s , entre X_{s+1} y X_s , para $s = k, k - 1$.
 - Expresar el coeficiente de correlación ρ_{ij} de las variables X_i, X_j únicamente en función de ρ_s , con $j \leq s < i$.
23. a) Sea $X = (X_1, \dots, X_n)$ un vector normal de componentes independientes centradas y de misma

varianza σ^2 . Determinar la ley de X condicionada a $\|X\|^2$.

b) Si las componentes de X tienen una misma media $\mu \neq 0$, determinar la ley de X condicionada a $(\bar{X}, \|X\|^2)$.

24. Si y_1, y_2, y_3 tienen como forma cuadrática $Q = 2y_1^2 + 3y_2^2 + 4y_3^2 + 2y_1y_2 - 2y_1y_3 - 4y_2y_3 - 6y_1 - 6y_2 + 10y_3 + 8$, determinar:

a) A b) μ c) $f(y_1/y_2, y_3)$ d) $f(y_1, y_3)$ e) ρ_{23} .

25. Sea Y una variable multinormal, con media $\mu = 0$ y covarianza $\Sigma = \begin{pmatrix} 4 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 3 \end{pmatrix}$, determinar:

a) A b) $f(y_1)$ c) Q d) $Z = 4Y_1 - 6Y_2 + Y_3$ su ley, media y varianza.

26. La función de densidad de una variable aleatoria binormal está dada por:

$$f(x, y) = \frac{1}{\sigma_x \sigma_y \sqrt{1 - \rho^2} 2\pi} e^{-\frac{1}{2(1-\rho^2)} \left(\left(\frac{x-\mu_x}{\sigma_x} \right)^2 - 2\rho \left(\frac{x-\mu_x}{\sigma_x} \right) \left(\frac{y-\mu_y}{\sigma_y} \right) + \left(\frac{y-\mu_y}{\sigma_y} \right)^2 \right)}.$$

Determinar:

a) A b) Σ c) $f(x/y)$ d) Probar que $\text{var}(X) = \sigma_x^2$, $\text{var}(Y) = \sigma_y^2$, $\text{cov}(X, Y) = \sigma_x \sigma_y \rho$.

27. Sea $X_1 \sim N(\mu, \sigma^2)$, $X_2/X_1 = x_1 \sim N(x_1, \sigma^2)$. Probar que:

$$(X_1, X_2) \sim N\left(\begin{pmatrix} \mu \\ \mu \end{pmatrix}, \begin{pmatrix} \sigma^2 & \sigma^2 \\ \sigma^2 & 2\sigma^2 \end{pmatrix}\right).$$

28. Sean $X_1 \sim N(\mu_1, \Sigma_1)$ y $X_2 \sim N(\mu_2, \Sigma_2)$ variables aleatorias independientes. Dar la ley de $Z = X_1 + X_2$, es decir determinar μ_Z y Σ_Z .

29. Dar la media y la varianza de un $\chi_n^2 = \Gamma\left(\frac{1}{2}n, \frac{1}{2\sigma^2}, \frac{m}{2\sigma^2}\right)$.

30. Dar la media y la varianza de una $F(p, q, \lambda) = \beta\left(\frac{1}{2}p, \frac{1}{2}q, \lambda\right) \frac{q}{p}$.

31. Si B es una matriz $n \times p$, $n > p$, $\text{rang}(B) = p$, probar que $A = B(B'B)^{-1}B'$ es idempotente.

32. Si $Y \sim N(0, I)$, determinar la distribución de:

a) $Y'B(B'B)^{-1}B'Y$. b) $Y'(I - B(B'B)^{-1}B')Y$. c) $E(Y'(I - B(B'B)^{-1}B')Y)$.

33. Sea $Y \sim N(B\beta, I)$, donde β es un vector $p \times 1$ desconocido, B es la matriz del problema 31. Determinar la distribución de:

a) $Q_1 = Y'(B(B'B)^{-1}B')Y$.

b) $Q_2 = Y'Y - Q_1$.

c) Probar que Q_1 y Q_2 son independientes.

d) Dar la ley de $U = \frac{n-p}{p} \frac{Q_1}{Q_2}$.

e) $E(U)$.

34. Sea $Y \sim N_3(0, I_3)$. Determinar $E(Y'AY)$ y $E(Y'BY)$ así como la distribución de $Y'AY$, $Y'BY$ y

$$U = Y'AY/Y'BY, \text{ si } A = \begin{pmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & -\frac{1}{3} & \frac{2}{3} \end{pmatrix}, \quad B = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}.$$

35. Sea $Y \sim N(0, I)$, determinar:

a) La matriz de $Q_1 = n\bar{Y}^2$. b) La matriz de $\sum_{i=1}^n (Y_i - \bar{Y})^2 = Q_2$.

c) La ley de Q_1 . d) La de Q_2 .

e) Q_1, Q_2 son independientes. f) La ley de Q_1/Q_2 .

g) Probar que $\mathbb{1}'Y$ y $\sum_{i=1}^n (Y_i - \bar{Y})^2$ son independientes.

h) $E(Q_1), E(Q_2)$.

36. Si $Y \sim N(\mu, \sigma^2 I)$, determinar $E(Y'AY)$, con A idempotente $\text{rang}(A) = k$.

37. Sea $Y \sim N(\mu, \Sigma)$, donde $\mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}$, $\Sigma = \begin{pmatrix} \sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_n^2 \end{pmatrix}$.

Probar que $\sum_{i=1}^n \frac{1}{\sigma_i^2} (Y_i - Y^*)^2 \sim \chi_{n-1}^2 = \Gamma\left(\frac{1}{2}(n-1), \frac{1}{2}, \lambda\right)$, donde $\lambda = 0$ si $\mu_1 = \mu_2 = \cdots = \mu_n$. El

término Y^* es igual a $Y^* = \frac{\sum_{i=1}^n (1/\sigma_i^2) Y_i}{\sum_{i=1}^n 1/\sigma_i^2}$.

38. Sean X_1, \dots, X_n una muestra de n variables aleatorias independientes, con media μ y varianza σ^2 .

Sea $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ y $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$.

a) Demostrar que $E(\bar{X}) = \mu$, $\text{var}(\bar{X}) = \frac{\sigma^2}{n}$.

b) Demostrar que $E(S^2) = \sigma^2$, $\text{var}(S^2) = \frac{1}{n} \left(m_4 - \frac{n-3}{n-1} \sigma^4 \right)$, donde m_4 es el momento centrado de orden 4.

c) Demostrar que $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$, $\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$.

d) Verificar que $\frac{\sqrt{n}(\bar{X} - \mu)}{S} \sim t_{n-1}$, de Student con $n - 1$ g.d.l.

39. Sean \bar{X}_1, \bar{X}_2 medias muestrales y si S_1^2, S_2^2 las varianzas muestrales de dos muestras independientes de tamaño n_1 y n_2 de distribución $N(\mu_1, \sigma^2)$ y $N(\mu_2, \sigma^2)$ respectivamente.

Probar que $\frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1 + n_2 - 2}}} \sim t_{n_1+n_2-2}$.

40. Sean $(X_{11}, \dots, X_{1n}), (X_{21}, \dots, X_{2n})$ muestras independientes de distribuciones $N(\mu_1, \sigma_1^2)$ y $N(\mu_2, \sigma_2^2)$ respectivamente.

Sean \bar{X}_1, \bar{X}_2 las medias muestrales y S_1^2, S_2^2 las varianzas muestrales y sea:

$$r = \frac{1}{(n-1)S_1S_2} \sum_{i=1}^n (X_{1i} - \bar{X}_1)(X_{2i} - \bar{X}_2),$$

el coeficiente de correlación entre las dos muestras. Probar que:

$$\frac{\sqrt{n}(X_1 - X_2 - (\mu_1 - \mu_2))}{\sqrt{S_1^2 + S_2^2 - 2rS_1S_2}} \sim t_{n-1}.$$

41. Sean \bar{X}_1, \bar{X}_2 medias muestrales de poblaciones independientes de tamaño n_1, n_2 , con igual varianza σ^2 . Sea \bar{X} la media de las dos poblaciones al considerarse conjuntamente (de tamaño $n_1 + n_2$). Probar que la varianza de $\bar{X} - \bar{X}_1$ es $\frac{\sigma^2 n_2}{n_1(n_1 + n_2)}$.

42. Si $X \sim N(\mu, \sigma^2)$ y si $F(x)$ es la función de repartición de X , probar que el coeficiente de correlación entre X y $F(X)$ es $\sqrt{3/\pi}$.

43. Se consideran k muestras de tamaño n_1, \dots, n_k independientes de distribución común $N(\mu, \sigma^2)$. Sean \bar{X}_i y S_i^2 la media y varianzas muestrales de la i^{ava} muestra $i = 1, \dots, k$. Sean:

$$U = \frac{1}{\sigma^2} [(n_1 - 1)S_1^2 + \dots + (n_k - 1)S_k^2]$$

$$V = \frac{1}{\sigma^2} [n_1(\bar{X}_1 - \bar{X})^2 + \dots + n_k(\bar{X}_k - \bar{X})^2]$$

$$W = \frac{1}{\sigma^2} n(\bar{X} - \mu),$$

donde \bar{X} es la media de todas las muestras en conjunto, $n = \sum_{i=1}^k n_i$.

Probar que U, V, W son independientes y de distribuciones $\chi_{n-k}^2, \chi_{k-1}^2, \chi_1^2$ respectivamente.

44. Sean Y_1, \dots, Y_n , n variables aleatorias independientes de distribuciones de la forma $N(\mu + \beta x_1, \sigma^2), \dots, N(\mu + \beta x_n, \sigma^2)$ respectivamente, donde $\mu, \beta, x_1, \dots, x_n$ son constantes.

Sean $\hat{\mu}$ y $\hat{\beta}$ los valores de μ y β que minimizan la suma de cuadrados $\sum_{i=1}^n (y_i - \mu - \beta x_i)^2$ i.e. $v = \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \hat{\mu} - \hat{\beta} x_i)^2$ es mínimo.

Probar que $\begin{pmatrix} \hat{\mu} \\ \hat{\beta} \end{pmatrix} \sim N_2\left(\begin{pmatrix} \mu \\ \beta \end{pmatrix}, V\right)$, donde $V = \begin{pmatrix} \frac{\sigma^2 \sum_{i=1}^n x_i^2}{na_n} & -\frac{\bar{x}\sigma^2}{a_n} \\ -\frac{\bar{x}\sigma^2}{a_n} & \frac{\sigma^2}{a_n} \end{pmatrix}$, $a_n = \sum_{i=1}^n (x_i - \bar{x})^2$ y que $v \sim \chi_{n-2}^2$. Probar además que $(\hat{\mu}, \hat{\beta})$ y v son independientes.

45. Sea X_1, \dots, X_n una muestra de una distribución de media μ y varianza σ^2 . Sea $f(x)$ una función derivable en $x = \mu$ tal que $f'(\mu) \neq 0$. Probar que $f(\bar{X})$ tiene distribución asintótica $N(f(\mu), (\sigma f'(\mu))^2/n)$, cuando $n \rightarrow \infty$.
46. Sea X_1, \dots, X_n una muestra de una distribución de Poisson $P(\mu)$, probar que $2\sqrt{\bar{X}}$ tiene distribución asintótica $N(2\sqrt{\mu}, \frac{1}{n})$, cuando $n \rightarrow \infty$.
47. Sea X_1, \dots, X_n , una muestra de distribución $U(0, \theta)$. Probar que $\sqrt{12} \log(2\bar{X})$ tiene una distribución asintótica $N(\sqrt{12} \log \theta, \frac{4}{n})$, cuando $n \rightarrow \infty$.
48. Si \bar{X}_i, S_i^2, n_i son la media muestral, la varianza muestral y el tamaño de la muestra i , de parámetros $\mu_i, \sigma_i^2, i = 1, 2$, tales que son independientes. Probar que:

$$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \xrightarrow{\text{ley}} N(0, 1).$$

Capítulo 2

Inferencia estadística

2.1 Introducción a la estadística inferencial

Dentro de la Estadística podemos distinguir dos grandes campos: La estadística descriptiva y La estadística decisional o inferencial.

La estadística descriptiva tiene por objeto describir los datos a través de de índices como la media, la varianza, el coeficiente de correlación, por medio de gráficos como histogramas o bien resúmenes de datos a través de los métodos en componentes principales.

La estadística inferencial por otro lado busca estimar parámetros, probar hipótesis y tomar decisiones considerando algunos factores tales como costos. Así se parte de observaciones y se trata de *aproximarse* a una ley de probabilidad. Por ejemplo consideramos la fabricación en serie de un objeto y supongamos que en una muestra de n objetos se observan t objetos defectuosos. Si se admite que los objetos defectuosos se distribuyen en el lote independientemente, con probabilidad p (a priori desconocida), la ley de probabilidad de la variable aleatoria $T = n^o$ de objetos defectuosos es una binomial $B(n, p)$. Observemos que como p es desconocida a priori, tenemos una familia de leyes binomiales $\{B(n, p)/0 \leq p \leq 1\}$.

Sabemos por otro lado, que T/n tiende en probabilidad a p , de donde si n es grande se puede pensar que T/n esté próximo de p . Esto es un razonamiento de tipo estadístico, que consiste en inferir sobre una ley de probabilidad a partir de una muestra u observación. Es claro que esta inferencia está siempre *más o menos* sujeta a un error. Es este error el que es necesario tratar de definir y conocer

por un lado y minimizar por otro lado.

En el problema tratado anteriormente se puede estimar el valor de p , con el estimador T/n (problema de estimación).

Una vez recolectados los datos se debe resolver el problema de decisión siguiente:

- se deja funcionar la máquina o
- se detiene para revisión,

pues si se deja funcionando hay pérdidas (objetos defectuosos, etc), pero si se detiene también hay pérdidas.

Así a estas operaciones están asociados costos; ¿cuál es el costo menor? Esta es un problema típico de decisión.

Por otro lado si dos máquinas fabrican objetos (o parte de ellos) en serie podríamos preguntar ¿cuál es la más eficiente? Para contestar a esto efectuamos dos muestras u observaciones para saber si p_1 de la máquina n^o1 es menor que p_2 de la máquina n^o2 i.e. $p_1 < p_2$. Esto es típico en pruebas de hipótesis.

2.2 Formalismo matemático

2.2.1 Definición de estructura estadística

Se llama estructura estadística al triplete $(\Omega, \mathfrak{A}, \mathfrak{P})$, donde \mathfrak{P} es una familia de leyes de probabilidad sobre (Ω, \mathfrak{A}) . A menudo \mathfrak{P} se indexa con un parámetro $\theta \in \Theta$; la estructura se escribe $(\Omega, \mathfrak{A}, \{P_\theta/\theta \in \Theta\})$.

Ejemplos

- Si se extraen al azar 100 piezas de un lote de objetos de fabricación en serie, definimos el fenómeno de esta manera:

$\Omega = \{0, 1, 2, \dots, 100\}$, $\mathcal{P}(\Omega)$, $\omega = n^o$ objetos defectuosos, $P_p(\omega) = \binom{100}{\omega} p^\omega (1-p)^{100-\omega}$, $\mathfrak{P} = \{P_p/p \in [0, 1]\}$.

- Una máquina empaqueta azúcar en bolsas de 500 gramos; si se asume normalidad la estructura del fenómeno sería, $\Omega = \mathbb{R}$, $\mathfrak{A} = \mathfrak{B}_{\mathbb{R}}$, $\mathfrak{P} = \{N(\mu, \sigma^2)/\mu = 500, \sigma > 0\}$.

Si no se acepta la normalidad, sino que se supone que la ley de probabilidad es absolutamente continua, con respecto a la medida de Borel¹ en \mathbb{R} , la estructura asociada sería $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}}, \mathfrak{P}_c)$, donde \mathfrak{P}_c es la familia de leyes de probabilidad sobre $\mathfrak{B}_{\mathbb{R}}$, dominadas por la medida de Borel.

2.3 Estructura dominada

Se dice que la estructura $(\Omega, \mathfrak{B}, \mathfrak{P})$ está dominada si existe una medida σ -finita μ sobre \mathfrak{B} tal que $\forall P \in \mathfrak{P}$ se tiene que $P \ll \mu$.

Decir que la estructura $(\Omega, \mathfrak{B}, \{P_\theta : \theta \in \Theta\})$ está dominada por μ , es decir que existe una densidad $(\omega \rightarrow f_\theta(\omega))$ de P_θ , con respecto a μ , para todo θ . Si se escoge una determinación de estas densidades, se llama función de verosimilitud la función $L(\theta, \omega)$ definida sobre $\Theta \times \Omega$ por:

$$L(\theta, \omega) = f_\theta(\omega).$$

La medida dominante de una estructura no es única. Si μ domina la estructura y $\mu \ll \mu'$, entonces μ' domina la estructura y se tiene:

$$\frac{dP_\theta}{d\mu'} = \frac{dP_\theta}{d\mu} \frac{d\mu}{d\mu'}.$$

El teorema siguiente muestra en particular que, si la estructura está dominada se puede siempre escoger como medida dominante una medida de probabilidad equivalente a \mathfrak{P} .

¹**Félix Edouard Justin Emile Borel (1871-1956)** Nace en Francia en 1871 e ingresa al Politécnico y a la Escuela Normal Superior a los 18 años. Fue profesor de la Facultad de Ciencias de París-Sorbona, Director Científico de la Escuela Normal Superior, diputado y Ministro de la Marina (1925).

Se le debe, financiado en parte por el barón de Rothschild, la creación (1928) del Instituto Henri Poincaré para la Investigación en Matemática Pura y Aplicada (hoy adjunto al C.N.R.S.). Miembro de la Oficina de las Medidas, fue miembro de la Academia de Ciencias (1921) y Presidente en 1934. Recibió la primera medalla de oro del CNRS (1954).

Borel es considerado como un matemático constructivista. Originó la teoría de los juegos estratégicos y la cibernética que desarrollaron Von Neumann y Morgenstern. Muy importantes son sus trabajos en teoría de conjuntos; topología y en teoría de la medida (1897), que anuncia la integral de Lebesgue (1901); las funciones de variable compleja (funciones meromorfas); cálculo de probabilidades, espacios probabilizados sobre una tribu de un espacio topológico; el concepto de σ -álgebra (tribu), la medida en el sentido topológico, que generaliza el concepto de medida de un segmento, o de una área, que es indisociable de la nueva teoría de integración de Lebesgue establecida en 1902.

Teorema 2.3.1 Una estructura estadística $(\Omega, \mathfrak{A}, \mathfrak{P})$ es dominada \iff existe una ley de probabilidad P^* sobre (Ω, \mathfrak{A}) llamada dominante privilegiada que domina la estructura estadística $(\Omega, \mathfrak{A}, \mathfrak{P})$ tal que:

- a) P^* es absolutamente continua con respecto a toda medida dominante de $(\Omega, \mathfrak{A}, \mathfrak{P})$.
- b) P^* es combinación estrictamente convexa de una subfamilia \mathfrak{P}' a lo sumo numerable de \mathfrak{P} , $P^* = \sum_{P \in \mathfrak{P}'} \alpha_P P$, $\alpha_P > 0$, $\sum_{P \in \mathfrak{P}'} \alpha_P = 1$.
- c) P^* es equivalente a \mathfrak{P} , es decir $\forall A \in \mathfrak{A}$, $((P(A) = 0, \forall P \in \mathfrak{P}) \iff P^*(A) = 0)$.

La demostración reposa en el lema siguiente.

Lema 2.3.1 La estructura estadística $(\Omega, \mathfrak{A}, \mathfrak{P})$ es dominada \iff existe una subfamilia a lo sumo numerable \mathfrak{P}' de \mathfrak{P} tal que $\forall A \in \mathfrak{A}$, $(P(A) = 0, \forall P \in \mathfrak{P}') \implies (P(A) = 0, \forall P \in \mathfrak{P})$.

Nota Nosotros admitiremos el lema. Los interesados pueden referirse a Tortrat–Hennequin: Theorie des probabilités et quelques applications–Masson 1965.

Demostración del teorema Sea P^* una combinación estrictamente convexa de las leyes de \mathfrak{P}' ; es una probabilidad sobre (Ω, \mathfrak{A}) que satisface las condiciones del teorema. En efecto si μ es una medida dominando a $(\Omega, \mathfrak{A}, \mathfrak{P})$ se tiene:

$$\begin{aligned} \mu(A) = 0 \implies (P(A) = 0, \forall P \in \mathfrak{P}) &\iff (P(A) = 0, \forall P \in \mathfrak{P}') \\ &\iff P^*(A) = 0. \end{aligned}$$

Nota La existencia de P^* es evidente si las medidas $P \in \mathfrak{P}$ son equivalentes: es suficiente tomar como P^* cualquiera de las probabilidades $P \in \mathfrak{P}$; sin embargo en general $P^* \notin \mathfrak{P}$. Por otro lado P^* no es única.

2.4 Estadísticas–Estadísticas Integrables

Definición 2.4.1 Sea $(\Omega, \mathfrak{A}, \mathfrak{P})$ una estructura estadística, se llama estadística sobre la estructura estadística $(\Omega, \mathfrak{A}, \mathfrak{P})$ una aplicación medible X de (Ω, \mathfrak{A}) en un espacio medible (Ω', \mathfrak{A}') , o sea

$$(\Omega, \mathfrak{A}, \mathfrak{P}) \xrightarrow{X} (\Omega', \mathfrak{A}').$$

- Si $\Omega' = \mathbb{R}$, X es una estadística real.
- Si $\Omega' = \mathbb{R}^n$, X es una estadística vectorial.

2.4.1 Estructura inducida

Llamamos estructura inducida por la estadística X , a la estructura estadística $(\Omega', \mathfrak{A}', \mathfrak{P}_X)$, donde $\mathfrak{P}_X = \{P_X/P \in \mathfrak{P}\}$.

Definición 2.4.2 Sean X_1, X_2 estadísticas definidas de $(\Omega, \mathfrak{A}, \mathfrak{P})$ en (Ω', \mathfrak{A}') y $(\Omega'', \mathfrak{A}'')$ respectivamente.

Se dice que X_1 es equivalente a X_2 si $X_1^{-1}(\mathfrak{A}') = X_2^{-1}(\mathfrak{A}'')$.

Observemos que esta definición no depende de \mathfrak{P} ; lo cual debe distinguirse de la definición de equivalencia de dos variables aleatorias en el sentido de igualdad c.p.d.

Si X_1 y X_2 son biyectivas bicontinuas, son equivalentes.

Definición 2.4.3 Se dice que un evento $A \in \mathfrak{A}$ es de medida \mathfrak{P} cero si $P(A) = 0$, $\forall P \in \mathfrak{P}$. Esta definición justifica la noción \mathfrak{P} -c.p.d. a menudo utilizada.

Definición 2.4.4 Sean X_1, X_2 estadísticas sobre $(\Omega, \mathfrak{A}, \mathfrak{P})$, con valores en (Ω', \mathfrak{A}') ; se dice que X_1 es \mathfrak{P} -equivalente a X_2 y se denota $X_1 \stackrel{\mathfrak{P}}{\sim} X_2$, si el evento $[X_1 \neq X_2]$ es de medida \mathfrak{P} -cero.

Ejercicio Sea $(\Omega, \mathfrak{A}, \mathfrak{P})$ una estructura estadística dominada, sea P^* una ley de probabilidad dominante privilegiada, un evento es de medida \mathfrak{P} -cero si y sólo si es de medida P^* -cero.

Definición 2.4.5 Se dice que las estadísticas X y Y son independientes, si $\forall P \in \mathfrak{P}$ las variables aleatorias X y Y son independientes.

Ejemplo Sobre la estructura estadística $(\mathbb{R}^2, \mathfrak{B}_{\mathbb{R}^2}, \{N(\mu, \sigma^2 I_2)/\mu \in \mathbb{R}^2, \sigma > 0\})$ las estadísticas $X(x, y) = x + y$, $Y(x, y) = x - y$ son independientes.

Definición 2.4.6 Se dice que la estadística real X sobre $(\Omega, \mathfrak{A}, \mathfrak{P})$ es integrable si $\forall P \in \mathfrak{P}$, la variable aleatoria X es integrable:

$$\forall P \in \mathfrak{P}, E_P(X) = \int_{\Omega} X dP \text{ existe.}$$

Definición 2.4.7 Si X es una estadística con valores en \mathbb{R}^n se dice que es integrable si cada componente es integrable.

Definición 2.4.8 Se dice que la estadística escalar integrable X es libre en media (respectivamente centrada) si $E_P(X)$ no depende de $P \in \mathfrak{P}$ (respectivamente $E_P(X) = 0, \forall P \in \mathfrak{P}$).

Ejemplo Sea la estructura estadística $(\mathbb{R}^2, \mathfrak{B}_{\mathbb{R}^2}, \mathfrak{P})$, con $\mathfrak{P} = \{P_\theta / \frac{dP_\theta}{d\mu} = e^{-\theta x - y/\theta}, x, y \geq 0, \theta > 0\}$.

La estadística $T: \mathbb{R}^2 \rightarrow \mathbb{R}$ tal que $T(x, y) = xy - 1$ es centrada.

Definición 2.4.9 Se llama imagen de la estadística integrable X sobre la estructura estadística $(\Omega, \mathfrak{A}, \{P_\theta : \theta \in \Theta\})$, la función $\beta_X: \Theta \rightarrow \mathbb{R}$

$$\theta \mapsto \beta_X(\theta) = E_{P_\theta}(X) = \int X dP_\theta.$$

2.5 Estructuras completas

Definición 2.5.1 Sea $(\Omega, \mathfrak{A}, \mathfrak{P})$ una estructura estadística y sea \mathfrak{B} sub-tribu de \mathfrak{A} , se dice que la estructura $(\Omega, \mathfrak{B}, \mathfrak{P})$ es completa, si toda estadística X \mathfrak{B} -medible tal que:

$$E_P(X) = 0, \quad \forall P \in \mathfrak{P}, \quad \text{es } \mathfrak{P}\text{-equivalente a cero i.e. } X = 0, \quad \mathfrak{P}\text{-c.p.d.}$$

Definición 2.5.2 Sea $(\Omega, \mathfrak{A}, \mathfrak{P})$ una estructura estadística y sea \mathfrak{B} sub-tribu de \mathfrak{A} , se dice que la estructura $(\Omega, \mathfrak{B}, \mathfrak{P})$ es cuasi-completa, si toda estadística X \mathfrak{B} -medible y acotada tal que:

$$E_P(X) = 0, \quad \forall P \in \mathfrak{P}, \quad \text{es } X = 0, \quad \mathfrak{P}\text{-c.p.d.}$$

Definición 2.5.3 Una estadística T se dice completa, si la estructura $(\Omega, T^{-1}(\mathfrak{A}'), \mathfrak{P})$ es completa. (Idem para cuasi-completa).

Es claro que si una estructura es completa, es cuasi-completa.

El recíproco no es cierto. En efecto consideremos la familia de leyes P_θ definidas sobre $\Omega = \mathbb{N} \cup \{-1\}$, $\mathfrak{A} = \mathcal{P}(\Omega)$, con $P_\theta(\{-1\}) = 0, P_\theta(\{n\}) = (1 - \theta)^2 \theta^n, n \geq 0$, la estructura estadística $(\Omega, \mathfrak{A}, \{P_\theta : \theta \in [0, 1]\})$ es cuasi-completa sin ser completa. En efecto (salvo por homotecia) la única estadística no nula centrada está definida por:

$$f(-1) = -1, \quad f(n) = n \quad \text{si } n \geq 0.$$

- \mathfrak{F} -c.p.d. \equiv la propiedad es cierta $\forall \omega \in \Omega_0 : P(\Omega_0) = 1, \forall P \in \mathfrak{F}$.
- Se dice también que la tribu \mathfrak{B} es completa (en vez de la estructura) pero es peligroso, pues es evidente que una estructura $(\Omega, \mathfrak{B}, \mathfrak{F})$ es más fácilmente completa en cuanto \mathfrak{B} es *pequeña* y en el límite $\mathfrak{B} = \{\emptyset, \Omega\}$ es siempre completa. El término completa se refiere más bien a la familia \mathfrak{F} , que debe ser suficiente basta.
- Sea $(\Omega, \mathfrak{A}, \{f_\theta : \theta \in \Theta\})$ una estructura estadística dominada por la ley privilegiada P^* . La estructura es completa si y sólo si $\forall \theta \in \Theta, \int X f_\theta dP^* = 0 \implies X = 0$ P^* -c.p.d.
- La estadística T es completa (respectivamente cuasi-completa) si para toda función medible $h : (\Omega', \mathfrak{A}') \longrightarrow (\mathbb{R}, \mathfrak{B}_{\mathbb{R}})$ tal que $h(T)$ sea integrable (respectivamente acotada) se tiene:

$$E_P(h(T)) = 0, \forall P \in \mathfrak{F} \implies h = 0, \mathfrak{F}_T\text{-c.p.d.}$$

Ejercicio

- a) Demostrar que las siguientes estructuras son completas:
- $(\Omega = \{0, 1, \dots, n\}, \mathcal{P}(\Omega), B(n, p) : p \in [0, 1])$
 - $(\mathbb{N}, \mathcal{P}(\mathbb{N}), P(\lambda) : \lambda > 0)$
 - $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}, N(\mu, 1); \mu \in \mathbb{R}})$
 - $(\mathbb{R}^+, \mathfrak{B}_{\mathbb{R}^+}, f_\theta(x) = \Gamma(p, \theta) : \theta \in \mathbb{R}^+, p > 0)$.
- b) Encontrar una ley dominante privilegiada P^* si existe.

2.6 Producto de estructuras

Definición 2.6.1 Sean $(\Omega, \mathfrak{A}, \mathfrak{F})$ y $(\Omega', \mathfrak{A}', \mathfrak{F}')$ dos estructuras estadísticas; se llama *producto de estructuras o estructura estadística producto a:*

$$(\Omega, \mathfrak{A}, \mathfrak{F}) \otimes (\Omega', \mathfrak{A}', \mathfrak{F}') = (\Omega \times \Omega', \mathfrak{A} \otimes \mathfrak{A}', \mathfrak{F} \otimes \mathfrak{F}'),$$

donde $\mathfrak{F} \otimes \mathfrak{F}' = \{P \otimes P' / P \in \mathfrak{F}, P' \in \mathfrak{F}'\}$.

Definición 2.6.2 Sean $(\Omega, \mathfrak{A}, \{P_\theta : \theta \in \Theta\})$ y $(\Omega', \mathfrak{A}', \{P'_\theta : \theta \in \Theta\})$ se llama *producto restringido a:*

$$(\Omega, \mathfrak{A}, \{P_\theta : \theta \in \Theta\}) \times (\Omega', \mathfrak{A}', \{P'_\theta : \theta \in \Theta\}) = (\Omega \times \Omega', \mathfrak{A} \otimes \mathfrak{A}', \{P_\theta \otimes P'_\theta : \theta \in \Theta\}).$$

2.6.1 Muestra empírica o muestra aleatoria

Es el producto restringido de n estructuras idénticas $(\Omega, \mathfrak{A}, \mathfrak{P})$ tal que:

$$(\Omega, \mathfrak{A}, \mathfrak{P})^n = (\Omega^n, \mathfrak{A}^n, \mathfrak{P}^n),$$

donde $\mathfrak{P}^n = \{P^n / P \in \mathfrak{P}\}$, con $P^n = \bigotimes_{i=1}^n P$.

Si se considera la variable aleatoria $X_i: (\Omega^n, \mathfrak{A}^n) \rightarrow (\Omega, \mathfrak{A})$ definida por $X_i(\omega) = \omega_i$ se verifica que los X_i son independientes.

Dentro del marco de producto de estructuras, la muestra es equivalente a una observación de n variables aleatorias independientes de igual ley.

Desde de un punto de vista práctico, la muestra es un modelo para el caso de n experiencias análogas independientes.

Ejemplo Si p es la probabilidad de que una pieza de un lote sea defectuosa tenemos que:

$$\Omega = \{B, M\}, \quad \mathfrak{A} = \mathcal{P}(\Omega) \quad \text{tal que} \quad P \in \mathfrak{P} \iff P(M) = p.$$

Si n piezas se fabrican independientemente unas de otras, entonces tenemos la muestra $(\{B, M\}^n, \mathcal{P}(\{B, M\}^n), \mathfrak{P}^n)$, con $P^n \in \mathfrak{P}^n \implies P^n(BMMB \dots) = p^b(1-p)^{n-b}$.

Sea X una variable aleatoria real = n° piezas defectuosas i.e. $X(BMMB \dots) = k$, la estructura indicada es $(\Omega' = \{0, 1, \dots, n\}, \mathcal{P}(\Omega'), \{B(n, p) : p \in [0, 1]\})$.

En el caso de estructuras dominadas, se escribe fácilmente las funciones de verosimilitud $L(\omega, \omega', \theta, \theta') = L(\omega, \theta) L(\omega', \theta')$.

En el caso restringido $L(\omega, \omega', \theta) = L(\omega, \theta) L(\omega', \theta)$. Así en el caso de una muestra empírica es cómodo de definir:

$$\ell(\omega, \theta) = \log L(\omega, \theta),$$

de donde

$$\ell(\omega^1, \dots, \omega^n, \theta) = \ell(\omega^1, \theta) + \dots + \ell(\omega^n, \theta).$$

Definición 2.6.3 Sea $(\Omega, \mathfrak{A}, \mathfrak{P})^n$ una muestra empírica, para todo $(\omega^1, \dots, \omega^n) \in \Omega^n$ se llama ley de

probabilidad empírica, la ley sobre (Ω, \mathfrak{A}) definida por:

$$\forall A \in \mathfrak{A}, \quad P^n(\omega^1, \dots, \omega^n, A) = \frac{1}{n} \sum_{i=1}^n I_A(\omega^i).$$

La ley fuerte de grandes números y otros teoremas de probabilidades demuestran que si P_0 es la ley de los elementos aleatorios independientes $\omega^1, \dots, \omega^n$, entonces $P^n \rightarrow P_0$, cuando $n \rightarrow \infty$, en varios sentidos. Este hecho es muy usado en estadística para estudiar nociones definidas sobre $(\Omega, \mathfrak{A}, \mathfrak{P})$, tales como momentos, leyes condicionales, etc. . . . Se considera este principio para definir la media, momentos, coeficientes de correlación, matrices de covarianza, ley marginal, etc. . . empíricos o muestrales.

2.7 Problemas de decisión

2.7.1 Pruebas de hipótesis

Definición 2.7.1 Sea $(\Omega, \mathfrak{A}, \mathfrak{P})$ una estructura estadística, se llama:

- **Hipótesis** Una parte no vacía de \mathfrak{P} .
- **Hipótesis simple** Hipótesis reducida a un sólo elemento.
- **Hipótesis compuesta** Hipótesis no simple.
- **Prueba** Estadística con valores en $([0, 1], \mathfrak{B}_{[0,1]})$.
- **Prueba de hipótesis** Se llama prueba de la hipótesis \mathfrak{P}_0 contra la hipótesis \mathfrak{P}_1 , con $\mathfrak{P}_0 \cap \mathfrak{P}_1 = \emptyset$, una prueba Φ tal que se rechaza \mathfrak{P}_0 (es decir \mathfrak{P}_0 falsa) con la probabilidad Φ .
- Se dice que la prueba Φ de \mathfrak{P}_0 contra \mathfrak{P}_1 es determinista, si Φ toma valores en $\{0, 1\}$. En este caso la regla de decisión es:
 - Rechazar \mathfrak{P}_0 (aceptar \mathfrak{P}_1) si $\Phi(\omega) = 1$.
 - Aceptar \mathfrak{P}_0 (rechazar \mathfrak{P}_1) si $\Phi(\omega) = 0$.
- La hipótesis \mathfrak{P}_0 se le llama hipótesis nula.
- Cuando Φ es determinista, se llama región crítica de prueba al conjunto $A = \Phi^{-1}(\{-1\})$.
- Se dice que la prueba Φ es estocástica si no es determinista.

Definición 2.7.2 Se llama potencia de la prueba Φ de la hipótesis \mathfrak{F}_0 contra la hipótesis \mathfrak{F}_1 , la restricción de la imagen de Φ a $\mathfrak{F}_0 \cup \mathfrak{F}_1$, es decir $\beta_\Phi(P) = \int \Phi dP$, con $P \in \mathfrak{F}_0 \cup \mathfrak{F}_1$. Algunos autores la definen como la restricción a \mathfrak{F}_1 .

Definición 2.7.3 Se llama nivel de la prueba de la hipótesis \mathfrak{F}_0 contra \mathfrak{F}_1 a:

- la restricción de la imagen de Φ a \mathfrak{F}_0 .
- $\sup_{P \in \mathfrak{F}_0} \beta_\Phi(P) = \alpha_\Phi$.

A grosso modo el nivel está relacionado con la probabilidad de rechazar \mathfrak{F}_0 , cuando \mathfrak{F}_0 es cierta.

Ejemplo Si Φ es determinista (es decir $\Phi = I_A$ con $A \in \mathfrak{A}$) se tiene $\beta_\Phi(P) = P(A)$, es decir $\beta_\Phi(P)$ es la probabilidad de A , cuando P recorre \mathfrak{F} .

Si Φ es una prueba de hipótesis de \mathfrak{F}_0 contra \mathfrak{F}_1 tenemos:

- rechazar \mathfrak{F}_0 si $\Phi(\omega) = 1 \iff \omega \in A$
- aceptar \mathfrak{F}_0 si $\Phi(\omega) = 0 \iff \omega \notin A$,

es decir $\beta_\Phi(P)$ es la probabilidad de rechazar \mathfrak{F}_0 , cuando P es la verdadera ley.

$1 - \beta_\Phi(P)$ es la probabilidad de aceptar \mathfrak{F}_0 , cuando la verdadera ley es P . Así tenemos:

- $P \in \mathfrak{F}_1$, $\beta_\Phi(P) = P(\text{rechazar } \mathfrak{F}_0, \text{ cuando } \mathfrak{F}_0 \text{ es falsa})$.
- $P \in \mathfrak{F}_0$, $\beta_\Phi(P) = P(\text{rechazar } \mathfrak{F}_0, \text{ cuando } \mathfrak{F}_0 \text{ es cierta})$.
- $\beta_\Phi(P)$ es una probabilidad de error (llamado error de primera especie) para $P \in \mathfrak{F}_0$, es decir es la probabilidad de rechazar \mathfrak{F}_0 sabiendo que \mathfrak{F}_0 es verdadera.

$$P(\text{rechazar } \mathfrak{F}_0 / \mathfrak{F}_0 \text{ es cierta}).$$

- $1 - \beta_\Phi(P)$ es una probabilidad de error (llamado error de segunda especie), cuando $P \in \mathfrak{F}_1$, es decir es la probabilidad de rechazar \mathfrak{F}_1 sabiendo que \mathfrak{F}_1 es verdadera.

$$P(\text{rechazar } \mathfrak{F}_1 / \mathfrak{F}_1 \text{ es verdadera}).$$

En la práctica para que Φ sea una *buena* prueba, es necesario que $\beta_\Phi(P)$ sea *pequeño* para $P \in \mathfrak{P}_0$ y *grande* para $P \in \mathfrak{P}_1$.

Ejemplo Una máquina fabrica piezas con una probabilidad p de tener una pieza no defectuosa y se define $\Omega = \{B, M\}^5$:

$$\mathfrak{P}_0 \iff p = 0.9$$

$$\mathfrak{P}_1 \iff p < 0.9.$$

Se escoge Φ tal que: $\Phi(BBBBB) = 0$ se acepta \mathfrak{P}_0

$$\Phi(\omega) = 1 \quad \text{si no}$$

i.e. se rechaza \mathfrak{P}_0 desde que hay una pieza defectuosa.

Ejercicio

- Buscar el nivel de la prueba y graficar $\beta_\Phi(P)$.
- Considerar $\{N(\mu_0, 1)\}$ contra $\{N(\mu, 1)/\mu > \mu_0\}$, $\alpha = 0.05$. Dar la región crítica y determinar la potencia.

2.7.2 Estimación

Se considera $(\Omega, \mathfrak{A}, \{P_\theta : \theta \in \Theta\})$ y suponemos que (Θ, \mathfrak{T}) es un espacio probabilisable.

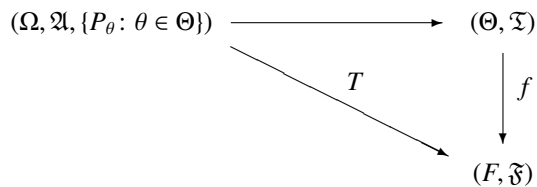
Definición 2.7.4 Estimación de θ Un estimador de θ es una estadística con valores en (Θ, \mathfrak{T}) , es decir una estadística $T : \Omega \rightarrow \Theta$

$$\omega \mapsto T(\omega).$$

En la práctica se busca que T esté próximo del verdadero valor θ (desconocido).

En forma general se considera un espacio probabilisable (F, \mathfrak{F}) y una función f medible de Θ en F .

Se denomina estimador de $f(\theta)$ a una estadística con valores en (F, \mathfrak{F}) .



Si $(F, \mathfrak{F}) = (\mathbb{R}^n, \mathfrak{B}_{\mathbb{R}^n})$, entonces diremos que T es un estimador sin sesgo de $f(\theta)$ si:

- T es P_θ -integrable, $\forall \theta \in \Theta$.
- $E_\theta(T) = f(\theta)$.

Sea $(\Omega, \mathfrak{A}, \{P_\theta: \theta \in \Theta\})^n$ una muestra empírica de tamaño n y sea $(F, \mathfrak{F}) = (\mathbb{R}, \mathfrak{B}_{\mathbb{R}})$, la estadística T es un estimador de $f(\theta)$ convergiendo en probabilidad, m.c. o c.s. si $T \rightarrow f(\theta)$, cuando $n \rightarrow \infty$, en probabilidad, m.c. o c.s.

Si $T \xrightarrow{Pr} f(\theta)$ se dice que T es consistente.

Ejemplo Se considera $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}, \mathfrak{P}})^n$, es decir n pruebas o experiencias independientes de igual ley de probabilidad $P \in \mathfrak{P}$.

Si se toma $X_i: \Omega^n \rightarrow \Omega$, entonces las variables aleatorias X_i son independientes de igual ley $P \in \mathfrak{P}$.

$$\omega \mapsto \omega_i$$

Se supone que $E_P(X_i)$ y $E_P(X_i^2)$ existen y que $E(X_i) = \mu_P, i = 1, \dots, n$. Se desea estimar μ_P .

Si T es la media de los valores observados, entonces:

$$T(\omega) = \frac{\omega_1 + \dots + \omega_n}{n},$$

es decir $T = \frac{1}{n} \sum_{i=1}^n X_i$ y se tiene que $E_P(T) = \mu_P$, es decir T es un estimador sin sesgo de μ_P y convergente (en probabilidad, m.c. y c.s.)

Ejemplo Sea la estructura estadística $(\{0, 1\}, \mathcal{P}(\{0, 1\}), \{P_p: p \in [0, 1]\})^n$, donde P_p es la probabilidad sobre $\{0, 1\}$ tal que $P_p(\{1\}) = p$ y $P_p(\{0\}) = 1 - p$.

Sea $X(\omega_1, \dots, \omega_n) = \omega_1 + \dots + \omega_n$ igual al número de veces que aparece 1 en la sucesión $(\omega_1, \dots, \omega_n)$.

Así tenemos que $T_n = \frac{X}{n}$ es un estimador sin sesgo y convergente a p .

Ejemplo Si tomamos $F = \mathbb{R}$, nos interesa cuantificar la distancia *media* de $T(\omega)$ y $f(\theta)$ considerando por ejemplo $E_\theta((T - f(\theta))^2)$.

Si T es cuadrado-integrable, esta cantidad es la varianza de T , si T es un estimador sin sesgo de $f(\theta)$.

Así T es un mejor estimador en cuanto a $E_\theta((T - f(\theta))^2)$ es pequeño; mejor en el sentido de más error

cuadrático medio. Si tomamos el primer ejemplo:

$$E_p(\bar{X}_n) = \mu_p, \quad \text{var}(\bar{X}_n) = \sigma_p^2/n.$$

Consideremos cualquier combinación convexa de los X_i , es decir:

$$T_n = \frac{1}{n} \sum_{i=1}^n \alpha_i X_i, \quad \alpha_i \geq 0, \quad \sum_{i=1}^n \alpha_i = nE(T_n) = \mu_p, \quad \text{var}(T_n) = \frac{\sigma^2}{n} \sum_{i=1}^n \alpha_i^2,$$

pero $\text{var}(\bar{X}_n) = \frac{\sigma^2}{n} \leq \frac{1}{n} \sigma^2 \sum_{i=1}^n \alpha_i^2$, por la desigualdad de Schwarz², por lo que para toda escogencia de los α_i , el estimador T_n no puede ser estrictamente mejor que \bar{X}_n , en el sentido del error cuadrático medio, para estimar sin sesgo a μ .

Cabe preguntarse por el mejor estimador sin sesgo de μ . Este es un problema general de la estadística matemática; que es el estudio de la existencia de un estimador óptimo en un cierto sentido, por ejemplo en el sentido del error cuadrático medio.

2.8 Noción de intervalo de confianza

Consideremos la estructura estadística $(\Omega, \mathfrak{A}, \{P_\theta: \theta \in \mathbb{R}\})^n$. Se desea encontrar un intervalo de \mathbb{R} (intervalo aleatorio) que contenga a θ , con una cierta probabilidad dada. Así pues, en la práctica es necesario encontrar dos estadísticas T_1 y T_2 , con $T_1 \leq T_2$ c.p.d. tales que:

$$P(T_1 \leq \theta \leq T_2) = 1 - \alpha = \gamma,$$

donde $0 \leq \alpha \leq 1$. Al valor γ se le llama coeficiente de seguridad del intervalo de confianza para θ .

²**Hermann Amandus Schwarz (1843-1921)** Nace en Hermsdorf, Silesia (hoy Polonia) y muere en Berlín, Alemania. Schwarz comenzó en Berlín sus estudios en química, pero fue influenciado por Kummer y Weierstrass y se interesa en la geometría. Continuó estudiando en Berlín supervisado por Weierstrass hasta su doctorado (en 1864) sobre superficies mínimas (superficies de menor área), un problema típico del cálculo de variaciones, que estableció un puente entre la teoría de superficies mínimas y la teoría de funciones analíticas. En 1869 fue designado como profesor de matemáticas en el Eidgenössische Technische Hochschule en Zurich y en 1875, aceptó un puesto de Matemáticas en la Universidad de Göttingen. Schwarz dio un método alternativo para resolver el problema de Dirichlet que pronto llegó a ser una técnica usual. Schwarz contestó la pregunta de si una superficie mínima dada rinde realmente un área mínima. Una idea de este trabajo, en que él construyó una función que usa aproximaciones sucesivas, Emile Picard la usó para la prueba de la existencia para soluciones de ecuaciones diferenciales. Contiene también la desigualdad para la integral ahora conocido como "la desigualdad de Cauchy-Schwarz o de Cauchy-Bunyakovsky-Schwarz". Creó métodos generales basado en su magnífica intuición geométrica. La desigualdad de Cauchy-Schwarz aparece en aritmética, geometría y en las formulaciones teóricas del trabajo de matemáticos tal como Bunyakovsky, Cauchy, Grassmann, von Neumann y Weyl.

Ejemplo Consideremos la estructura estadística $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}}, \{N(\mu, 1)/\mu \in \mathbb{R}\})^n$ y deseamos un intervalo de confianza para μ , con un coeficiente de seguridad $100(1 - \alpha)\%$.

Sabemos que $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ estima a μ y converge c.s. (i.e. en ley y probabilidad). Además, $\bar{X} - \mu \sim N\left(0, \frac{1}{n}\right)$, entonces $P\left(\left|\frac{\bar{X} - \mu}{1/\sqrt{n}}\right| < t_\alpha\right) = 1 - \alpha$, donde t_α depende de α y debe determinarse en una tabla de distribución normal para un valor α dado. Notemos que:

$$P\left(\bar{X} - \frac{t_\alpha}{\sqrt{n}} < \mu < \bar{X} + \frac{t_\alpha}{\sqrt{n}}\right) = 1 - \alpha.$$

Si $\alpha = 0.05$, entonces $t_\alpha = 1.96$ i.e.

$$P\left(\bar{X} - \frac{1.96}{\sqrt{n}} < \mu < \bar{X} + \frac{1.96}{\sqrt{n}}\right) = 0.95$$

i.e. el intervalo de confianza es $\bar{X} \pm \frac{1.96}{\sqrt{n}}$ para μ , con un coeficiente de seguridad del 95%.

- El caso $\theta \in \Theta \subset \mathbb{R}^n$ se discutirá más adelante.
- α es la probabilidad de μ de caer fuera del intervalo.
- Cuando α decrece, la longitud del intervalo crece.

2.9 Noción de suficiencia estadística

Consideremos una urna con bolas rojas y negras. Se realizan n extracciones (con reposición), habiendo un porcentaje p de bolas rojas desconocido $p \in]0, 1[$.

El resultado de la prueba se presenta como una sucesión de n elementos. El problema que se presenta es el siguiente: Sabemos que la sucesión es útil en su conjunto para dar información sobre p . Cabe preguntarse si se puede extraer datos más concisos, resumidos, dando una información del mismo valor, conteniendo toda la información sobre p .

Si se observa la prueba $\omega = (N, R, R, N, \dots, R)$, la información está contenida dentro de ω . Consideremos $X(\omega) = n^\circ$ de bolas rojas en ω , entonces se tiene que $P(\omega) = p^k(1 - p)^{n-k}$, donde k es el número de bolas rojas en el total de las n extracciones.

Por otro lado, $P(X = k) = \binom{n}{k} p^k(1 - p)^{n-k}$ y si consideramos $P(\omega/X = k) = 1/\binom{n}{k}$, se observa que si se tiene $X = k$, el emplazamiento de los R en la sucesión de resultados no da información

adicional sobre p en la medida en que la probabilidad de los diversos órdenes es *independiente* de p , es decir X contiene toda la información sobre p y diremos que X es una *estadística suficiente*. Vamos a formalizar las ideas precedentes y pasar al modelo matemático general.

Definición 2.9.1 Sea $(\Omega, \mathfrak{A}, \mathfrak{P})$ una estructura estadística y sea \mathfrak{B} una sub-tribu de \mathfrak{A} , entonces diremos que \mathfrak{B} es una tribu suficiente sii $\forall A \in \mathfrak{A}$, existe una determinación $P^{\mathfrak{B}}(A)$ que sea la misma para todo $P \in \mathfrak{P}$.

Definición 2.9.2 Sea $(\Omega, \mathfrak{A}, \mathfrak{P})$ una estructura estadística y sea T una estadística con valores en $(\mathcal{T}, \mathcal{C})$, entonces T es una estadística suficiente sii la tribu $T^{-1}(\mathcal{C})$ es suficiente (i.e. $\forall A \in \mathfrak{A}$ existe una determinación de $P^T(A)$ común a todas las leyes $P \in \mathfrak{P}$).

Nota Toda estadística equivalente a una estadística suficiente, es suficiente. Por ejemplo $X = n^\circ$ bolas rojas y $Y = n^\circ$ bolas negras son equivalentes y suficientes.

Teorema 2.9.1 Sea $(\mathbb{R}^n, \mathfrak{B}_{\mathbb{R}^n}, \mathfrak{P})$ una estructura estadística y sea T una estadística con valores en \mathbb{R}^k , entonces T es suficiente sii existe una versión regular de la ley de probabilidad sobre $\mathfrak{B}_{\mathbb{R}^n}$ condicionada a T , que sea la misma para todo $P \in \mathfrak{P}$.

Prueba La prueba es inmediata por el teorema de Jifina.

2.9.1 Teorema de factorización

Tal como se ha definido la suficiencia estadística es difícil de manejar. Así vamos a dar un criterio simple para poder reconocer las estadísticas suficientes.

Teorema 2.9.2 Sea $(\Omega, \mathfrak{A}, \{P_\theta: \theta \in \Theta\})$ una estructura estadística dominada por una ley de probabilidad dominante privilegiada P^* . Sea \mathfrak{B} sub-tribu de \mathfrak{A} , entonces \mathfrak{B} es suficiente sii $\forall \theta \in \Theta$ existe una determinación f_θ de la densidad de P_θ , con respecto a P^* , que sea \mathfrak{B} -medible. Se puede entonces escoger por determinación común de las probabilidades condicionadas a \mathfrak{B} de \mathfrak{A} , una determinación de $P^{*\mathfrak{B}}(A)$.

Teorema 2.9.3 Sea $(\Omega, \mathfrak{A}, \{P_\theta: \theta \in \Theta\})$ una estructura estadística dominada por una medida μ , σ -finita. Sea $L(\cdot, \theta)$ densidad de P_θ , con respecto a μ y sea T estadística con valores en $(\mathcal{T}, \mathcal{C})$, entonces

T es una estadística suficiente sii existe

- una función numérica h definida sobre Ω , positiva, \mathfrak{A} -medible.
- $\forall \theta \in \Theta$ existen funciones numéricas g_θ definidas sobre \mathfrak{Z} , \mathfrak{C} -medibles tales que:

$$\forall \theta \in \Theta, L(\cdot, \theta) = (g_\theta \circ T)h \text{ } \mu\text{-c.p.d.}$$

$$\text{i.e. } \forall \theta \in \Theta, L(\omega, \theta) = g_\theta(T(\omega))h(\omega) \text{ } \mu\text{-c.p.d.}$$

Nota

- $h: \Omega \rightarrow \mathbb{R}^+$ no depende de θ .
- h, g_θ no son únicas.
- El teorema es conocido como el teorema de factorización de Neyman–Fisher. Fisher en 1922 probó la condición suficiente y Neyman³ en 1935 la condición necesaria.

Demostración del Teorema 2.9.2:

(\implies) Supongamos que \mathfrak{B} es suficiente, por definición existe una determinación de la probabilidad condicional $P_\theta^{\mathfrak{B}}(A)$ de A a \mathfrak{B} que no depende de θ . Denotémosla $P^{\mathfrak{B}}(A)$. Se tiene entonces que:

- $P^{\mathfrak{B}}(A)$ función numérica \mathfrak{B} -medible,
- $\forall A \in \mathfrak{A}, \forall \theta \in \Theta, \forall B \in \mathfrak{B}, \int_B P^{\mathfrak{B}}(A) dP_{\theta\mathfrak{B}} = P_\theta(A \cap B),$ (1)

donde $P_{\theta B}$ es la restricción de P_θ a \mathfrak{B} .

Tenemos así en particular, tomando $B = \Omega$:

$$\forall A \in \mathfrak{A}, \forall \theta \in \Theta, P_\theta(A) = \int P^{\mathfrak{B}}(A) dP_{\theta\mathfrak{B}}. \quad (2)$$

³**Jerzy Neyman (1894-1981)** Nace en Rusia en 1894 y realiza sus estudios de física y matemática en la Universidad de Kharkov, En 1921, regresa a Polonia, país de origen de su familia, donde trabaja como estadístico en el Instituto Nacional de Agricultura de Bydgoszcz.

En 1924 realiza una estadía en Londres, donde estudia en el University College bajo la dirección de K. Pearson y tuvo la oportunidad de trabajar con E.S. Pearson, W.S. Gosset y R.A. Fisher.

En 1937, viaja a Estados Unidos a realizar una serie de conferencias y acepta un puesto de profesor en la Universidad de Berkeley, donde crea el Departamento de Estadística y termina su brillante carrera.

Es uno de los fundadores de la estadística moderna y junto con E.S. Pearson, da las bases de la Teoría de Pruebas de Hipótesis.

Por otro lado existe P^* verificando:

$$P^* = \sum_{i \in I} c_i P_{\theta_i}, \quad I \text{ numerable, } c_i > 0, \quad \sum_{i \in I} c_i = 1.$$

Por lo tanto $P_\theta \ll P^*$ implica que $P_{\theta \mathfrak{B}} \ll P_{\mathfrak{B}}^*$, donde $P_{\mathfrak{B}}^*$ designa la restricción de P^* a \mathfrak{B} . Por el teorema de Radon⁴-Nikodým⁵ existe una densidad f_θ de $P_{\theta \mathfrak{B}}$ con respecto a $P_{\mathfrak{B}}^*$, i.e. $f_\theta = \frac{dP_{\theta \mathfrak{B}}}{dP_{\mathfrak{B}}^*}$,

⁴**Johann Karl Radon (1887-1956)** Nace en Bohemia (hoy Decin, República Checa) el 16 de diciembre de en 1887 y muere el 25 mayo de 1956 en Viena, Austria. Va a la escuela en Leitmeritz (hoy Litomerice) en Bohemia entre 1897 y 1905. Entra a la Universidad de Viena y obtiene el doctorado in 1910 sobre cálculo de variaciones. En 1911 que termina en Göttingen y trabaja como profesor en la Universidad de Brünn (hoy Brno) por un año y se va al Technische Hochschule en Viena.

En 1919 Radon trabaja como profesor en Hamburgo y en Greifswald en 1922. Enseñó en Erlangen en 1925 y desde 1928 a 1945 trabajó en la Universidad de Breslau. Se estableció en la Universidad de Viena en 1947 donde se queda hasta su muerte. Radon aplicó el cálculo de variaciones a la geometría diferencial, lo que llevó a aplicaciones en teoría de números. Aportó una generalización de la teoría de la medida a los espacios topológicos localmente compactos. La medida de Radon, extiende la teoría de la integración de este último a las funciones continuas en un compacto (funciones y partes medibles). Sus trabajos coronan la teoría de integración de Lebesgue.

⁵**Otton Marcin Nikodým (1887-1974)** Nace el 13 de agosto de 1887 en Zablutow, Galicia, Austria-Hungría (hoy Ucrania) y muere el 4 de mayo de 1974 en Utica, EE.UU. Otton Nikodým (escrito a veces Otto Martin) nació en Zablutow cerca de Kolomyja (en la provincia de Stanislawow). Su padre Otton Boguslaw murió en un accidente cuando tenía un año y pronto su madre Marianna Cyprian también murió. Los abuelos Marcin Cyprian (del origen italiano) y Amalia Bott de Pirot (del origen francés) lo criaron. Después de que una estancia corta en Viena, la familia volvió a Lvov en 1897, donde terminó la secundaria. Se preparó para los exámenes en latín y griego y obtuvo un diploma clásico de secundaria. Estudió matemática en la Universidad Jan Kazimierz en Lvov bajo la dirección de W. Sierpinski, J. Puzyna y M. Smoluchowski. En 1911 se graduó en matemática y también recibió una licencia para enseñar matemática y física en secundaria. Inmediatamente después comenzó su trabajo en la secundaria en Cracovia, donde permanece hasta 1924. En 1915, E. Tarnawski era su estudiante de cuarto año de secundaria y en una carta a W. Orlicz escribió: "Otto Nikodým queda en mi memoria como personalidad algo rara... Le veo como delgado y pelo oscuro con barba. Aunque tenía cerca de 30 años, era diferente de otras personas de su edad; casi impersonal como si no cambiara con la edad,... estéticamente neutro y distante..., físicamente débil, no levantaba la voz, desapasionado siempre y sin embargo audible... Sus conferencias eran interesantes por su contenido... Presentó la ciencia como era, sin incorporar su propia personalidad que desaparecía de la vista."

Otro pupilo en 1922-1924 M. Miesowicz escribió: "Él podía evocar el aprecio, la admiración y el entusiasmo de los estudiantes, por su precisión y elegancia de expresar las leyes físicas en una forma matemática rigurosa."

El 2 de abril de 1919, se fundó a la Sociedad Matemática Polaca con dieciséis matemáticos, entre ellos Otton Nikodým. En 1924, bajo fuerte presión de Sierpinski, Nikodým tomó su examen doctoral en la Universidad de Varsovia. Parece que no se cuidó mucho para la tesis y su respuesta a Sierpinski fue: ¿Voy a ser más sabio debido a esto?

Después que se trasladó a la Universidad de Jagiellonian en Cracovia y comenzó a publicar en 1925. Nikodým (y su esposa Stanislawna) pasó el año académico 1926-27 en la Sorbonne y volvieron a Polonia, en junio de 1927, donde fue habilitado para enseñar en la Universidad de Varsovia. En el período 1930-1945, Nikodým vivió en Varsovia y hasta el inicio de la II guerra mundial, ambos enseñaron en la Universidad. Durante este período Nikodým publicó 32 artículos y cuatro libros de texto. El nombre de Nikodým se conoce sobre todo en la Teoría de la medida (teorema de la derivada de Radon–Nikodým, el teorema de la convergencia de Nikodým, el teorema de Nikodým–Grothendieck), en el análisis funcional (propiedad de Radon–Nikodým de un espacio de Banach, el espacio métrico de Frechet–Nikodým, el conjunto Nikodým), proyección sobre conjuntos convexos con aplicaciones al problema de Dirichlet, las soluciones generalizadas de ecuaciones diferenciales, teoría

positiva y \mathfrak{B} -medible. Por (2) se tiene:

$$\forall \theta \in \Theta, P_\theta(A) = \int P^{\mathfrak{B}}(A) f_\theta dP_{\mathfrak{B}}^* \quad (3)$$

Se va a demostrar que f_θ es también densidad de P_θ con respecto a P^* (i.e. una determinación).

Sea X una v.a.r. \mathfrak{B} -medible, positiva, P_{θ_i} -integrable para todo $i \in I$ y sea $\phi_i = \frac{dP_{\theta_i, \mathfrak{B}}}{dP_{\mathfrak{B}}^*}$ una determinación de la densidad de $P_{\theta_i, \mathfrak{B}}$, con respecto a $P_{\mathfrak{B}}^*$.

de conjuntos descriptiva y los fundamentos de la mecánica cuántica.

Nikodým demostró en 1927 cómo producir un subconjunto N , en el cuadrado unitario con $\text{área}(N) = 1$, tal que para cada punto $(x, y) \in N$, hay una línea que interseca N en un solo punto (x, y) . Este sistema paradójico en el plano, que para ciertos problemas juega un papel similar a al conjunto de Besicovitch, se llama conjunto de Nikodým.

En 1945 Nikodým fue profesor en la Universidad Técnica de Cracovia y en año académico 1945-46 enseñó matemática. En 1946 Nikodým y su esposa Stanislawa fueron para Bélgica y Francia en donde él comenzó su trabajo sobre los fundamentos matemáticos de la mecánica de cuántica. A partir la 1948 a 1965 trabajó en los EE.UU. en el Kenyon College, en Gambier, Ohio, donde se pensionó. Después que se retira en 1966, Nikodým va a Utica, Nueva York, donde continúa su investigación, patrocinada en parte por la Atomic Energy Commission and National Science Foundation. Después de 1947, escribió cerca de 50 artículos de investigación. Dio conferencias en Italia, Bélgica, Francia, Inglaterra, Alemania, Rumania, Canadá y en universidades en los EE.UU. En 1965 la Universidad de Nápoles, Italia, lo invitó por un semestre para que diera un curso de Teoría de la medida.

Nikodým escribió tres volúmenes de "Didactics of pure mathematics in high school," pero publicó solamente los dos primeros volúmenes en 1930 y 1938. La pobre respuesta de los profesores a sus métodos de mejorar la enseñanza, lo desalentó y quemó el manuscrito del tercer volumen. Presentó conferencias populares en radio por ejemplo: Lógica e intuición en la Ciencia, Sobre el infinito, Sobre paradojas en lógica, ¿Qué bueno es el álgebra?, Diversas clases de espacios, El misterio de la gravitación, Sobre la importancia de la teoría y éstas fueron publicadas en 1946 como un libro "Let's look deeply inside the mind" (Spojrzmy w glebiny mysli). Algunos de sus otros libros fueron: Introducción al cálculo diferencial, (Varsovia, 1936) común con su esposa, Teoría de tensores con aplicaciones a la geometría y a la física matemática I, (Varsovia, 1938), Ecuaciones diferenciales, (Poznán, 1949). Tres de sus otros libros fueron: el segundo volumen de Teoría de tensores y dos volúmenes de Mecánica, desaparecidos durante la II Guerra Mundial. Cuando Nikodým oyó que el trabajo de muchos años se perdió, dijo solamente: En ese caso no tendré que corregir las pruebas. Ante la cuestión de que si deseaba escribir estos libros otra vez, contestó: "Hay tantos nuevos problemas que no puedo pasar más tiempo en los que ya he acabado." Su último libro "The Mathematical Apparatus for Quantum-Theories", basado en la "Theory of Boolean Lattices" publicado en 1966 by Springer-Verlag, contiene casi mil páginas, sobre el formalismo matemático para la mecánica cuántica, o más exactamente un estudio detallado de los sub-álgebras booleanas de la lógica de subespacios cerrados de un espacio complejo de Hilbert.

Nikodým gustaba de la buena literatura y cuentos de hadas, pero sobre todo la música. Podía tocar por horas el piano. Conocía bien (aparte del polaco) inglés, francés, alemán e italiano. Enseñó en todos estos idiomas. The National Science Foundation en los EE.UU. le pidió levantar el nivel de las matemáticas en los EE.UU. y contestó: "Debemos invitar a los conferenciantes entusiastas puesto que el entusiasmo es contagioso. Él mismo fue un entusiasta hasta el final de su vida. En 1971 recibió una descarga eléctrica y por dos años y diez meses no recuperó el sentido. Él murió el 4 de mayo de 1974 y lo enterraron en el "Cemetery for the meritorious" en Doylestown, Pennsylvania en la Capilla de nuestra señora de Czestochowa. Después de su muerte Nelson Dunford escribió en una carta a la esposa de Nikodým: "Otton era y será siempre, un gran hombre. Estoy feliz de haberlo conocido. Sus descubrimientos fueron muy profundos y vivirán para siempre tanto como el teorema de Pitágoras, que ha sobrevivido por siglos."

Se tiene que $\forall B \in \mathfrak{B}$:

$$\begin{aligned} P_{\mathfrak{B}}^*(B) &= \sum_{i \in I} c_i P_{\theta_i}(B) = \sup_{n \in I} \sum_{i=1}^n c_i P_{\theta_i}(B) = \\ &= \sup_{n \in I} \sum_{i=1}^n \int_B c_i \phi_i dP_{\mathfrak{B}}^* = \int_B (\sup_{n \in I} \sum_{i=1}^n c_i \phi_i) dP_{\mathfrak{B}}^*, \end{aligned}$$

de donde $\sum_{i \in I} c_i \phi_i = 1$, $P_{\mathfrak{B}}^*$ -c.p.d. Así tenemos que:

$$\begin{aligned} \sum_{i \in I} c_i \int X dP_{\theta_i \mathfrak{B}} &= \sup_{n \in I} \sum_{i=1}^n \int c_i X \phi_i dP_{\mathfrak{B}}^* = \sup_{n \in I} \int X \left(\sum_{i=1}^n c_i \phi_i \right) dP_{\mathfrak{B}}^* = \\ &= \int X (\sup_{n \in I} \sum_{i=1}^n c_i \phi_i) dP_{\mathfrak{B}}^* = \int X dP_{\mathfrak{B}}^*. \end{aligned}$$

Denotemos $P'(B) = P^*(A \cap B)$. Utilizando (1) y que si $X = I_B P^{\mathfrak{B}}(A)$ se deduce: $\forall A \in \mathfrak{A}, \forall \theta \in \Theta, \forall B \in \mathfrak{B}$:

$$P'(B) = P^*(A \cap B) = \sum_{i \in I} c_i P_{\theta_i}(A \cap B) = \sum_{i \in I} c_i \int_B P^{\mathfrak{B}}(A) dP_{\theta_i \mathfrak{B}} = \int_B P^{\mathfrak{B}}(A) dP_{\mathfrak{B}}^*.$$

$P^{\mathfrak{B}}(A)$ es una determinación de la densidad de P' con respecto a $P_{\mathfrak{B}}^*$. Por (3) tenemos:

$$\forall A \in \mathfrak{A}, \forall \theta \in \Theta, P_{\theta}(A) = \int f_{\theta} dP' \quad (4)$$

Consideremos una variable aleatoria real X \mathfrak{B} -medible, positiva, integrable tal que X es límite de una sucesión creciente de funciones escalonadas sobre \mathfrak{B} :

$$X = \sup_n \sum_{j \in J_n} b_{n_j} I_{B_{n_j}}, \quad J_n \text{ finito}, B_{n_j} \in \mathfrak{B}, b_{n_j} \geq 0.$$

Se tiene así:

$$\begin{aligned} \int X dP' &= \int \sup_n (\sum_j b_{n_j} I_{B_{n_j}}) dP' = \sup_n \int \left(\sum_j b_{n_j} I_{B_{n_j}} \right) dP' \\ &= \sup_n \sum_j b_{n_j} P'(B_{n_j}) = \sup_n \sum_j b_{n_j} P^*(A \cap B_{n_j}) \\ &= \sup_n \int_A \sum_j b_{n_j} I_{B_{n_j}} dP^* = \int_A X dP^* \end{aligned}$$

Utilizando $X = f_{\theta}$ se tiene en (4), $\forall A \in \mathfrak{A}, \forall \theta \in \Theta$:

$$P_{\theta}(A) = \int_A f_{\theta} dP^*$$

i.e. $\forall \theta \in \Theta, f_{\theta}$ es una determinación de \mathfrak{B} -medible por construcción de $\frac{dP_{\theta}}{dP^*}$.

(\Leftarrow) Recíprocamente se supone que f_θ es una determinación de P_θ con respecto a P^* , que es \mathfrak{B} -medible $\forall \theta \in \Theta$. Vamos a demostrar que $P^{*\mathfrak{B}}(A)$ puede escogerse para determinación de $P_\theta^{\mathfrak{B}}(A)$, $\forall \theta \in \Theta$, lo que prueba la suficiencia de la tribu \mathfrak{B} . En efecto, $\forall A \in \mathfrak{A}$, $\forall \theta \in \Theta$, $\forall B \in \mathfrak{B}$:

$$\begin{aligned} P_\theta(A \cap B) &= \int_{A \cap B} f_\theta dP^* = \int I_A I_B f_\theta dP^* = E_{P^*}(I_A I_B f_\theta) \\ &= E_{P^*}(E_{P^*}^{\mathfrak{B}}(I_A I_B f_\theta)) = E_{P^*}(I_B f_\theta E_{P^*}^{\mathfrak{B}}(I_A)) = \int_B f_\theta P^{*\mathfrak{B}}(A) dP^* \\ &= \int_B P^{*\mathfrak{B}}(A) dP_\theta = P_\theta(A \cap B), \end{aligned}$$

pero $P^{*\mathfrak{B}}(A)$ es \mathfrak{B} -medible y se puede escoger a $P^{*\mathfrak{B}}(A)$ como una determinación de $P_\theta^{\mathfrak{B}}(A)$, $\forall A \in \mathfrak{A}$ y $\forall \theta \in \Theta$.

Demostración del teorema 2.9.3 Sea P^* una ley de probabilidad dominante privilegiada; se tiene que $\forall \theta \in \Theta$, $P_\theta \ll P^*$ y como $P^* \ll \mu$, entonces:

$$L(\cdot, \theta) = \frac{dP_\theta}{d\mu} = \frac{dP_\theta}{dP^*} \frac{dP^*}{d\mu} = f_\theta h.$$

T es suficiente sii $T^{-1}(\mathfrak{C})$ es suficiente sii se puede escoger f_θ , $T^{-1}(\mathfrak{C})$ -medible $\forall \theta \in \Theta$, como una determinación de la densidad de P_θ con respecto a P^* sii $\forall \theta \in \Theta$ existe una función numérica g_θ , \mathfrak{C} -medible definida sobre $(\mathfrak{X}, \mathfrak{C})$ tal que $f_\theta = g_\theta \circ T$.

Así T es suficiente es equivalente a la existencia de h y g_θ teniendo las propiedades indicadas en el teorema i.e.:

$$L(\cdot, \theta) = (g_\theta \circ T)h \quad \mu\text{-c.p.d.}$$

2.9.2 Aplicaciones al reconocimiento de estadísticas suficientes

1. Consideremos la estructura estadística $(\Omega = \{0, 1\}^n, \mathcal{P}(\Omega), \mathfrak{P})$, donde \mathfrak{P} es la familia de probabilidades P_θ sobre $\{0, 1\}^n$ tales que:

$$P_\theta(\{\omega_1, \dots, \omega_n\}) = \theta^{\sum_{i=1}^n \omega_i} (1 - \theta)^{n - \sum_{i=1}^n \omega_i}, \quad \forall \omega \in \Omega.$$

P_θ está dominada por la medida de conteo μ . Así tenemos que:

$$L(\omega, \theta) = \theta^{\sum_{i=1}^n \omega_i} (1 - \theta)^{n - \sum_{i=1}^n \omega_i}$$

i.e. si definimos $T(\omega) = \sum_{i=1}^n \omega_i$, tenemos que T es una estadística suficiente. En efecto, si definimos $g_\theta(t) = \theta^t(1 - \theta)^{n-t}$ y $h(\omega) = 1$, entonces $L(\omega, \theta) = g_\theta(T(\omega))h(\omega)$ i.e. T es suficiente.

2. Sea la muestra empírica $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}, \{N(\mu, \sigma^2) / \mu \in \mathbb{R}, \sigma > 0\}})^n$. La estructura estadística está dominada por la medida de Borel en $(\mathbb{R}^n, \mathfrak{B}_{\mathbb{R}^n})$ i.e. la densidad se escribe:

$$L(x, \mu, \sigma) = \frac{1}{\sigma^n (\sqrt{2\pi})^n} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right\},$$

con $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$, $\mu \in \mathbb{R}$, $\sigma > 0$, pero:

$$\sum_{i=1}^n (x_i - \mu)^2 = \sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2,$$

con $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$, de donde:

$$L(\mathbf{x}, \mu, \sigma) = \frac{1}{\sigma^n (\sqrt{2\pi})^n} \exp \left\{ -\frac{1}{2\sigma^2} \left(\sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2 \right) \right\}.$$

Sea la estadística $T = (\bar{X}, S^2)$ con:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

i.e. $T: \mathbb{R}^n \rightarrow \mathbb{R}^2$. Así tenemos que:

$$L(\mathbf{x}, \mu, \sigma) = g_{(\mu, \sigma)}(T(\mathbf{x}))h(\mathbf{x}),$$

con $h(\mathbf{x}) = \frac{1}{(\sqrt{2\pi})^n}$ y $g_{(\mu, \sigma)}(u, v) = \frac{1}{\sigma^n} \exp \left\{ -\frac{1}{2}((n-1)v + n(u - \mu)^2)/\sigma^2 \right\}$ i.e. T es una estadística suficiente.

3. Consideremos una muestra empírica $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}, \{U(0, \theta) : \theta > 0\}})^n$, con $U(0, \theta)$ ley uniforme en $[0, \theta]$, la cual está dominada por la medida de Borel:

$$L(\mathbf{x}, \theta) = \begin{cases} \frac{1}{\theta^n} & \text{si } \mathbf{x} \in [0, \theta]^n \\ 0 & \text{si no} \end{cases}$$

Sea la estadística con valores en \mathbb{R} , (i.e. $T: \mathbb{R}^n \rightarrow \mathbb{R}$) tal que $T(\mathbf{x}) = X_{(n)}(\mathbf{x}) = \sup x_i$. Demostremos que T es suficiente.

En efecto, sea $g_\theta(u) = \begin{cases} \frac{1}{\theta^n} & \text{si } u \leq \theta \\ 0 & \text{si } u > \theta \end{cases}$ y sea $h(\mathbf{x}) = \begin{cases} 1 & \text{si } x_i > 0, i = 1, \dots, n \\ 0 & \text{si no.} \end{cases}$

Así, $L(\mathbf{x}, \theta) = g_\theta(T(\mathbf{x}))h(\mathbf{x})$, $\forall \mathbf{x} \in \mathbb{R}^n$, $\forall \theta > 0$ y se deduce que T es suficiente.

Ejercicio Considerar el ejemplo en la familia de leyes de probabilidad uniformes $U(\theta_1, \theta_2)$, con $\theta_1 \in \mathbb{R}$, $\theta_2 \in \mathbb{R}$, $\theta_2 > \theta_1$ y demostrar que la estadística $T = (X_{(1)}, X_{(n)})$ es suficiente.

Observación Si T es una estadística suficiente permite trabajar más fácilmente sobre la estructura inducida por T , pues ella contiene toda la información de la estructura inicial. En particular, en los ejemplos anteriores, T toma valores en \mathbb{R} o \mathbb{R}^2 cualquiera que sea el tamaño de la muestra; i.e. se obtiene una reducción en los datos que puede ser considerable.

- Demostrar que la definición de suficiencia estadística es equivalente a decir que para toda estadística integrable X existe una determinación $E_P(X/\mathfrak{B})$ común a todas las leyes $P \in \mathfrak{P}$.
- Demostrar en el teorema de factorización de Neyman–Fisher que la función $g_\theta = \frac{dP_{\theta T}}{dP_T^*}$, donde $P_{\theta T}$ y P_T^* son las medidas inducidas por T .

2.10 Proyección de una estadística. Teorema de Rao–Blackwell

Consideremos una estructura estadística $(\Omega, \mathfrak{A}, \{P_\theta: \theta \in \Theta\})$ y \mathfrak{B} una sub-tribu de \mathfrak{A} . Al tomar una estadística X que sea integrable, $E_\theta^{\mathfrak{B}}(X)$ no es en general una estadística pues depende de $\theta \in \Theta$.

Definición 2.10.1 Sea $(\Omega, \mathfrak{A}, \mathfrak{P})$ una estructura estadística, diremos que X se proyecta sobre \mathfrak{B} , si $E_\theta^{\mathfrak{B}}(X) = \int X dP_\theta^{\mathfrak{B}}$ no depende de $\theta \in \Theta$, $\mathfrak{P} = \{P_\theta: \theta \in \Theta\}$ y lo denotamos $E^{\mathfrak{B}}(X)$.

- Es claro que $E_\theta^{\mathfrak{B}}(X)$ es una estadística, si X se proyecta sobre \mathfrak{B} .
- Si \mathfrak{B} es suficiente, toda estadística real X integrable se proyecta sobre \mathfrak{B} , pues la esperanza de X condicionada a la tribu \mathfrak{B} , con respecto a la ley P_θ , es independiente de θ .

En efecto, si \mathfrak{B} es suficiente, existe una determinación de $P_\theta^{\mathfrak{B}}(A) = E_\theta^{\mathfrak{B}}(I_A)$ independiente de $\theta \in \Theta$, $\forall A \in \mathfrak{A}$, de donde por linealidad, existe una determinación independiente de θ de $E_\theta^{\mathfrak{B}}(X)$ para una estadística real escalonada X .

Toda estadística real integrable $X(= X^+ - X^-)$, X^+ (resp. X^-) es límite creciente de funciones escalonadas X_n^+ (resp. X_n^-) y se tiene $E_\theta^{\mathfrak{B}}(X^+) = E_\theta^{\mathfrak{B}}(\sup X_n^+) = \sup E_\theta^{\mathfrak{B}}(X_n^+)$, de donde se deduce la propiedad.

Inversamente, si toda estadística real integrable se proyecta sobre \mathfrak{B} , entonces \mathfrak{B} es suficiente. En efecto el resultado se deduce considerando en particular las indicatrices.

Teorema 2.10.1 Teorema de Rao-Blackwell

Sea $(\Omega, \mathfrak{A}, \{P_\theta: \theta \in \Theta\})$ una estructura estadística, \mathfrak{B} una subtribu de \mathfrak{A} y X una estadística real perteneciendo a L_2 , tal que X se proyecta sobre \mathfrak{B} (es el caso en particular si \mathfrak{B} es suficiente). Sea $Y = E^{\mathfrak{B}}(X)$ esta proyección, entonces $\theta \in \Theta$, $\text{var}_\theta(Y) \leq \text{var}_\theta(X)$.

Sea $(\Omega, \mathfrak{A}, \{P_\theta: \theta \in \Theta\})$, \mathfrak{B} una sub-tribu de \mathfrak{A} y X una estadística vectorial tal que sus componentes son cuadrado integrable. Si

$$Y = E^{\mathfrak{B}}(X) = \begin{bmatrix} E^{\mathfrak{B}}(X_1) \\ \vdots \\ E^{\mathfrak{B}}(X_n) \end{bmatrix}, \text{ con } X = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix},$$

entonces $\forall \lambda = (\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n$, $\forall \theta \in \Theta$ $\text{var}_\theta(\lambda'Y) \leq \text{var}_\theta(\lambda'X)$ i.e. $\forall \lambda \in \mathbb{R}^n$, $\forall \theta \in \Theta$, $\lambda' \Sigma_Y^\theta \lambda \leq \lambda' \Sigma_X^\theta \lambda$, con Σ_Z^θ la matriz de varianza-covarianza de la variable Z con respecto a la ley P_θ .

Demostración La primera parte es una aplicación inmediata de las propiedades de esperanzas condicionadas (desigualdad de Jensen).

Para la segunda parte se tiene que:

$$E_\theta^{\mathfrak{B}}(\lambda'X) = E_\theta^{\mathfrak{B}}(\lambda_1 X_1 + \dots + \lambda_n X_n) = \sum_{i=1}^n \lambda_i E_\theta^{\mathfrak{B}}(X_i) = \lambda'Y,$$

de donde $\text{var}_\theta(\lambda'Y) \leq \text{var}_\theta(\lambda'X)$, $\forall \lambda \in \mathbb{R}^n$, $\forall \theta \in \Theta$ por la parte primera. Notemos que $\text{var}(\lambda_1 Z_1 + \dots + \lambda_n Z_n) = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \text{cov}(Z_i, Z_j) = \lambda' \Sigma_Z \lambda$.

2.10.1 Aplicación a la estimación

El teorema de Rao-Blackwell tiene una aplicación inmediata en la teoría de estimación. Supongamos que X es un estimador sin sesgo de $f(\theta) \in \mathbb{R}$ i.e. $E_\theta(X) = f(\theta)$. Si \mathfrak{B} es una tribu suficiente (o simplemente si X se proyecta sobre \mathfrak{B}) y si $Y = E^{\mathfrak{B}}(X)$, entonces Y es un estimador sin sesgo de $f(\theta)$ pues:

$$E_\theta(Y) = E_\theta(E^{\mathfrak{B}}(X)) = E_\theta(E_\theta^{\mathfrak{B}}(X)) = E_\theta(X) = f(\theta).$$

Además si $X \in L_2$ se tiene $\text{var}_\theta(Y) \leq \text{var}_\theta(X)$, es decir que Y es un estimador mejor que X en el sentido de error cuadrático medio.

- Si \mathfrak{B} es la tribu engendrada por la estadística suficiente T , $Y = E^{\mathfrak{B}}(X) = E^T(X) = g \circ T$, es decir, Y es función de la estadística suficiente T . Se ha proyectado X sobre la estadística suficiente T . Así, dado un estimador X (cuadrado integrable) sin sesgo de $f(\theta)$, se puede encontrar siempre un estimador, al menos tan *bueno* como X en función solamente de la estadística suficiente.
- La noción de suficiencia y estructura estadística completa nos conduce a un resultado interesante en la estimación.

Sea $(\Omega, \mathfrak{B}, \{P_\theta: \theta \in \Theta\})$ una estructura estadística completa y \mathfrak{B} una tribu suficiente. Sea Z un estimador sin sesgo de $f(\theta)$, cuadrado integrable y $Y' = E^{\mathfrak{B}}(Z)$, $Y = E^{\mathfrak{B}}(X)$. Sabemos que $Y - Y'$ es \mathfrak{B} -medible y que $\forall \theta \in \Theta$, $E_\theta(Y - Y') = f(\theta) - f(\theta) = 0$ y por la propiedad de completitud $Y = Y'$, \mathfrak{B} -c.p.d. i.e. $\text{var}_\theta(Y) = \text{var}_\theta(Y') \leq \text{var}_\theta(Z)$.

Así todo estimador sin sesgo de $f(\theta)$ tiene la misma proyección sobre \mathfrak{B} , \mathfrak{B} -c.p.d. y esta proyección es el *mejor* estimador sin sesgo de $f(\theta)$ (en el sentido de varianza mínima).

2.11 Método de máxima verosimilitud: una técnica de estimación

Se considera una estructura estadística $(\Omega, \mathfrak{A}, \{\mathfrak{P}_\theta: \theta \in \Theta\})$ dominada por μ y de función de verosimilitud $L(\omega, \theta)$. Para cada ω se puede encontrar en general un valor θ que maximice $L(\omega, \theta)$. De esta manera tenemos una correspondencia $\Omega \xrightarrow{\hat{\theta}} \Theta$ i.e. si $\hat{\theta}$ es una función medible, es una estadística $\omega \mapsto \hat{\theta}(\omega)$

y se puede considerar como un estimador de θ . Des este modo diremos que $\hat{\theta}$ es el *estimador de máxima verosimilitud* i.e. $L(\omega, \hat{\theta}(\omega))$ es un máximo.

Nota No necesariamente el estimador $\hat{\theta}$ es sin sesgo.

Ejemplos

1. Sea $L(\mathbf{x}, \theta) = f_\theta(\mathbf{x}) = \theta^{\sum_{i=1}^n x_i} (1 - \theta)^{n - \sum_{i=1}^n x_i}$.

$$\frac{\partial \ln L(\mathbf{x}, \theta)}{\partial \theta} = \frac{\sum_{i=1}^n x_i}{\theta} - \frac{n - \sum_{i=1}^n x_i}{1 - \theta} = 0 \iff \theta = \frac{1}{n} \sum_{i=1}^n x_i \text{ (verificar que es máximo), de donde } \hat{\theta}(\mathbf{x}) =$$

$$\frac{1}{n} \sum_{i=1}^n x_i \text{ i.e. } \hat{\theta} = \frac{1}{n} \sum_{i=1}^n X_i.$$

$$2. (\mathbb{R}, \mathfrak{B}_{\mathbb{R}, \{N(\mu, \sigma^2); \mu \in \mathbb{R}, \sigma > 0\}})^n, \quad \ln L(\mathbf{x}, \mu, \sigma) = -n \ln \sigma + k - \frac{1}{2} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^2,$$

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu), \quad \frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^n (x_i - \mu)^2.$$

- Si σ es fijo (conocido), $\frac{\partial \ln L}{\partial \mu} = 0$ sii $\hat{\mu}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n x_i$
- Si μ conocido, $\frac{\partial \ln L}{\partial \sigma} = 0$ sii $\widehat{\sigma^2}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$.
- Si μ y σ son desconocidos, $\frac{\partial \ln L}{\partial \mu} = 0 \implies \hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}$ y $\frac{\partial \ln L}{\partial \sigma} = 0 \implies \widehat{\sigma^2} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$.

$$E_{(\mu, \sigma)}(\widehat{\sigma^2}) = \frac{n-1}{n} \sigma^2 \text{ y } S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \text{ es un estimador insesgado de } \sigma^2.$$

3. $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}, \{U(\theta); \theta > 0\}})^n$ está dominado por la medida de Borel sobre \mathbb{R}^n .

$$L(\mathbf{x}, \theta) = \begin{cases} \frac{1}{\theta^n} & \text{si } x_i \in [0, \theta], \quad i = 1, \dots, n \\ 0 & \text{si no.} \end{cases}$$

$\hat{\theta} = \sup X_i = X_{(n)}$ y se considera $T = \frac{n+1}{n} \hat{\theta}$ un estimador sin sesgo de θ .

Si existe una estadística suficiente, el estimador está en función de la estadística suficiente.

2.12 Noción de libertad

Sea $(\Omega, \mathfrak{A}, \mathfrak{P})$ una estructura estadística; se dice que:

- el evento A es libre si $P(A)$ es constante $\forall P \in \mathfrak{P}$.
- la subtribu \mathfrak{B} de \mathfrak{A} es libre si cada uno de los eventos es libre.
- la estadística T con valores en $(\mathfrak{T}, \mathfrak{C})$ es libre sii $T^{-1}(\mathfrak{C})$ es libre i.e. sii la ley de probabilidad de T es constante para todo $P \in \mathfrak{P}$ (o sea \mathfrak{P}_T tiene un solo elemento)
- la estadística T real integrable es libre en media sii $E_P(T)$ no depende de $P \in \mathfrak{P}$.
- la estadística T real integrable es libre en media sobre $\mathfrak{P}_0 \subset \mathfrak{P}$ sii $E_P(T)$ no depende de $P \in \mathfrak{P}_0$.

Teorema 2.12.1 *Sea $(\Omega, \mathfrak{A}, \mathfrak{P})$ una estructura estadística y sea $\mathfrak{B} \subset \mathfrak{A}$ una sub-tribu, \mathfrak{B} es completa (i.e. cuasi completa) si y sólo si toda estadística (resp. acotada) libre en media que se proyecta sobre \mathfrak{B} es libre en media condicionada a \mathfrak{B} .*

Demostración

(\Leftarrow) Si X es libre en media, se proyecta sobre \mathfrak{B} y se define $Y = E^{\mathfrak{B}}(X) - E(X)$, \mathfrak{B} -medible, integrable (resp. acotada), centrada y como \mathfrak{B} es completa (resp. cuasi completa), $Y = 0$ \mathfrak{P} -c.p.d. y se tiene $E^{\mathfrak{B}}(X) = E(X)$, \mathfrak{P} -c.p.d., o sea X es libre en media condicionada a \mathfrak{B} .

(\Rightarrow) Si \mathfrak{B} no es completa (resp. cuasi completa) existe una estadística real Y no \mathfrak{P} -c.p.d. equivalente a 0, \mathfrak{B} -medible, integrable (resp. acotada) y centrada, entonces Y es \mathfrak{B} -medible y se proyecta sobre \mathfrak{B} , $E_P^{\mathfrak{B}}(Y) = Y$, $\forall P \in \mathfrak{P}$ y no se puede tener $E_P^{\mathfrak{B}}(Y) = E(Y) = 0$, \mathfrak{P} -c.p.d.

Proposición 2.12.1 *Sea \mathfrak{B} una subtribu suficiente y cuasi-completa por la estructura estadística $(\Omega, \mathfrak{A}, \mathfrak{P})$ y sea \mathfrak{C} una tribu libre para la misma estructura, entonces \mathfrak{B} y \mathfrak{C} son independientes para toda ley $P \in \mathfrak{P}$.*

Demostración Sea $C \in \mathfrak{C}$, entonces tenemos que $P(C/\mathfrak{B}) = P(C)$, $\forall P \in \mathfrak{P}$ pues \mathfrak{B} es cuasi-completa, por lo tanto:

$$\forall B \in \mathfrak{B}, P(B \cap C) = \int_B P^{\mathfrak{B}}(C) dP = P(C)P(B), \forall P \in \mathfrak{P}.$$

i.e. \mathfrak{B} y \mathfrak{C} son tribus independientes para toda ley $P \in \mathfrak{P}$.

Corolario 2.12.1 *Sea $(\Omega, \mathfrak{A}, \mathfrak{P})$ una estructura estadística y sea T una estadística suficiente con valores en $(\mathfrak{T}, \mathfrak{C})$. Si la estructura $(\mathfrak{T}, \mathfrak{C}, \mathfrak{P}_T)$ es cuasi-completa, toda estadística libre por la estructura $(\Omega, \mathfrak{A}, \mathfrak{P})$ es independiente de T .*

Demostración Sea $\mathfrak{B} = T^{-1}(\mathfrak{C})$ la tribu suficiente y cuasi-completa por la estructura $(\Omega, \mathfrak{A}, \mathfrak{P})$. Sea \mathfrak{D} la tribu generada por la estadística libre X , entonces \mathfrak{B} y \mathfrak{D} son independientes para toda ley $P \in \mathfrak{P}$.

2.13 Ejercicios

1. Sea \bar{X} la media de una muestra de tamaño 5 de una distribución $N(0, 125)$. Determinar c tal que $P(\bar{X} < c) = 0.9$.
2. Sea \bar{X} la media de una muestra de tamaño n de una distribución normal con media μ y varianza 100. Calcular n para el cual $P(\mu - 5 < \bar{X} < \mu + 5) = 0.954$.

3. Sea X_1, \dots, X_{25} y Y_1, \dots, Y_{25} dos muestras independientes normales de distribución $N(0, 16)$ y $N(1, 9)$ respectivamente. Sean \bar{X} y \bar{Y} las medias muestrales. Determinar la $P(\bar{X} > \bar{Y})$.
4. Determinar la media y la varianza de $S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ donde X_1, \dots, X_n es una muestra de distribución $N(\mu, \sigma^2)$.
5. Sea S^2 la varianza de una muestra de tamaño 6 de distribución $N(\mu, 12)$. Determinar $P(2.3 < S^2 < 22.2)$.
6. Determinar la densidad de la varianza de la varianza muestral S^2 , cuyas observaciones provienen de una población normal $N(\mu, \sigma^2)$.
7. Sea \bar{X} y S^2 la media y la varianza muestral de una muestra de tamaño 25 de distribución $N(3, 100)$. Evaluar $P([0 < \bar{X} < 6])$ y $[55.2 < S^2 < 145.6]$.
8. Sea X una variable aleatoria continua tal que $E(X) = \mu$. Sea $Y = v(X)$ integrable donde $v(x) = v(\mu) + v'(\mu)(x - \mu) + v''(\eta)(x - \mu)^2/2$, con η entre μ y x .
Si $v''(x) \geq 0, \forall x$, probar que $E(v(X)) \geq v(\mu) = v(E(X))$ y
 $v''(x) \leq 0, \forall x$, probar que $E(v(X)) \leq v(\mu) = v(E(X))$.
9. Sea X una variable aleatoria con función de densidad $f(x) = e^{-x}, 0 < x < \infty$. Calcular la probabilidad de que el intervalo aleatorio $(X, 3X)$ contenga el punto $x = 3$. ¿Cuál es el valor esperado de la longitud de ese intervalo?
10. Sean X_1, X_2 variables aleatorias con densidad $f(x) = \frac{1}{4}$, para $x = 1, 2, 3, 4$. Calcular la probabilidad de que el intervalo aleatorio $(1, X_1 + X_2 - \frac{1}{2})$ contenga el punto 3.
11. Sea X_1, X_2 una muestra de tamaño 2 de distribución uniforme en $[0, 1]$. Calcular la probabilidad de que el intervalo $(X_1/(3X_2), 2X_1/X_2)$ contenga el punto $\frac{2}{3}$.
12. Sea \bar{X} la media de una muestra de tamaño n de distribución $N(\mu, \sigma^2)$. Determinar la probabilidad de que el intervalo aleatorio $[n(\bar{X} - \mu)^2/5.02, n(\bar{X} - \mu)^2/0.001]$ incluya el punto σ^2 . ¿Cuál es el valor esperado de la longitud de ese intervalo? Note que este valor es el mismo para todo $n \in \mathbb{N}$.
13. Sea X_1, \dots, X_n una muestra de distribución $N(\mu, \sigma^2)$, donde σ^2 es conocida. Determinar la probabilidad de que el intervalo $(\bar{X} \pm 2\sigma/\sqrt{n})$ incluya el punto μ . Si $n = 40$ y $\sigma^2 = 10$ y $\bar{x} = 7.164$, determinar

un intervalo con un coeficiente de seguridad del 80%.

14. Sea X_1, \dots, X_n y Y_1, \dots, Y_m dos muestras de distribuciones respectivas $N(\mu_1, \sigma^2)$ y $N(\mu_2, \sigma^2)$, con parámetros desconocidos. Demostrar que la estadística:

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{nS_1^2 + mS_2^2}{n+m-2} \left(\frac{1}{n} + \frac{1}{m}\right)}} \sim t_{n+m-2}$$

y sirve para encontrar un intervalo de confianza que contenga a μ , con un coeficiente de seguridad $1 - \alpha$. Determinar el intervalo para el caso $n = 10$, $m = 7$, $\bar{x} = 4.2$, $\bar{y} = 3.4$, $S_1^2 = 49$, $S_2^2 = 32$, $\alpha = 0.1$

15. a) Discutir el problema anterior para el caso en que σ_1^2 y σ_2^2 son conocidas pero no necesariamente iguales.
 b) Discutir el problema anterior para el caso en que σ_1^2 y σ_2^2 son desconocidas y no necesariamente iguales ¿Qué resultado propondría?
16. Sea X_1, \dots, X_n una muestra de tamaño n de distribución $N(\mu, \sigma^2)$. Si σ no es conocida, determinar un intervalo de confianza para μ basado en la estadística $\sqrt{n}(\bar{X} - \mu)/S$, con un coeficiente de 0.95.
17. Sea X_1, \dots, X_n, X_{n+1} una muestra de tamaño $n + 1$ ($n > 1$) de distribución $N(\mu, \sigma^2)$. Sea $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ y $S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$. Determinar la constante c tal que la estadística $c(\bar{X} - X_{n+1})/S \sim t_{n-1}$. Si $n = 8$, determinar k tal que $P(\bar{X} - kS < X_9 < \bar{X} + kS) = 0.8$.
18. Sea X_1, \dots, X_n una muestra de distribución $N(0, \theta)$, con $\theta > 0$. Probar que $\sum_{i=1}^n X_i^2$ es una estadística suficiente de θ .
19. Sabemos que si X_1, \dots, X_n es una muestra de distribución uniforme $U(0, \theta)$, $X_{(n)}$ es una estadística suficiente para θ . Generalizar este resultado considerando una muestra con distribución $f(x, \theta) = Q(\theta)M(x)$, si $0 < x < \theta$, $0 < \theta < \infty$. Note que $\int_0^\theta M(x)dx = Q(\theta)$.
20. Probar que si $a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0 = 0$ para más de n puntos, entonces $a_0 = \dots = a_n = 0$. Use este resultado para probar que la familia $\{B(n, \theta)/0 < \theta < 1\}$ es completa.
21. Si la serie $a_0 + a_1 z + \dots + a_n z^n + \dots$ converge a cero en todo punto de un intervalo, entonces $a_0 = a_1 = \dots = 0$. Use este hecho para probar $\{P(\lambda): \lambda > 0\}$ es completa.

22. Sea $Y_1 < Y_2 < \dots < Y_n$ estadística de orden de una muestra de tamaño n , de distribución $f(x, \theta) = \frac{1}{\theta}$, con $0 < x < \theta$, $0 < \theta < \infty$. Probar que la familia de probabilidades de Y_n es completa.
23. Sea X una variable aleatoria con densidad $f(x, \theta) = Q(\theta)M(x)$, $0 < x < \theta$, $\theta > 0$, donde $M(x)$ es continua. Sea $Y_1 < \dots < Y_n$ estadísticas de orden de una muestra de tamaño n de distribución igual a X . Probar que la familia de densidades de Y_n es completa (note que Y_n es suficiente).
24. Probar que para cada una de las siguientes familias $\{f(x, \theta): 0 < \theta < +\infty\}$ no es completa, definiendo al menos una función $U(x) \neq 0$ tal que $E_\theta(U(X)) = 0$, $\forall \theta > 0$:
- $f(x, \theta) = \frac{1}{2\theta}$, si $-\theta < x < \theta$
 - $N(0, \theta)$.
25. Probar que la primera estadística de orden Y_1 de una muestra de tamaño n , de distribución $f(x, \theta) = e^{-(x-\theta)}$, $\theta < x < +\infty$, $-\infty < \theta < +\infty$, es una estadística suficiente y completa para θ . Determinar la única estadística función de Y_1 que sea el mejor estimador de θ .
26. Sea X_1, \dots, X_n muestra con distribución $f(x, \theta) = \frac{1}{6\theta^4} x^3 e^{-x/\theta}$, $0 < x < +\infty$, $0 < \theta < +\infty$ i.e. distribución gamma $\Gamma(4, \frac{1}{\theta})$. Determinar una estadística completa y suficiente T de θ y la única función $\Phi(T)$ que es el mejor estimador de θ . ¿Es $\Phi(T)$ una estadística suficiente y completa?
27. Sea X_1, \dots, X_n una muestra de distribución:
- $f(x, \theta) = \theta x^{\theta-1}$, $0 < x < 1$, $0 < \theta < +\infty$
 - $f(x, \theta) = \frac{1}{2} e^{-|x-\theta|}$, $-\infty < x < +\infty$, $-\infty < \theta < +\infty$
 - $f(x, \theta) = e^{-(x-\theta)}$, $\theta < x < +\infty$, $-\infty < \theta < +\infty$.
- En cada caso determinar la estadística $\hat{\theta}$ de máxima verosimilitud de θ . Si una estadística suficiente Y existe, expresar $\hat{\theta}$ en función de Y .
28. Sea X_1, \dots, X_n una muestra con densidad $f(x, \theta_1, \theta_2) = \frac{1}{\theta_2} e^{-\frac{(x-\theta_1)}{\theta_2}}$, si $\theta_1 \leq x < +\infty$, $-\infty < \theta_1 < +\infty$, $0 < \theta_2 < +\infty$. Determinar una estadística de máxima verosimilitud de θ_1 y θ_2 . Encontrar estimadores insesgados de θ_1 y θ_2 si existen.
29. Sea X_1, \dots, X_n una muestra con función de densidad $f(x, \theta)$, $\theta \in \Theta$. Sea $L(\omega, \theta)$ la función de

verosimilitud i.e. $L(\omega, \theta) = f(x_1, \theta) \cdots f(x_n, \theta)$, con $\omega = (x_1, \dots, x_n)$. Sea $T = u(X_1, \dots, X_n)$ una estadística suficiente para θ . Si $g(t, \theta_0)$ es la función de densidad de T en el punto $\theta = \theta_0$, entonces la función de densidad $g(t, \theta)$ de la estadística suficiente T para todo $\theta \in \Theta$, es dada por:

$$g(t, \theta) = g(t, \theta_0) \frac{L(\omega, \theta)}{L(\omega, \theta_0)},$$

con $T(\omega) = T(x_1, \dots, x_n) = t$. El resultado prueba que si se puede definir la función de densidad de la estadística suficiente en todo punto especial, se puede definir la función de densidad de la estadística suficiente en todo punto $\theta \in \Theta$. El punto θ puede representar cualquier número de parámetros.

30. Consideremos una muestra X_1, \dots, X_n , con densidad $f(x, \theta)$, $\theta \in \Theta$. Se considera $\hat{\theta} = u(X_1, \dots, X_n)$ la estadística de máxima verosimilitud de θ . Si $h(\theta)$ es función de θ , probar que $h(u(x_1, \dots, x_n))$ es el valor de $h(\theta)$ en el cual se obtiene el máximo. La estadística resultante $h(u(X_1, \dots, X_n))$ es llamada la estadística de máxima verosimilitud de $h(\theta)$ y se denota $\hat{h}(\theta) = h(\hat{\theta})$; este hecho es llamado la propiedad de invarianza de la estadística de máxima verosimilitud. El resultado se puede extender al caso de más de un parámetro.

31. Sea X_1, \dots, X_n una muestra de distribución $N(\theta_1, \theta_2)$. Determinar la estadística de máxima verosimilitud para las siguientes funciones:

a) $h(\theta_1) = \theta_1^2$

b) $h(\theta_2) = \sqrt{\theta_2}$

c) $h(\theta_1, \theta_2)$, con $P(X < h(\theta_1, \theta_2)) = 0.9$

d) $h(\theta_1, \theta_2) = P(X < c)$, con c una constante conocida.

32. Sea X una variable aleatoria con función de densidad $f(x, \theta) = \theta x^{\theta-1}$, $0 < x < 1$, $\theta \in \{1, 2\}$. Para probar la hipótesis $H_0: \theta = 1$ contra $H_1: \theta = 2$ se utiliza la muestra X_1, X_2 y se define la región crítica por $C = \{(x_1, x_2) / \frac{3}{4}x_1 < x_2\}$. Determinar la potencia de la prueba.

33. Sea X una variable aleatoria con distribución de Poisson de parámetro θ . Sea $H_0: \theta = \frac{1}{2}$ y $H_1: \theta < \frac{1}{2}$ i.e. $\theta = \{\theta / 0 \leq \theta < \frac{1}{2}\}$. Sea X_1, \dots, X_{12} una muestra de distribución igual a X . Se rechaza H_0 si $X_1 + \dots + X_{12} \leq 2$. Sea $\beta(\theta)$ la potencia de la prueba, encontrar $\beta(\frac{1}{2}), \beta(\frac{1}{4}), \beta(\frac{1}{6})$ y $\beta(\frac{1}{12})$ y graficar la

función β . ¿Cuál es el nivel de la prueba?

34. Sean X_1, \dots, X_n una muestra de tamaño n de ley uniforme $U[0, \theta]$. Se considera $V_n = \sup X_i$, $U_n = \inf X_i$
- Estudiar la convergencia de V_n .
 - Estudiar la convergencia de U_n .
 - Encontrar un estimador T de θ en función de V_n teniendo “buenas” propiedades i.e. que sea sin sesgo y convergente.
 - Considere \bar{X} y defina $Z = 2\bar{X}$. Comparar Z y T .
35. Sea $X \sim f(x) = \frac{1}{\theta} e^{-x/\theta}$, si $x \geq 0$ i.e. $X \sim \Gamma(1, \frac{1}{\theta})$.

Nota En la práctica X es la duración de una máquina, θ es la duración “vida promedio” de la máquina. Sea X_1, \dots, X_n observaciones de X (i.e. una muestra de tamaño n).

- Encontrar un estimador sin sesgo y convergente del parámetro θ .
- Se desea escoger entre las hipótesis $H_0: \theta \leq \theta_0$, (θ_0 dado) y $H_1: \theta > \theta_0$. Para esto se utiliza al estimador T de θ y se escoge como regla:

$$\Phi = 1 \quad \text{si } T > \ell$$

$$\Phi = 0 \quad \text{si } T \leq \ell.$$

Determinar al imponer el nivel $\alpha = \sup_{\theta \in H_0} E_\theta(\Phi)$, que T es equivalente a un χ_{2n}^2 . Encontrar ℓ para $n = 15$, $\alpha = 0.05$, $\theta_0 = 4$.

- Graficar la función potencia de Φ (i.e. $E_\theta(\Phi)$, $\theta \in H_0 \cup H_1$).
- Recordemos que X es la duración “vida” de la máquina. Sea Z el número de máquinas que fallan “mueren” entre 0 y t_0 .

Se considera la prueba:

$$\Psi = 1 \quad \text{si } Z = 0$$

$$\Psi = 0 \quad \text{si } Z \geq 1.$$

- Determinar t_0 para θ_0 , n y α dados. (Verificar que $Z = \sum_{i=1}^n \mathbb{1}_{[X_i < t_0]}$).

d.2) Estudiar la potencia de Ψ .

36. Sea X una variable aleatoria continua uniforme en $[\theta - \frac{1}{2}, \theta + \frac{1}{2}]$, θ parámetro real. Sean X_1, \dots, X_n , n variables aleatorias independientes de misma ley que X , se asocia a esta muestra las estadísticas de orden $X_{(1)}, \dots, X_{(n)}$.

a) Calcular $E(X)$, $\text{var}(X)$, $E(X_{(1)})$, $\text{var}(X_{(1)})$, $\text{var}(X_{(n)})$, $\text{cov}(X_{(1)}, X_{(n)})$.

b) Estudiar la estadística $S = \frac{1}{n} \sum_{i=1}^n X_i$, en tanto se considere estimador de θ (sesgo, convergencia).

c) ¿Cuál es la distribución asintótica de $\sqrt{n}(S - \theta)$, cuando $n \rightarrow \infty$?

d) Encontrar un estimador de θ de la forma $aX_{(1)} + bX_{(n)} + c$ (a, b, c constantes) que sea sin sesgo. Entre éstos estimadores encontrar el de varianza mínima y denótelo T .

e) Comparar T y S .

37. Sea la variable aleatoria real X con función de repartición

$$F(x) = \begin{cases} 0 & \text{si } x < \theta \\ 1 - e^{-(x-\theta)} & \text{si } x \geq \theta. \end{cases}$$

Sean X_1, \dots, X_n una muestra de n variables aleatorias independientes distribuidas como X . Sea $X_{(1)} = \inf X_i$.

a) Determinar la función de densidad y la función de repartición de $X_{(1)}$. Calcular $E(X_{(1)})$ y deducir un estimador U de θ que sea sin sesgo y convergente.

b) Sea $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$. Calcular $E(\bar{X})$ y deducir un estimador V , sin sesgo y convergente de θ . Comparar U y V .

c) Para realizar una prueba de hipótesis $H_0: \theta = \theta_0$ contra $H_1: \theta \neq \theta_0$ se adopta la regla siguiente: Se rechaza H_0 si $X_{(1)} < \theta_0$ o $X_{(1)} > \ell_\alpha$.

c.1) Determinar ℓ_α en función de α, θ_0, n .

c.2) Calcular la imagen de la prueba. Estudiar las variaciones de la potencia de la prueba según los valores de θ . ¿La prueba es sin sesgo? ¿Es convergente?

38. El problema consiste en estudiar algunas propiedades de la estructura estadística (2) definida abajo.

En la primera parte se obtienen resultados preliminares, en la segunda parte se efectúa una prueba de hipótesis y al final se estudia la estimación del parámetro.

a) Sobre un espacio de probabilidad dado, se define una variable aleatoria real X con densidad respecto a la medida de Borel f_θ tal que:

$$f_\theta(x) = \begin{cases} \theta x - \frac{\theta}{2} + 1 & \text{si } x \in [0, 1] \\ 0 & \text{si no.} \end{cases} \quad (1)$$

θ un parámetro real tal que $\theta \in [-2, 2]$.

a.1) Calcular en función de θ la media y la varianza de X y de $\log X$.

a.2) ¿Cuál es la ley de $-\log X$ y de $-2 \log X$ cuando $\theta = 0$?

b) Se considera la estructura estadística:

$$(\mathbb{R}, \mathfrak{B}_{\mathbb{R}, \{P_\theta: \theta \in [-2, 2]\}})^n \quad (2)$$

donde P_θ es la probabilidad de densidad f_θ definida en (1), $n \in \mathbb{N}^*$.

Se considera la hipótesis $H_0 = \{P_0\}$, $H' = \{P_\theta: \theta \in]0, 2]\}$ y se define una prueba por:

$$\Phi(\mathbf{x}) = 1 \quad \text{si } \prod_{i=1}^n x_i > k$$

$$\Phi(\mathbf{x}) = 0 \quad \text{si } \prod_{i=1}^n x_i \leq k,$$

donde $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$.

Demostrar que se puede determinar k a partir de las tablas de ley χ^2 para que Φ como prueba de H_0 contra H' sea de nivel α .

Como aplicación precisar la prueba Φ para $\alpha = 0.1$ y $n = 10$.

c) Siempre bajo la estructura estadística (2), nos interesamos en la estimación de θ y la calidad de los estimadores insesgados por medio de sus varianzas.

c.1) Sea $X_i (i = 1, \dots, n)$ las aplicaciones coordenadas y se define:

$$T = 12 \left(\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{2} \right).$$

Demostrar que T es un estimador sin sesgo y convergente de θ . Dar la ley límite de $\sqrt{n}(T - \theta)$, cuando $n \rightarrow \infty$.

c.2) Encontrar una función lineal U de $\frac{1}{n} \sum_{i=1}^n \log X_i$ que sea un estimador sin sesgo de θ . Comparar los estimadores U y T .

Capítulo 3

Pruebas de hipótesis

3.1 Generalidades

Consideramos siempre una estructura estadística $(\Omega, \mathfrak{A}, \mathfrak{P})$ y una prueba Φ con valores en $[0, 1]$. Una prueba de la hipótesis \mathfrak{P}_0 contra \mathfrak{P}_1 es una prueba para la cual está asociada la regla de decisión:

Rechazar \mathfrak{P}_0 con la probabilidad $\Phi(\omega)$, si ω es el resultado de la experiencia.

Se demuestra que $\beta_\Phi(P)$ se interpreta como la probabilidad de rechazar la hipótesis \mathfrak{P}_0 (aceptar \mathfrak{P}_1) si P es la verdadera ley de probabilidad.

Definición 3.1.1 *Dos pruebas Φ y Φ' de \mathfrak{P}_0 contra \mathfrak{P}_1 son equivalentes sii:*

$$\beta_\Phi(P) = \beta_{\Phi'}(P), \quad \forall P \in \mathfrak{P}_0 \cup \mathfrak{P}_1 \text{ y se denota } \Phi \sim \Phi'.$$

Nota No debe confundirse esta noción de equivalencia con la noción definida para estadísticas; esto significa, en particular, que el valor estadístico de una prueba entre dos hipótesis está enteramente representada por su función potencia.

Recordemos que existen dos maneras de escoger una hipótesis falsa:

Escoger \mathfrak{P}_1 cuando $P \in \mathfrak{P}_0$, eventualidad a la cual está asociada la función $\beta_\Phi(P)$, con $P \in \mathfrak{P}_0$.

Escoger \mathfrak{P}_0 cuando $P \in \mathfrak{P}_1$, eventualidad a la cual está asociada $1 - \beta_\Phi(P)$, con $P \in \mathfrak{P}_1$.

Definición 3.1.2 *Sea Φ una prueba de \mathfrak{P}_0 contra \mathfrak{P}_1 , se llama nivel de significación de Φ el número*

α_Φ :

$$\alpha_\Phi = \sup\{\beta_\Phi(P) / P \in \mathfrak{F}_0\}.$$

Se dice que Φ es sin sesgo, si $\alpha_\Phi \leq \beta_\Phi(P)$, $\forall P \in \mathfrak{F}_1$.

Se dice que Φ es trivial si es constante sobre Ω .

Se dice que Φ es ineficaz si su función potencia es constante.

Proposición 3.1.1 Si Φ es una prueba de hipótesis de \mathfrak{F}_0 contra \mathfrak{F}_1 sobre la estructura $(\Omega, \mathfrak{A}, \mathfrak{P})$ y U es la ley uniforme sobre $[0, 1]$ entonces existe una prueba Ψ sobre $(\Omega \times [0, 1], \mathfrak{A} \otimes \mathfrak{B}_{[0,1]}, \mathfrak{P} \otimes U)$ tal que Ψ es determinista y de misma imagen que Φ .

Prueba Notemos que si $P' \in \mathfrak{P} \otimes U$, $P' = P \otimes U$

$$\text{Sea } \Psi(w, u) = \begin{cases} 1 & \text{si } u < \Phi(w) \\ 0 & \text{si } u \geq \Phi(w) \end{cases} \quad \text{i.e. es determinista sobre } \Omega \times [0, 1] \text{ y:}$$

$$\beta_\Psi(P') = \int_{\Omega} \int_0^1 \Psi(w, u) dU dP = \int_{\Omega} \Phi dP = \beta_\Phi(P).$$

Proposición 3.1.2 Sea Φ una prueba de \mathfrak{F}_0 contra \mathfrak{F}_1 ; si \mathfrak{B} es una tribu suficiente sobre la estructura $(\Omega, \mathfrak{A}, \mathfrak{P}_0 \cup \mathfrak{P}_1)$, la proyección de Φ sobre \mathfrak{B} , $E(\Phi/\mathfrak{B})$ define una prueba equivalente a Φ .

Prueba Sabemos que $E^{\mathfrak{B}}(\Phi)$ es una estadística y como $0 \leq \Phi \leq 1$ entonces $0 \leq E^{\mathfrak{B}}(\Phi) \leq 1$ i.e. $\Phi' = E^{\mathfrak{B}}(\Phi)$ es una prueba. Además:

$$\beta_{\Phi'}(P) = E_P(\Phi') = \beta_\Phi(P), \quad \forall P \in \mathfrak{P}_0 \cup \mathfrak{P}_1.$$

Nota Proyectando una prueba determinista sobre una tribu suficiente, se obtiene en general una prueba estocástica.

Si \mathfrak{B} es la tribu generada por la estadística T entonces $\Phi' = E^{\mathfrak{B}}(\Phi)$ se puede escribir $\Phi' = E^{\mathfrak{B}}(\Phi) = g \circ T$, i.e. si T es suficiente, para toda prueba Φ se puede construir Φ' que depende solo de T tal que $\Phi \sim \Phi'$.

3.2 Relación de preorden sobre las pruebas de hipótesis

El problema esencial de la teoría de pruebas de hipótesis es determinar, si es posible, la mejor prueba de una hipótesis contra otra. Es por eso conveniente abordar el estudio de relaciones de preorden de estas pruebas.

Definición 3.2.1 Sea $(\Omega, \mathfrak{A}, \mathfrak{P})$ una estructura estadística, sean \mathfrak{P}_0 y \mathfrak{P}_1 hipótesis y sea $\mathfrak{T} = \{\Phi/\Phi$ es una prueba de \mathfrak{P}_0 contra $\mathfrak{P}_1\}$. Sean dos pruebas Φ, Φ' de \mathfrak{T} ; se dice que Φ es mejor que Φ' y se denota $\Phi' \leq \Phi$ sii:

$$\beta_{\Phi}(P) \leq \beta_{\Phi'}(P), \quad \forall P \in \mathfrak{P}_0$$

$$\beta_{\Phi'}(P) \leq \beta_{\Phi}(P), \quad \forall P \in \mathfrak{P}_1.$$

Φ es estrictamente mejor que Φ' , ($\Phi > \Phi'$) si $\Phi \geq \Phi'$ y $\Phi \neq \Phi'$.

Se dice que Φ es **admisibles** si no existe una prueba estrictamente mejor.

Se dice que Φ es **cuasi-admisibles** sii $\forall \Psi \in \mathfrak{T}$ tal que $\beta_{\Psi}(P) \leq \beta_{\Phi}(P), \forall P \in \mathfrak{P}_0$

$$\implies \left\{ \begin{array}{l} \beta_{\Psi}(P) = \beta_{\Phi}(P), \quad \forall P \in \mathfrak{P}_1, \\ \text{o bien} \\ \text{existe } P_1 \in \mathfrak{P}_1 \text{ tal que } \beta_{\Phi}(P_1) > \beta_{\Psi}(P_1). \end{array} \right.$$

Observación Es claro que una prueba admisible es cuasi-admisibles pero el inverso no necesariamente es cierto.

Si Φ es admisible como prueba de \mathfrak{P}_0 contra \mathfrak{P}_1 , $1 - \Phi$ es admisible de \mathfrak{P}_1 contra \mathfrak{P}_0 .

Si $\mathfrak{P}'_0 \subset \mathfrak{P}_0$ y Φ es cuasi-admisibles como prueba de \mathfrak{P}'_0 contra \mathfrak{P}_1 .

Si Φ es cuasi-admisibles como prueba de \mathfrak{P}_0 contra \mathfrak{P}_1 y si $1 - \Phi$ es cuasi-admisibles como prueba de \mathfrak{P}_1 contra \mathfrak{P}_0 , entonces Φ es admisible.

Ejemplo Si $\mathfrak{P}_0 = \{\mathfrak{P}_0\}$ y $\mathfrak{P}_1 = \{\mathfrak{P}_1\}$ y si Ψ es admisible se tiene:

$$\beta_{\Psi}(P_0) \leq \beta_{\Phi}(P_0) \implies \beta_{\Psi}(P_1) \leq \beta_{\Phi}(P_1)$$

$$\text{o si } \beta_{\Phi}(P_1) \leq \beta_{\Psi}(P_1) \implies \beta_{\Phi}(\mathfrak{P}_0) \leq \beta_{\Psi}(\mathfrak{P}_0).$$

3.3 Pruebas óptimas

Definición 3.3.1 Sea Φ una prueba de \mathfrak{F}_0 contra \mathfrak{F}_1 la prueba Φ se dice que es uniformemente la más potente U.M.P. si:

$$\forall \Psi \in \mathfrak{T} (\beta_\Psi(P) \leq \beta_\Phi(P), \forall P \in \mathfrak{F}_0) \implies (\beta_\Psi(P) \leq \beta_\Phi(P), \forall P \in \mathfrak{F}_1).$$

Definición 3.3.2 Se dice que Φ es estrictamente U.M.P. si es U.M.P. y admisible.

Nota Una prueba U.M.P es cuasi-admisibile y sin sesgo (pues es mejor que una prueba trivial de igual nivel de significación).

Si \mathfrak{F}_0 y \mathfrak{F}_1 son hipótesis simples, las nociones U.M.P y cuasi-admisibile coinciden.

Teorema 3.3.1 Sea $\mathfrak{F}'_0 \subset \mathfrak{F}_0$ y Φ una prueba de \mathfrak{F}_0 contra \mathfrak{F}_1 de nivel de significación α_Φ . Si como prueba de \mathfrak{F}'_0 contra \mathfrak{F}_1 , Φ es U.M.P de nivel α_Φ , entonces Φ es U.M.P. como prueba de \mathfrak{F}_0 contra \mathfrak{F}_1 .

Demostración Sea Φ' una prueba de \mathfrak{F}_0 contra \mathfrak{F}_1 tal que $\alpha_{\Phi'} \leq \alpha_\Phi$ entonces:

$$\sup\{\beta_{\Phi'}(P)/P \in \mathfrak{F}'_0\} \leq \sup\{\beta_{\Phi'}(P)/P \in \mathfrak{F}_0\} = \alpha_{\Phi'} \leq \alpha_\Phi$$

implica por hipótesis $\beta_{\Phi'}(P) \leq \beta_\Phi(P), \forall P \in \mathfrak{F}_1$.

Teorema 3.3.2 Una prueba Φ de \mathfrak{F}_0 contra \mathfrak{F}_1 es U.M.P. sii $\forall P_1 \in \mathfrak{F}_1$, Φ es U.M.P. como prueba de \mathfrak{F}_0 contra $\{P_1\}$.

La demostración es inmediata.

Definición 3.3.3 Una prueba Φ de \mathfrak{F}_0 contra \mathfrak{F}_1 es libre si su potencia es constante en \mathfrak{F}_0 i.e. $\beta_\Phi(P) = \alpha_\Phi, \forall P \in \mathfrak{F}_0$ (o sea Φ es libre en media sobre $(\Omega, \mathfrak{A}, \mathfrak{F}_0)$).

Teorema 3.3.3 Si una prueba libre Φ de \mathfrak{F}_0 contra la hipótesis simple $\{P_1\}$ es cuasi-admisibile (resp. admisible), entonces Φ es U.M.P. (resp. estrictamente U.M.P.) como prueba de \mathfrak{F}_0 contra $\{\mathfrak{F}_1\}$.

Demostración Si Φ es libre, entonces para toda prueba Φ' se tiene:

$$\alpha_{\Phi'} \leq \alpha_{\Phi} \implies \beta_{\Phi'}(P) \leq \beta_{\Phi}(P) = \alpha_{\Phi}, \forall P \in \mathfrak{P}_0.$$

Como Φ es cuasi-admisibles se tiene que $\beta_{\Phi'}(P_1) \leq \beta_{\Phi}(P_1)$. Si Φ es admisible por definición es estrictamente U.M.P.

Teorema 3.3.4 *Sea Φ una prueba de \mathfrak{P}_0 contra \mathfrak{P}_1 tal que $\forall P_0 \in \mathfrak{P}_0, \forall P_1 \in \mathfrak{P}_1, \Phi$ es cuasi-admisibles como prueba de $\{P_0\}$ contra $\{P_1\}$, entonces es U.M.P. como prueba de \mathfrak{P}_0 contra \mathfrak{P}_1 .*

Demostración Sea $P_0 \in \mathfrak{P}_0, P_1 \in \mathfrak{P}_1$ y sea Φ cuasi-admisibles como prueba de $\{P_0\}$ contra $\{P_1\}$. Sea Φ' una prueba de \mathfrak{P}_0 contra \mathfrak{P}_1 , entonces se tiene que $\forall P_0 \in \mathfrak{P}_0, (\beta_{\Phi'}(P_0) \leq \beta_{\Phi}(P_0)) \implies \beta_{\Phi'}(P_1) \leq \beta_{\Phi}(P_1)$, pero la implicación es válida $\forall P_1 \in \mathfrak{P}_1$ i.e. Φ es U.M.P. como prueba de \mathfrak{P}_0 contra \mathfrak{P}_1 .

Definición 3.3.4 *La prueba Φ de \mathfrak{P}_0 contra \mathfrak{P}_1 es uniformemente la más potente entre las pruebas sin sesgo (U.M.P.I.), si para toda prueba Φ' insesgada de \mathfrak{P}_0 contra \mathfrak{P}_1 se tiene:*

$$(\beta_{\Phi'}(P) \leq \beta_{\Phi}(P), \forall P \in \mathfrak{P}_0) \implies (\beta_{\Phi'}(P) \leq \beta_{\Phi}(P), \forall P \in \mathfrak{P}_1).$$

Se dice que Φ es estrictamente U.M.P.I. si es U.M.P.I. y admisible.

Nota Es evidente que una prueba U.M.P. es U.M.P.I.

Una prueba U.M.P.I. es sin sesgo y cuasi-admisibles.

Una prueba Φ es U.M.P.I. como prueba de \mathfrak{P}_0 contra \mathfrak{P}_1 sii para toda ley $P_1 \in \mathfrak{P}_1, \beta_{\Phi}(P_1)$ es el máximo entre todas las pruebas sin sesgo de \mathfrak{P}_0 contra \mathfrak{P}_1 .

3.4 Determinación de pruebas óptimas

3.4.1 Caso de dos hipótesis simples

Se considera una estructura estadística $(\Omega, \mathfrak{A}, \mathfrak{P})$, $\mathfrak{P}_0 = \{P_0\}$ y $\mathfrak{P}_1 = \{P_1\}$; se tiene que $\{P_0, P_1\}$ está dominada siempre por una medida μ (por ejemplo $\mu = \frac{1}{2}(P_0 + P_1)$) i.e. se tienen las funciones de verosimilitud: $L(\cdot, P_0)$ y $L(\cdot, P_1)$.

Lema 3.4.1 Lema de Neyman-Pearson¹

Sea Φ una prueba definida sobre $(\Omega, \mathfrak{A}, \mathfrak{P})$ por:

$$\Phi(\omega) = \begin{cases} 1 & \text{si } L(\omega, P_1) > kL(\omega, P_0) \\ \pi & \text{si } L(\omega, P_1) = kL(\omega, P_0) \\ 0 & \text{si } L(\omega, P_1) < kL(\omega, P_0), \end{cases}$$

entonces se tiene:

- a) $\forall \alpha \in [0, 1]$ existe $k \geq 0$ y $\pi \in [0, 1]$ tales que Φ sea de nivel α .
- b) Si Ψ es una prueba de nivel inferior a α , entonces Ψ tiene una potencia inferior a la de Φ . (i.e. $\forall \Psi \in \mathfrak{T}, \beta_\Psi(P_0) \leq \beta_\Phi(P_0) \implies \beta_\Psi(P_1) \leq \beta_\Phi(P_1)$) es decir, Φ es una prueba de nivel α y U.M.P.

Prueba

- a) Verificar el caso $\alpha = 0, \alpha = 1$. Supongamos que $\alpha \in]0, 1[$; así tenemos que $\beta_\Phi(P_0) = \alpha \iff P_0(\{\omega \in \Omega / L(\omega, P_1) > kL(\omega, P_0)\}) + \pi P_0(\{\omega \in \Omega / L(\omega, P_1) = kL(\omega, P_0)\}) = \alpha$, i.e.

$$\alpha = P_0(\Phi = 1) + \pi P_0(\Phi = \pi) \quad (1)$$

Notemos que si P_0 es una medida de probabilidad, $C = \{\omega \in \Omega : L(\omega, P_0) > 0\}$ es tal que $P_0(C) = 1$ i.e. $\forall A \in \mathfrak{A}, P_0(A \cap C) = P_0(A)$.

$$\text{Así, } \alpha = P_0 \left(\left\{ \omega / \frac{L(\omega, P_1)}{L(\omega, P_0)} > k \right\} \right) + \pi P_0 \left(\left\{ \omega / \frac{L(\omega, P_1)}{L(\omega, P_0)} = k \right\} \right).$$

Sea Y la v.a.r. definida por $\omega \mapsto Y(\omega) = \frac{L(\omega, P_1)}{L(\omega, P_0)}$, entonces $Y \geq 0$ y Y es finita P_0 -c.p.d.

Sea G la función de repartición de Y : $G(0) = 0, G(+\infty) = 1$. La ecuación (1) nos dice:

$$\alpha = P_0(Y > k) + \pi P_0(Y = k) = 1 - G(k+0) + \pi(G(k+0) - G(k)). \quad (2)$$

– Supongamos que existe $k \in \mathbb{R}^+$ tal que $G(k+0) = 1 - \alpha$, entonces si $\pi = 0$, la ecuación (2) se satisface (si Y es continua respecto a la medida de Borel), entonces G es continua y existe un k

¹**Egon Shape Pearson (1895-1980)** Nace en Londres en 1895 y fue hijo del estadístico de Karl Pearson. Realizó sus estudios de matemática en el Trinity College de Cambridge. En 1921 entra al Departamento de Estadística del University College de Londres. Junto con Neyman y Gosset en 1924, trabajan en Teoría de Hipótesis. Cuando su padre deja el Departamento de Estadística Aplicada, asume la dirección y en 1936 asume la dirección de la Revista Biometrika. Se consagró igualmente a los problemas de Control de Calidad y de Investigación de Operaciones.

satisfaciendo la ecuación (2)).

– Supongamos que no existe k tal que $G(k+0) = 1 - \alpha$.

Sea $k = \sup\{h \in \mathbb{R}: G(h) < 1 - \alpha\}$ i.e. $G(k) < 1 - \alpha$; así $\forall \epsilon > 0, G(k + \epsilon) > 1 - \alpha$, de donde se tiene $G(k+0) \geq 1 - \alpha > G(k)$, pero $G(k+0) \neq 1 - \alpha$ y $G(k+0) > 1 - \alpha > G(k)$.

Sea ahora $\pi = \frac{G(k+0) - (1 - \alpha)}{G(k+0) - G(k)}$, entonces $\pi \in]0, 1[$ y la relación (2) se satisface.

b) Supongamos que $\Psi \in \mathfrak{T}: \beta_\Psi(P_0) \leq \alpha$. Así tenemos:

$$\beta_\Phi(P_1) - \beta_\Psi(P_1) = \int \Phi dP_1 - \int \Psi dP_1 = \int (\Phi - \Psi)L(\cdot, P_1) d\mu.$$

Sea $D_1 = \{\omega \in \Omega/L(\omega, P_1) > kL(\omega, P_0)\}$, $D_2 = \{\omega \in \Omega/L(\omega, P_1) = kL(\omega, P_0)\}$ y sea $D_3 = \{\omega \in \Omega/L(\omega, P_1) < kL(\omega, P_0)\}$, entonces:

$$\beta_\Phi(P_1) - \beta_\Psi(P_1) = \int_{D_1} (\Phi - \Psi)(L(\cdot, P_1) - kL(\cdot, P_0)) d\mu + \int_{D_2} (\Phi - \Psi)(L(\cdot, P_1) - kL(\cdot, P_0)) d\mu +$$

$$\int_{D_3} (\Phi - \Psi)(L(\cdot, P_1) - kL(\cdot, P_0)) d\mu + k(\int \Phi L(\cdot, P_0) d\mu - \int \Psi L(\cdot, P_0) d\mu).$$

Se verifica que todas las integrales del miembro de la derecha son positivas, de donde $\beta_\Phi(P_1) \geq \beta_\Psi(P_1)$.

– En la construcción de la prueba Ψ , si existe una estadística T suficiente, Ψ se expresa en función de la estadística T .

– Se puede demostrar que si $k > 0$, Ψ es admisible.

Ejemplo 1 Se considera la estructura estadística $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}, \{N(\mu, 1); \mu \in \mathbb{R}\}}^n)$ y sea $\mathfrak{P}_0 = \{N(\mu_0, 1)\}$, $\mathfrak{P}_1 = \{N(\mu_1, 1)\}$ y queremos probar \mathfrak{P}_0 contra \mathfrak{P}_1 . Se tiene que:

$$L(\mathbf{x}, \mu) = (\sqrt{2\pi})^{-n} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (x_i - \mu)^2 \right\}$$

por Neyman-Pearson, existe Φ prueba tal que:

$$\Phi(\mathbf{x}) = \begin{cases} 1 & \text{si } L(\mathbf{x}, P_1) > kL(\mathbf{x}, P_0) \\ \pi & \text{si } L(\mathbf{x}, P_1) = kL(\mathbf{x}, P_0) \\ 0 & \text{si } L(\mathbf{x}, P_1) < kL(\mathbf{x}, P_0), \end{cases}$$

con k y π escogidos para que Φ sea de nivel α . Así:

$$\begin{aligned}\Phi(\mathbf{x}) = 1 &\iff \frac{L(\mathbf{x}, \mu_1)}{L(\mathbf{x}, \mu_0)} = \exp\left\{-\frac{1}{2} \left(\sum_{i=1}^n (x_i - \mu_1)^2 - \sum_{i=1}^n (x_i - \mu_0)^2 \right)\right\} > k \\ &\iff \sum_{i=1}^n (x_i - \mu_1)^2 - \sum_{i=1}^n (x_i - \mu_0)^2 < k' \\ &\iff (\mu_0 - \mu_1) \sum_{i=1}^n (2x_i - (\mu_1 + \mu_0)) < k'' \\ &\iff (\mu_0 - \mu_1)(2\bar{x} - (\mu_1 + \mu_0)) < k'''.\end{aligned}$$

1er caso Si $\mu_1 > \mu_0$ se obtiene $\Phi(\mathbf{x}) = 1 \iff 2\bar{x} - (\mu_1 + \mu_0) > h \iff \bar{x} > c$, i.e. si Φ es de nivel α ($\beta_\Phi(P_0) = \alpha$).

$$\alpha = E_{P_0}(\Phi) = P_0(\Phi = 1) + \pi P_0(\Phi = \pi) + 0 \cdot P_0(\Phi = 0).$$

Sea \bar{X} la variable aleatoria tal que $\bar{X}(\mathbf{x}) = \bar{x}$, por lo tanto:

$$P_0(\Phi = \pi) = 0 = P_0(\bar{X} = c) \text{ i.e. } \alpha = P_0(\Phi = 1) = P_0(\bar{X} > c),$$

pero $\bar{X} \sim N\left(\mu, \frac{1}{n}\right)$, de donde:

$$P_0(\bar{X} > c) = P_0(\sqrt{n}(\bar{X} - \mu_0) > \sqrt{n}(c - \mu_0)).$$

Sea F la función de repartición de la normal $N(0, 1)$, i.e. $F(u) = \int_{-\infty}^u e^{-x^2/2} dx / \sqrt{2\pi}$, de donde:

$$\alpha = 1 - F(\sqrt{n}(c - \mu_0)) \iff c = \mu_0 + F^{-1}(1 - \alpha) / \sqrt{n}.$$

Así definimos:

$$\Phi_1(\mathbf{x}) = \begin{cases} 1 & \text{sii } \bar{x} > \mu_0 + F^{-1}(1 - \alpha) / \sqrt{n} \\ 0 & \text{sii } \bar{x} \leq \mu_0 + F^{-1}(1 - \alpha) / \sqrt{n}. \end{cases}$$

$\Phi_1(x)$ es una prueba U.M.P. de \mathfrak{F}_0 contra \mathfrak{F}_1 para $\mu_1 > \mu_0$.

2do caso Si $\mu_1 < \mu_0$, Φ_2 es una prueba tal que:

$$\Phi_2(\mathbf{x}) = 1 \iff \bar{x} < d \text{ y } \Phi_2(\mathbf{x}) = 0 \iff \bar{x} \geq d.$$

Si el nivel es $\alpha = P_0(\Phi_2 = 1) = P_0(\bar{X} < d) = F(\sqrt{n}(d - \mu_0))$ i.e.

$$\Phi_2(\mathbf{x}) = \begin{cases} 1 & \iff \bar{x} < \mu_0 + F^{-1}(\alpha) / \sqrt{n}, \\ 0 & \iff \bar{x} \geq \mu_0 + F^{-1}(\alpha) / \sqrt{n}. \end{cases}$$

Φ_2 es una prueba U.M.P. de \mathfrak{F}_0 contra \mathfrak{F}_1 para $\mu_1 < \mu_0$.

– Observemos que Φ_1 y Φ_2 dependen de la estadística suficiente \bar{X} .

Ejemplo 2 Se consideran las hipótesis $\mathfrak{F}_0 = \{P_{\theta_0}\}$, $\mathfrak{F}_1 = \{P_{\theta_1}\}$ dentro de la estructura estadística $(\{0, 1\}, \mathfrak{F}(\{0, 1\}), \{P_\theta: \theta \in [0, 1]\})^n$, con $P_\theta(\{1\}) = \theta$. La estructura está dominada por la medida de conteo i.e.

$$L(\mathbf{x}, \theta) = \theta^{\sum_{i=1}^n x_i} (1 - \theta)^{n - \sum_{i=1}^n x_i},$$

con $\mathbf{x} = (x_1, \dots, x_n) \in \{0, 1\}^n$, $\sum_{i=1}^n x_i = t = T(\mathbf{x})$.

Así, $\Phi(\mathbf{x}) = 1 \iff \left(\frac{\theta_1}{\theta_0} \frac{1 - \theta_0}{1 - \theta_1}\right)^t > k'$.

1er caso $\theta_1 > \theta_0$

$$\Phi(\mathbf{x}) = 1 \iff T(\mathbf{x}) = t > \ell$$

$$\Phi(\mathbf{x}) = \pi \iff T(\mathbf{x}) = \ell$$

$$\Phi(\mathbf{x}) = 0 \iff T(\mathbf{x}) < \ell.$$

Si $\ell \in \{0, 1, \dots, n\}$ no se puede dejar el caso $\Phi(\mathbf{x}) = \pi$.

Nivel es $\alpha = P_0(\Phi = 1) + \pi P_0(\Phi = \pi) = P_0(T > \ell) + \pi P_0(T = \ell)$, donde $T \sim B(n, \theta_0)$.

2do caso $\theta_1 < \theta_0$

$$\Phi(\mathbf{x}) = \begin{cases} 1 & \iff T(\mathbf{x}) < \ell' \\ \pi & \iff T(\mathbf{x}) = \ell' \\ 0 & \iff T(\mathbf{x}) > \ell', \end{cases}$$

donde ℓ' y π se deben determinar.

Ejemplo numérico $n = 8$, $\theta_0 = \frac{1}{2}$, $\theta_1 < \theta_0$, $\alpha = 0.05$. Los valores para $T = 0, 1, 2$ están dados por $P_0(T = 0) = \frac{1}{256}$, $P_0(T = 1) = \frac{8}{256}$, $P_0(T = 2) = \frac{28}{256}$.

Así debemos tener $P_0(T < \ell') + \pi P_0(T = \ell') = \alpha$.

Si $\ell' = 2$, $P_0(T < \ell') = \frac{9}{256} < 0.05 < P_0(T < \ell' + 1) = \frac{37}{256}$ i.e. $\frac{9}{256} + \pi \frac{28}{256} = 0.05 \iff \pi = 0.136 \dots$

Nota Como T es suficiente, la prueba depende de T .

3.4.2 \mathfrak{P}_0 simple, \mathfrak{P}_1 compuesta

Si tenemos una prueba Φ U.M.P. se tiene:

$$\forall \Psi \in \mathfrak{T}, \beta_{\Psi}(P_0) \leq \beta_{\Phi}(P_0) \iff (\beta_{\Psi}(P_1) \leq \beta_{\Phi}(P_1), \forall P_1 \in \mathfrak{P}_1).$$

Discusión de la existencia de una prueba U.M.P.

Si consideramos una prueba P_i fija: $P_i \in \mathfrak{P}_1$, se puede probar la hipótesis $\{P_0\}$ contra $\{P_i\}$ y por Neyman-Pearson encontrar la mejor prueba Φ_i . Igualmente para la hipótesis $\{P_0\}$ contra $\{P_j\}$ se encuentra Φ_j . Cuando Φ_i funciona para P_j tiene una potencia más pequeña que Φ_j pues esta prueba es U.M.P.

1er caso Existe una prueba Φ equivalente a todo Φ_i , ($i \in I$).

Φ_i es una prueba U.M.P. de \mathfrak{P}_0 contra $\{P_i\}$, $\mathfrak{P}_1 = \bigcup_{i \in I} \{P_i\}$ y por tanto Φ es una prueba U.M.P. de \mathfrak{P}_0 contra \mathfrak{P}_1 .

2do caso: En general no existe una prueba U.M.P. de \mathfrak{P}_0 contra \mathfrak{P}_1 .

Ejemplo 1 Si consideramos $\mathfrak{P}_0 = \{N(\mu_0, 1)\}$ y $\mathfrak{P}_1 = \{N(\mu, 1) : \mu > \mu_0\}$ la prueba Φ_1 no depende de μ_1 , $\forall \mu_1 > \mu_0$ i.e. Φ_1 es U.M.P. de \mathfrak{P}_0 contra \mathfrak{P}_1 .

– Verificar que $\beta_{\Phi_1}(\mu) = 1 - F(\sqrt{n}(\mu_0 - \mu) + F^{-1}(1 - \alpha))$ para $\mu > \mu_0$. Trazar la gráfica de la función potencia.

Ejemplo 2 Si $\mathfrak{P}_0 = \{N(\mu_0, 1)\}$ y $\mathfrak{P}_2 = \{N(\mu, 1)/\mu < \mu_0\}$, entonces Φ_2 es U.M.P. y $\beta_{\Phi_2}(\mu) = F(\sqrt{n}(\mu_0 - \mu) + F^{-1}(\alpha))$.

Ejemplo 3 Si $\mathfrak{P}_0 = \{N(\mu_0, 1)\}$ y $\mathfrak{P}_2 = \{N(\mu, 1)/\mu \neq \mu_0\}$ no existe una prueba U.M.P. Φ_1 no es “sin sesgo” como prueba de \mathfrak{P}_0 contra \mathfrak{P}_1 , pues para serlo debe ser superior al nivel.

Ejemplo 4 Existe una prueba Φ_3 de \mathfrak{P}_0 contra \mathfrak{P}_3 que es estrictamente U.M.P.I. de nivel α .

Teorema 3.4.1 Sean P_0, P_1, \dots, P_N leyes de probabilidad sobre (Ω, \mathfrak{A}) y sean f_0, f_1, \dots, f_N , sus densidades con respecto a una medida μ positiva; para todas las constantes no negativas $\lambda_0, \dots, \lambda_{N-1}$

(resp. positivas), toda prueba Φ tal que:

$$\begin{aligned} f_N(\omega) &> \sum_{j=0}^{N-1} \lambda_j f_j(\omega) \implies \Phi(\omega) = 1 \\ f_N(\omega) &\leq \sum_{j=0}^{N-1} \lambda_j f_j(\omega) \implies \Phi(\omega) = 0, \end{aligned}$$

es cuasi-admisibles (resp. admisible) como prueba de $\mathfrak{P}_0 = \{P_0, \dots, P_{N-1}\}$ contra $\mathfrak{P}_1 = \{P_N\}$.

Demostración Observemos que Φ es U.M.P. como prueba de \mathfrak{P}_0 contra \mathfrak{P}_1 . En efecto sea $\Psi \in \mathfrak{T}$ tal que $\beta_\Psi(P) \leq \beta_\Phi(P)$, $\forall P \in \mathfrak{P}_0$, entonces se tiene:

$$\begin{aligned} \beta_\Psi(P_j) &\leq \beta_\Phi(P_j) \text{ para } j = 0, 1, \dots, N-1, \text{ i.e.} \\ \beta_\Phi(P_N) - \beta_\Psi(P_N) &\geq \sum_{j=0}^{N-1} \lambda_j (\beta_\Phi(P_j) - \beta_\Psi(P_j)) \geq 0 \text{ i.e. } \beta_\Psi(P_N) \leq \beta_\Phi(P_N), \end{aligned}$$

por lo tanto Ψ es cuasi-admisibles. Supongamos ahora que las constantes λ_j son estrictamente positivas, entonces para toda prueba Ψ se tiene:

$$\beta_\Phi(P_N) - \beta_\Psi(P_N) \geq \sum_{j=0}^{N-1} \lambda_j (\beta_\Phi(P_j) - \beta_\Psi(P_j)),$$

por tanto si una de las desigualdades:

$$\beta_\Phi(P_j) \geq \beta_\Psi(P_j), \quad j = 0, \dots, N-1,$$

es estricta entonces $\beta_\Phi(P_N) > \beta_\Psi(P_N)$, lo que demuestra que Φ es admisible.

Generalización del teorema de Neyman-Pearson

Sea (Ω, \mathfrak{A}) un espacio probabilizable, μ una medida y sean h, g_1, g_2, \dots, g_m funciones reales μ -integrables sobre (Ω, \mathfrak{A}) .

Sea $\Phi: (\Omega, \mathfrak{A}) \rightarrow ([0, 1], \mathfrak{B}_{[0,1]})$ la aplicación tal que:

$$\begin{aligned} \Phi(\omega) &= 1 \quad \text{si} \quad h(\omega) > \sum_{i=1}^m k_i g_i(\omega) \\ \Phi(\omega) &= 0 \quad \text{si} \quad h(\omega) < \sum_{i=1}^m k_i g_i(\omega), \end{aligned}$$

donde los k_i constantes no negativas, entonces para toda función Ψ medible $\Psi: \Omega \rightarrow [0, 1]$ tal que $\int \Psi g_j d\mu \leq \int \Phi g_j d\mu$, $j = 1, \dots, m$, entonces $\int \Psi h d\mu \leq \int \Phi h d\mu$.

Demostración Sabemos que $(\Phi - \Psi)(h - \sum_{i=1}^n k_i g_i) \geq 0$, por lo que:

$$\int \Phi h d\mu - \int \Psi h d\mu - \int (\Phi - \Psi) \sum_{i=1}^n k_i g_i d\mu \geq 0,$$

luego $\int \Phi h d\mu - \int \Psi h d\mu \geq \sum_{i=1}^m k_i (\int \Phi g_i d\mu - \int \Psi g_i d\mu) \geq 0$.

3.5 Métodos en ausencia de pruebas óptimas

3.5.1 Restricción de la clase de pruebas consideradas

Sobre la estructura $(\Omega, \mathfrak{A}, \mathfrak{P})$ se puede limitar a pruebas \mathfrak{B} -medibles, con \mathfrak{B} sub-tribu de \mathfrak{A} . Esto se traduce en cambiar la estructura y considerar la estructura $(\Omega, \mathfrak{B}, \mathfrak{P})$. Un ejemplo de éste método es el estudio de pruebas homogéneas en el problema de Behrens-Fisher.

3.5.2 Prueba del cociente de verosimilitud

Sea $L(\omega, \theta)$ la función de verosimilitud de la estructura y sea Φ una prueba de la hipótesis $\mathfrak{P}_0 = \{P_\theta: \theta \in \Theta_0\}$ contra $\mathfrak{P}_1 = \{P_\theta: \theta \in \Theta_1\}$, se dice que Φ es la prueba del cociente de verosimilitud si existe una constante λ tal que:

$$\frac{\sup\{L(\omega, \theta)/\theta \in \Theta_0\}}{\sup\{L(\omega, \theta)/\theta \in \Theta\}} \geq \lambda \implies \Phi = 0$$

$$\frac{\sup\{L(\omega, \theta)/\theta \in \Theta_0\}}{\sup\{L(\omega, \theta)/\theta \in \Theta\}} < \lambda \implies \Phi = 1,$$

donde $\Theta = \Theta_0 \cup \Theta_1$. Si bien el principio del cociente de verosimilitud no tiene ninguna propiedad optimal bien definida, nos conduce a procedimientos satisfactorios en numerosos casos particulares.

En el caso de una muestra empírica $(\Omega, \mathfrak{A}, \mathfrak{P})^n$ un estudio asintótico es a menudo útil. Sean \mathfrak{P}_0 y \mathfrak{P}_1 dos hipótesis, se dice que la sucesión de pruebas Φ_n de \mathfrak{P}_0 contra \mathfrak{P}_1 es consistente, si cuando $n \rightarrow \infty, \beta_{\Phi_n}(P) \rightarrow 0, \forall P \in \mathfrak{P}_0$ y $\beta_{\Phi_n}(P) \rightarrow 1, \forall P \in \mathfrak{P}_1$.

Ejemplo 1 Sea X una variable aleatoria normal tal que $X \sim N(\theta_1, \theta_2)$ y sea la estructura estadística $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}, \{N(\theta_1, \theta_2): (\theta_1, \theta_2) \in \Theta\}})^n$, con $\Theta = \{(\theta_1, \theta_2) / -\infty < \theta_1 < +\infty, 0 < \theta_2 < +\infty\}$. Consideramos la hipótesis: $H_0: \theta_1 = 0, \theta_2 > 0$ y $H_1: \theta_1 \neq 0, \theta_2 > 0$.

Nos interesamos en hacer una prueba de H_0 contra H_1 .

Sean X_1, \dots, X_n observaciones, con $n > 1$ de misma ley que X . La función de verosimilitud es:

$$L(\mathbf{x}, \theta_1, \theta_2) = \left(\frac{1}{2\pi\theta_2}\right)^{\frac{n}{2}} e^{-\frac{1}{2\theta_2} \sum_{i=1}^n (x_i - \theta_1)^2} = L(\mathbf{x}, \theta), \quad \theta \in \Theta, \quad \mathbf{x} \in \mathbb{R}^n.$$

Sea $(\theta_1, \theta_2) = (0, \theta_2) \in H_0$, así:

$$L(\mathbf{x}, 0, \theta_2) = \left(\frac{1}{2\pi\theta_2}\right)^{\frac{n}{2}} e^{-\frac{1}{2\theta_2} \sum_{i=1}^n x_i^2} = L(\mathbf{x}, \theta_0), \quad \theta_0 \in H_0, \quad \mathbf{x} \in \mathbb{R}^n.$$

Tenemos que determinar el máximo de $L(\mathbf{x}, \theta)$, con $\theta \in \Theta$ y con $\theta \in \Theta_0 = H_0$. Denotamos estos máximos por $L(\mathbf{x}, \hat{\theta})$ y $L(\mathbf{x}, \hat{\theta}_0)$ respectivamente. El criterio del cociente nos determina un valor $\Lambda(x_1, \dots, x_n) = \Lambda = \frac{L(\mathbf{x}, \hat{\theta}_0)}{L(\mathbf{x}, \hat{\theta})}$. Así, $\frac{\partial}{\partial \theta_2} \log L(\mathbf{x}, \theta_0) = -\frac{n}{2\theta_2} + \frac{1}{2\theta_2} \sum_{i=1}^n x_i^2 = 0$, tiene por solución $\hat{\theta}_2 = \frac{1}{n} \sum_{i=1}^n x_i^2$ y este número maximiza $L(\mathbf{x}, \theta_0)$. De este modo:

$$L(\mathbf{x}, \hat{\theta}_0) = \left(\frac{1}{2\pi \sum_{i=1}^n x_i^2 / n}\right)^{\frac{n}{2}} e^{-\frac{\sum_{i=1}^n x_i^2}{2 \sum_{i=1}^n x_i^2 / n}} = \left(\frac{ne^{-1}}{2\pi \sum_{i=1}^n x_i^2}\right)^{\frac{n}{2}}.$$

Sabemos que el máximo de $L(\mathbf{x}, \theta)$ se obtiene reemplazando θ_1 y θ_2 por \bar{x} y $\sum_{i=1}^n (x_i - \bar{x})^2 / n$, es decir:

$$L(\mathbf{x}, \hat{\theta}) = \left[\frac{1}{2\pi \sum_{i=1}^n (x_i - \bar{x})^2 / n}\right]^{\frac{n}{2}} e^{-\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{2 \sum_{i=1}^n (x_i - \bar{x})^2 / n}} = \left[\frac{ne^{-1}}{2\pi \sum_{i=1}^n (x_i - \bar{x})^2}\right]^{\frac{n}{2}},$$

o sea

$$\Lambda = \left[\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sum_{i=1}^n x_i^2}\right]^{\frac{n}{2}}.$$

Como $\sum_{i=1}^n x_i^2 = \sum_{i=1}^n (x_i - \bar{x})^2 + n\bar{x}^2$, tenemos así que se rechaza H_0 si $0 \leq \Lambda \leq \lambda_0$ y la prueba se define por:

$$\Phi = \begin{cases} 1 & \text{si } \Lambda < \lambda_0 \\ 0 & \text{si } \Lambda \geq \lambda_0. \end{cases}$$

De este modo, $\frac{\sqrt{n}|\bar{x}|}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 / (n-1)}} \geq \sqrt{(n-1)(\lambda_0^{-2/n} - 1)} = c$ i.e. si H_0 es cierta, la estadística de

la prueba sigue un t_{n-1} de Student con $n-1$ g.d.l.

Ejemplo 2 Sean X y Y variables aleatorias independientes de distribuciones respectivas $N(\theta_1, \sigma^2)$ y $N(\theta_2, \sigma^2)$, de varianza común σ^2 desconocida. Así:

$$\Theta = \{(\theta_1, \theta_2, \sigma^2) / -\infty < \theta_1 < \infty, -\infty < \theta_2 < \infty, 0 < \sigma < \infty\}.$$

Sean X_1, \dots, X_n y Y_1, \dots, Y_m dos muestras aleatorias independientes de estas distribuciones. Se considera $H_0: \theta_1 = \theta_2, \sigma > 0$ y $H_1: \theta_1 \neq \theta_2, \sigma > 0$. Así tenemos $n + m > 2$ variables aleatorias independientes i.e. su función de verosimilitud es:

$$L(x_1, \dots, x_n, y_1, \dots, y_m, \theta_1, \theta_2, \sigma) = \frac{1}{(2\pi\sigma^2)^{(n+m)/2}} e^{-\frac{1}{2\sigma^2} \left(\sum_{i=1}^n (x_i - \theta_1)^2 + \sum_{i=1}^m (y_i - \theta_2)^2 \right)}.$$

Si $\frac{\partial \ln L(\cdot, \theta_1, \sigma)}{\partial \theta_1} = 0$ y $\frac{\partial \ln L(\cdot, \theta_1, \sigma)}{\partial \sigma^2} = 0$, bajo H_0 tenemos:

$$\begin{aligned} \sum_{i=1}^n (x_i - \theta_1) + \sum_{i=1}^m (y_i - \theta_1) &= 0 \\ -(n+m) + \frac{1}{\sigma^2} \left(\sum_{i=1}^n (x_i - \theta_1)^2 + \sum_{i=1}^m (y_i - \theta_1)^2 \right) &= 0. \end{aligned}$$

Las soluciones para θ_1 y σ^2 son respectivamente:

$$u = \frac{n\bar{x} + m\bar{y}}{n+m}, \quad \omega = \frac{\sum_{i=1}^n (x_i - u)^2 + \sum_{i=1}^m (y_i - u)^2}{n+m}$$

y (u, ω) maximiza la función de verosimilitud bajo H_0 . El máximo es:

$$L(\cdot, \hat{\theta}_0) = \left(\frac{e^{-1}}{2\pi\omega} \right)^{(n+m)/2}.$$

De igual manera si consideramos las derivadas de $\ln L(\cdot, \theta)$, con respecto a $\theta_1, \theta_2, \sigma^2$ y las igualamos

a cero, entonces las soluciones son respectivamente $u_1 = \bar{x}, u_2 = \bar{y}, \omega' = \frac{\sum_{i=1}^n (x_i - u_1)^2 + \sum_{i=1}^m (y_i - u_2)^2}{n+m}$,

o sea u_1, u_2 y ω' maximiza $L(\cdot, \theta)$. Así el máximo sobre Θ es:

$$L(\cdot, \hat{\theta}) = \left(\frac{e^{-1}}{2\pi\omega'} \right)^{(n+m)/2},$$

i.e. $\Lambda(x_1, \dots, x_n, y_1, \dots, y_m) = \Lambda = \left(\frac{\omega'}{\omega}\right)^{(n+m)/2}$. La variable aleatoria $\Lambda^{2/(n+m)}$ es:

$$\frac{\sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{i=1}^m (Y_i - \bar{Y})^2}{\sum_{i=1}^n \left(X_i - \frac{n\bar{X} + m\bar{Y}}{n+m}\right)^2 + \sum_{i=1}^m \left(Y_i - \frac{n\bar{X} + m\bar{Y}}{n+m}\right)^2} = \frac{1}{1 + \frac{\frac{nm}{n+m}(\bar{X} - \bar{Y})^2}{\sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{i=1}^m (Y_i - \bar{Y})^2}}.$$

Si la hipótesis $H_0: \theta_1 = \theta_2, \sigma > 0$ es cierta, la variable aleatoria:

$$T = \frac{\sqrt{\frac{nm}{n+m}}(\bar{X} - \bar{Y})}{\sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{i=1}^m (Y_i - \bar{Y})^2}{n+m-2}}},$$

es un t de Student con $n+m-2$ g.d.l. Así la variable aleatoria:

$$\Lambda^{2/(n+m)} = \frac{n+m-2}{(n+m-2) + T^2},$$

es decir, la prueba de H_0 contra H_1 está en función de la estadística t_{n+m-2} . Se rechaza H_0 si $0 \leq \Lambda \leq \lambda_0 < 1$ y su nivel α es $\alpha = P_{H_0}(\Lambda \leq \lambda_0)$. Pero $\Lambda \leq \lambda_0$ es equivalente a que $|T| \geq c$ i.e. $\alpha = P_{H_0}(|T| \geq c)$.

Si $n = 10, m = 6$, y $\alpha = 0.05$, entonces $c = 2.145$

La potencia $\beta_\Phi(\theta)$, con $\theta \in H_1$ (i.e. $\theta_1 \neq \theta_2$), está en función de un t de Student descentrada con $n+m-2$ g.d.l. con parámetro de descentrage $\delta = \sqrt{nm/(n+m)}(\theta_1 - \theta_2)/\sigma$, pues:

$$Z = \sqrt{\frac{nm}{n+m}}(\bar{X} - \bar{Y})/\sigma \sim N\left(\sqrt{\frac{nm}{n+m}}(\theta_1 - \theta_2)/\sigma, 1\right)$$

$$V = \left(\sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{i=1}^m (Y_i - \bar{Y})^2\right)/\sigma^2 \sim \chi_{n+m-2}^2.$$

Ejemplo 3 Prueba de independencia de dos poblaciones normales

Sea (X, Y) una variable aleatoria normal $N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\right)$.

Las variables son independientes si $\rho = 0$, i.e. si hacemos la hipótesis $H_0: \rho = 0$ contra $H_1: \rho \neq 0$, el criterio del cociente de verosimilitud puede utilizarse.

Sea $\Theta = \{(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho) / -\infty < \mu_i < +\infty, 0 < \sigma_i < +\infty, i = 1, 2, -1 < \rho < 1\}$, es decir $H_0 = \{(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho) / \rho = 0, -\infty < \mu_i < \infty, 0 < \sigma_i < +\infty, i = 1, 2\}$.

Sea $(X_1, Y_1), \dots, (X_n, Y_n)$ una muestra de tamaño $n > 2$; la función de verosimilitud es:

$$L(\cdot, \theta) = \left(\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \right)^n e^{-\frac{1}{2}\sum_{i=1}^n q_i},$$

$$\text{donde } q_i = \frac{1}{1-\rho^2} \left[\left(\frac{x_i - \mu_1}{\sigma_1} \right)^2 - 2\rho \left(\frac{x_i - \mu_1}{\sigma_1} \right) \left(\frac{y_i - \mu_2}{\sigma_2} \right) + \left(\frac{y_i - \mu_2}{\sigma_2} \right)^2 \right].$$

La función de verosimilitud bajo H_0 se determina haciendo $\rho = 0$ en la función anterior.

Si definimos:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i, v_1 = \sum_{i=1}^n (x_i - \bar{x})^2, v_2 = \sum_{i=1}^n (y_i - \bar{y})^2, v = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}),$$

la función de verosimilitud bajo H_0 se maximiza sustituyendo $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ por $\bar{x}, \bar{y}, v_1/n, v_2/n$ respectivamente. Este máximo es:

$$L(\cdot, \hat{\theta}_0) = \left(\frac{ne^{-1}}{2\pi\sqrt{v_1v_2}} \right)^n.$$

Similarmente si sustituimos $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ y ρ por $\bar{x}, \bar{y}, v_1/n, v_2/n$ y $v/\sqrt{v_1v_2}$ respectivamente, el máximo es:

$$L(\cdot, \hat{\theta}) = \left(\frac{ne^{-1}}{2\pi\sqrt{v_1v_2}\sqrt{1-\frac{v^2}{v_1v_2}}} \right)^n.$$

Así se tiene que $\Lambda = \left(1 - \frac{v^2}{v_1v_2}\right)^{n/2}$. La estadística definida por:

$$R = \frac{V}{\sqrt{V_1V_2}} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}},$$

es llamada el coeficiente de correlación de la muestra (X_i, Y_i) . Como $0 \leq \Lambda \leq 1$, se tiene que $0 \leq R^2 \leq 1$, o sea $-1 \leq R \leq 1$. Se rechaza H_0 si $\Lambda \leq \lambda_0$ y el nivel de la prueba es:

$$\alpha = P_{H_0}((1 - R^2)^{n/2} \leq \lambda_0) = P_{H_0}(|R| \geq c),$$

donde $c = \sqrt{1 - \lambda_0^{2/n}}$. Para determinar el valor de c es necesario determinar la distribución de R bajo

H_0 . Sea $X_1 = x_1, \dots, X_n = x_n$, con $n > 2$, donde x_1, \dots, x_n y \bar{x} son números fijos.

Dado que Y_1, \dots, Y_n son independientes y como $\rho = 0$, también son independientes de X_1, \dots, X_n

i.e. la densidad es:

$$f(y_1, \dots, y_n) = (\sqrt{2\pi}\sigma_2)^{-n} e^{-\frac{1}{2\sigma_2^2} \sum_{i=1}^n (y_i - \mu_2)^2}.$$

La distribución condicionada de $\frac{1}{\sigma_2^2} \sum_{i=1}^n (Y_i - \bar{Y})^2$ a $X_1 = x_1, \dots, X_n = x_n$ es un χ_{n-1}^2 . Además la distribución condicional de W con:

$$W = \frac{\sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}} = \frac{\sum_{i=1}^n (x_i - \bar{x})Y_i}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}} \sim N(0, \sigma_2^2).$$

La distribución de W^2/σ_2^2 condicionada a $X_1 = x_1, \dots, X_n = x_n$ es un χ_1^2 y además:

$$\frac{1}{\sigma_2^2} \sum_{i=1}^n (Y_i - \bar{Y})^2 = \frac{W^2}{\sigma_2^2} + \frac{1}{\sigma_2^2} \sum_{i=1}^n \left(Y_i - \bar{Y} - \frac{\sum_{k=1}^n (x_k - \bar{x})(Y_k - \bar{Y})}{\sum_{k=1}^n (x_k - \bar{x})^2} (x_i - \bar{x}) \right)^2 = \frac{W^2}{\sigma_2^2} + \frac{U^2}{\sigma_2^2},$$

pero $\frac{1}{\sigma_2^2} U^2$ es independiente de W y sigue un χ_{n-2}^2 . Así se tiene:

$$\frac{W/\sigma_2}{\sqrt{\frac{U}{\sigma_2^2(n-2)}}} = \frac{W\sqrt{n-2}}{\sqrt{U}} \sim t_{n-2},$$

i.e. es un t de Student con $n-2$ g.d.l. condicionado a $X_1 = x_1, \dots, X_n = x_n$. Sea:

$$R_x = \frac{\sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}},$$

entonces tenemos:

$$\frac{W\sqrt{n-2}}{\sqrt{U}} = \frac{R_x\sqrt{n-2}}{\sqrt{1-R_x^2}}.$$

Notemos que la densidad $g(t)$ de esta distribución t no depende de x_1, \dots, x_n . Sea $Z = R\sqrt{n-2}/\sqrt{1-R^2}$,

donde:

$$R = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}}.$$

Así, $f(z, x_1, \dots, x_n) = f(z/x_1, \dots, x_n)f(x_1, \dots, x_n) = g(t)f(x_1, \dots, x_n)$, o sea:

$$f(z) = \int_{\mathbb{R}^n} f(z, x_1, \dots, x_n) dx_1 \cdots dx_n = g(t),$$

donde $f(z)$ es la densidad de $R \sqrt{n-2} / \sqrt{1-R^2}$.

Cuando $\rho = 0$ tenemos que R es independiente de X_1, \dots, X_n . Repitiendo el proceso de condicionamiento a Y_1, \dots, Y_n , obtenemos que R es independiente de Y_1, \dots, Y_n .

Como $Z = R \sqrt{n-2} / \sqrt{1-R^2} \sim t_{n-2}$ haciendo un cambio de variable tenemos que la función de densidad de R está dada por:

$$f(r) = \frac{\Gamma(\frac{1}{2}(n-1))}{\Gamma(\frac{1}{2})\Gamma(\frac{1}{2}(n-2))} (1-r^2)^{(n-4)/2}, \quad -1 \leq r \leq 1.$$

El criterio del cociente de verosimilitud de la hipótesis $H_0: \rho = 0$ contra la hipótesis $H_1: \rho \neq 0$ se puede basar sobre la estadística R o sobre la estadística $Z = R \sqrt{n-2} / \sqrt{1-R^2}$. En cualquier caso el nivel:

$$\alpha = P_{H_0}(|R| \geq c_1) = P_{H_0}(|Z| \geq c_2).$$

Para obtener la potencia $\beta_\theta(P)$ es necesario la distribución de R para $\rho \neq 0$. Sea $\theta = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ y sea la estadística suficiente $T = (\bar{X}, \bar{Y}, V_1, V_2, R)$, entonces la función de densidad de T es (utilizando el teorema de factorización $L(\omega, \theta) = g_\theta(t)h(\omega)$).

$$g_\theta(t) = g_\theta(\bar{x}, \bar{y}, v_1, v_2, r). \quad (1)$$

Para obtener la función de densidad de R es necesario integrar (1) con respecto a $\bar{x}, \bar{y}, v_1, v_2$ i.e.

$$f(r) = \frac{(1-\rho^2)^{(n-1)/2} (1-r^2)^{\frac{1}{2}(n-4)}}{\sqrt{\pi} \Gamma(\frac{1}{2}(n-1)) \Gamma(\frac{1}{2}(n-2))} \sum_{k=0}^{\infty} \frac{(2\rho r)^k}{k!} \Gamma^2\left(\frac{n-1+k}{2}\right).$$

Nota Si definimos $Z = \frac{1}{2} \log \frac{1+R}{1-R}$ y $\gamma = \frac{1}{2} \log \frac{1+\rho}{1-\rho}$ para valores pequeños de n , la ley de Z está próxima a la ley $N\left(\gamma + \frac{\rho}{2(n-1)}, \frac{1}{n-3}\right)$.

Un resultado parecido fue obtenido por Fisher (1925) quien demuestra que R se distribuye asintóticamente como una distribución $N\left(\rho, \frac{1}{n}(1-\rho^2)^2\right)$ y además $\frac{1}{2} \log \left(\frac{1+R}{1-R}\right)$ para n grande, se distribuye como $N\left(\gamma, \frac{1}{n}\right)$.

3.6 Indicaciones sobre la búsqueda de una prueba U.M.P.I.

Se supone que \mathfrak{P} está indexada por $\theta \in \Theta \subset \mathbb{R}$ (se supone que Θ es un intervalo). Sea $\mathfrak{P}_0 = \{P_{\theta_0}\}$, con $\theta_0 \in \overset{\circ}{\Theta}$ y $\mathfrak{P}_1 = \{P_\theta: \theta \in \Theta \setminus \{\theta_0\}\}$. En este caso se dice que la prueba es bilateral. Se desea encontrar Φ U.M.P.I. de nivel $\alpha = \int \Phi dP_{\theta_0}$.

Si \mathfrak{P} está dominada por una medida μ σ -finita, entonces denotamos $L(\cdot, \theta)$ la densidad de P_θ , con respecto a μ y tendremos:

$$(1) \quad \begin{cases} \int \Phi L(\cdot, \theta_0) d\mu = \alpha \\ \beta_\Phi(P_\theta) \geq \alpha, \forall P_\theta \in \mathfrak{P}_1. \end{cases}$$

Suponemos que $\beta_\Phi(P_\theta) = \beta_\Phi(\theta)$ es derivable y como Φ es sin sesgo y $\theta \in \overset{\circ}{\Theta}$ implica que $\beta'_\Phi(\theta_0) = 0$.

La condición sin sesgo se escribe $\beta'_\Phi(\theta_0) = 0$ y si se puede derivar bajo el signo de integral:

$$(2) \quad \beta'_\Phi(\theta_0) = \int \Phi \left(\frac{dL}{d\theta} \right)_{\theta_0} d\mu = \int \Phi \left(\frac{d \log L}{d\theta} \right)_{\theta_0} L(\cdot, \theta_0) d\mu = 0$$

Dado que se toma $\theta \in \Theta \setminus \{\theta_0\}$, para tener una prueba la más potente es necesario que:

$$(3) \quad \int \Phi L(\cdot, \theta) d\mu \text{ sea máximo, } \forall \theta \in \Theta \setminus \{\theta_0\}, \Phi \text{ fijo,}$$

entonces el problema viene a ser, encontrar Φ tal que la integral (3) sea máxima con las restricciones

(1) y (2).

Así después de la generalización de Neyman-Pearson para $m = 2$:

$$g_1 = L(\cdot, \theta_0), \quad g_2 = \left(\frac{dL}{d\theta} \right)_{\theta_0} = \left(\frac{d \log L}{d\theta} \right)_{\theta_0} L(\cdot, \theta_0), \quad h = L(\cdot, \theta),$$

se puede tomar:

$$\begin{aligned} \Phi(\omega) &= 1 && \text{si } L(\omega, \theta) > k_1 L(\omega, \theta_0) + k_2 L(\omega, \theta_0) \left(\frac{d \log L}{d\theta} \right)_{\theta_0} \\ \Phi(\omega) &= 0 && \text{si } L(\omega, \theta) < k_1 L(\omega, \theta_0) + k_2 L(\omega, \theta_0) \left(\frac{d \log L}{d\theta} \right)_{\theta_0}, \end{aligned}$$

o sea:

$$\begin{aligned} \Phi(\omega) &= 1 && \text{si } \frac{L(\omega, \theta)}{L(\omega, \theta_0)} > k_1 + k_2 \left(\frac{d \log L}{d\theta} \right)_{\theta_0} \\ \Phi(\omega) &= 0 && \text{si } \frac{L(\omega, \theta)}{L(\omega, \theta_0)} < k_1 + k_2 \left(\frac{d \log L}{d\theta} \right)_{\theta_0}. \end{aligned}$$

Nota Si se encuentra Φ que responde a éste desarrollo, no se está seguro de tener la prueba U.M.P. entre las pruebas sin sesgo; pero si es, debe verificar las condiciones.

Ejemplo Sea la estructura estadística $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}, \{N(\mu, 1)/\mu \in \mathbb{R}\}})^n$. Se desea probar:

$$\mathfrak{P}_0 = \{N(\mu_0, 1)\} \text{ y } \mathfrak{P}_3 = \{N(\mu, 1)/\mu \neq \mu_0\}.$$

La estructura está dominada por la medida de Borel y se tiene la función de verosimilitud $L(\mathbf{x}, \mu) = (\sqrt{2\pi})^{-n} e^{-\frac{1}{2} \sum_{i=1}^n (x_i - \mu)^2}$, $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$, $\frac{d \ln L}{d\mu} = n(\bar{x} - \mu)$, con $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$, $\frac{L(\mathbf{x}, \mu)}{L(\mathbf{x}, \mu_0)} = k e^{(\mu - \mu_0)n\bar{x}}$, $k > 0$, de donde:

$$\Phi(\mathbf{x}) = 1 \text{ si } k e^{(\mu - \mu_0)n\bar{x}} > k_1 + k_2 n(\bar{x} - \mu_0) \iff e^{(\mu - \mu_0)n\bar{x}} > k'_1 + k'_2 \bar{x}.$$

$$\boxed{\mu > \mu_0}$$

1er caso: Si $k'_2 < 0$, entonces $\Phi(\mathbf{x}) = 1$, si $\bar{x} > c'$, pero la prueba no es sin sesgo.

2do caso: Si $k'_2 > 0$, entonces $\Phi(\mathbf{x}) = 1$, si $\bar{x} < c$ o $\bar{x} > d$, $c < d$.

Si $\mu < \mu_0$, las condiciones son las mismas y en definitiva tenemos:

$$\Phi(\mathbf{x}) = 1 \quad \text{si } \bar{x} < c \text{ o } \bar{x} > d, \quad c < d,$$

$$\Phi(\mathbf{x}) = 0 \quad \text{si } c \leq \bar{x} \leq d.$$

La prueba Φ es determinista. Es necesario determinar c y d que satisfagan (1) y (2). Así tenemos que $\beta_\Phi(\mu) = P_\mu(\bar{X} > d \text{ o } \bar{X} < c)$.

Nivel $\alpha = P_{\mu_0}(\sqrt{n}(\bar{X} - \mu_0) < \sqrt{n}(c - \mu_0)) \text{ o } \sqrt{n}(\bar{X} - \mu_0) > \sqrt{n}(d - \mu_0))$, entonces:

$$\beta_\Phi(\mu) = F(\sqrt{n}(c - \mu)) + 1 - F(\sqrt{n}(d - \mu))$$

$$\beta'_\Phi(\mu_0) = 0 \iff f(\sqrt{n}(c - \mu_0)) = f(\sqrt{n}(d - \mu_0)) \iff c - \mu_0 = -(d - \mu_0),$$

es decir $\alpha = 2F(\sqrt{n}(d - \mu_0)) \iff \sqrt{n}(d - \mu_0) = F^{-1}\left(\frac{\alpha}{2}\right)$.

Ejemplo Si $\alpha = 0.05$, $n = 100$, $\mu_0 = 0$, se tiene $F^{-1}\left(\frac{\alpha}{2}\right) = 1.96$, por lo que:

$$c = -0.196 \text{ y } d = 0.196$$

Observemos que Φ es sin sesgo. Estudie la función potencia.

3.6.1 Indicaciones en el caso \mathfrak{P}_0 compuesta

Veamos algunos ejemplos.

Ejemplo 1 $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}, \mathfrak{P}})^n$, con $\mathfrak{P} = \{N(\mu, 1) / \mu \in \mathbb{R}\}$, $\mathfrak{P}_0 = \{N(\mu, 1) : \mu \leq \mu_0\}$, $\mathfrak{P}_1 : \mu > \mu_0$.

Ejemplo 2 $\mathfrak{P} = \{N(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma > 0\}$, $\mathfrak{P}_0 = \{N(\mu_0, \sigma^2) : \sigma > 0\}$, $\mathfrak{P}_1 = \{N(\mu, \sigma^2) : \mu \neq \mu_0, \sigma > 0\}$.

En ambos ejemplos \mathfrak{P}_0 es compuesta.

Si Φ es una prueba, $\beta_\Phi(P)$ en general no es constante para $P \in \mathfrak{P}_0$ y si se define el nivel $\alpha =$

$\sup_{P \in \mathfrak{P}_0} \beta_\Phi(P)$ es difícil de manejar.

Nos podríamos preguntar si es posible restringirnos a pruebas Φ tales que $\beta_\Phi(P) = k$ (constante) $\forall P \in \mathfrak{P}_0$. Para el ejemplo 1 esto no es posible; sin embargo en el ejemplo 2 sí es posible hacerlo. En este ejemplo al parámetro σ^2 se le llama parámetro fantasma.

Recordemos que si se tiene una prueba Φ tal que $\beta_\Phi(P) = \text{cte}$ sobre \mathfrak{P}_0 , Φ es una *prueba libre* sobre \mathfrak{P}_0 .

3.6.2 Pruebas libres deducidas de pruebas condicionadas (Estructuras de Neyman)

Si se considera una estructura del tipo $(\Omega, \mathcal{A}, \{P_\theta : \theta \in \Theta\})$ y se considera $\Theta = \Theta_1 \times \Theta_2$ i.e. $\theta \in \Theta \iff \theta = (\theta_1, \theta_2)$.

Si tenemos $\mathfrak{P}_0 = \{P_\theta : \theta_1 = \theta_1^0, \theta_2 \in \Theta_2\}$, $\mathfrak{P}_1 = \{P_\theta : \theta_1 \neq \theta_1^0, \theta_2 \in \Theta_2\}$ y que además existe una estadística T con valores en $(\mathfrak{T}, \mathfrak{C})$ que sea suficiente para la estructura $(\Omega, \mathfrak{A}, \mathfrak{P}_0)$, la ley $P \in \mathfrak{P}_0$ condicionada a T no depende de θ_2 . Por tanto, se considera la estructura $(\Omega, \mathfrak{A}, \mathfrak{P}^{T=t})$.

La hipótesis \mathfrak{P}_0 es entonces $\mathfrak{P}_0^{T=t} = \{P_\theta^{T=t} : \theta_1 = \theta_1^0, \theta_2 \in \Theta_2\}$, la cual es una *hipótesis simple* que no depende de θ_2 . Sobre esta estructura se puede fabricar una prueba (condicionada) Φ_t y que se lleva al problema anterior para encontrar la prueba óptima.

Decir que la prueba es de nivel α es decir que $E_{\theta_1 = \theta_1^0}^{T=t}(\Phi_t) = \alpha$.

Así la prueba Φ de \mathfrak{P}_0 contra \mathfrak{P}_1 tal que $\Phi(\omega) = \Phi_{T(\omega)}(\omega) = \Phi_t(\omega)$, es una prueba libre de nivel α .

En efecto:

$$E_{\theta_1^0, \theta_2}(\Phi) = E_{\theta_1^0, \theta_2}(E_{\theta_1^0, \theta_2}^T(\Phi)) = E_{\theta_1^0, \theta_2}(E_{\theta_1^0, \theta_2}^{T=t}(\Phi_t)) = \alpha,$$

i.e. $E_{\theta_1^0, \theta_2}^T(\Phi)$ no depende de θ_2 .

3.7 Indicaciones sobre los problemas de estimación puntual

3.7.1 Desigualdad de Cramer-Rao

Vamos a demostrar que bajo ciertas condiciones un estimador de un parámetro real, no puede tener varianza menor que un cierto valor. Sea $(\Omega, \mathfrak{A}, \{\mathfrak{P}_\theta : \theta \in \Theta \subset \mathbb{R}^m\})$ una estructura estadística y sea T una estadística con valores en $(\mathbb{R}^m, \mathfrak{B}_{\mathbb{R}^m})$. Se supone que la estructura está dominada por una medida μ σ -finita, de densidad con respecto a μ , denotada f_θ la cual verifica:

$$\int_{\Omega} f_\theta d\mu = 1, \quad (1)$$

$$\int \frac{\partial \ln f_\theta}{\partial \theta_i} f_\theta d\mu = 0, \quad i = 1, \dots, m. \quad (2)$$

Si el conjunto $[f_\theta > 0] = \{\omega \in \Omega / f_\theta(\omega) > 0\}$ no depende de $\theta \in \Theta$, entonces (2) expresa que (1) se deriva bajo el signo de integral, pues $\frac{\partial \ln f_\theta}{\partial \theta_i} = \frac{1}{f_\theta} \frac{\partial f_\theta}{\partial \theta_i}$.

Recíprocamente, cuando $[f_\theta > 0]$ depende de θ , (2) no se cumple. Así se tiene que (2) $\iff E_\theta \left(\frac{\partial \ln f_\theta}{\partial \theta_i} \right) = 0$.

Definición 3.7.1 Si $\frac{\partial \ln f_\theta}{\partial \theta_i} \in L_2$ se llama información (en el sentido de Fisher) la matriz

$$I_\theta = \left\| \text{cov}_\theta \left(\frac{\partial \ln f_\theta}{\partial \theta_i}, \frac{\partial \ln f_\theta}{\partial \theta_j} \right) \right\|.$$

Cuando $\theta \in \mathbb{R}$, I_θ se reduce a $\text{var}_\theta \left(\frac{d \ln f_\theta}{d\theta} \right)$.

Podemos derivar (2) bajo el signo de integral:

$$\int \frac{\partial^2 \ln f_\theta}{\partial \theta_i \partial \theta_j} f_\theta d\mu + \int \frac{\partial \ln f_\theta}{\partial \theta_i} \cdot \frac{\partial \ln f_\theta}{\partial \theta_j} f_\theta d\mu = 0, \quad (2')$$

de donde:

$$E_\theta \left(\frac{\partial \ln f_\theta}{\partial \theta_i} \cdot \frac{\partial \ln f_\theta}{\partial \theta_j} \right) = \text{cov}_\theta \left(\frac{\partial \ln f_\theta}{\partial \theta_i}, \frac{\partial \ln f_\theta}{\partial \theta_j} \right) = E_\theta \left(\frac{\partial^2 \ln f_\theta}{\partial \theta_i \partial \theta_j} \right). \quad (3)$$

$$- \text{Si } \theta \in \mathbb{R} \quad E_{\theta} \left(\left[\frac{d \ln f_{\theta}}{d\theta} \right]^2 \right) = -E_{\theta} \left(\frac{d^2 \ln f_{\theta}}{d\theta^2} \right) = \text{var}_{\theta} \left(\frac{d \ln f_{\theta}}{d\theta} \right) = I_{\theta}.$$

– La densidad f_{θ} depende de μ , por lo tanto “a priori” las fórmulas (1), (2), (3) dependen de μ . Sin embargo, si ν es otra medida dominante $\frac{dP_{\theta}}{d\nu} = \frac{dP_{\theta}}{d\mu} \cdot \frac{d\mu}{d\nu}$, por lo que $\log \frac{dP_{\theta}}{d\nu} = \log \frac{dP_{\theta}}{d\mu} + \log \frac{d\mu}{d\nu}$. Tomando la derivada del logaritmo, se encuentra que $\frac{\partial \ln f_{\theta}}{\partial \theta_i}$ no depende de μ .

Teorema 3.7.1 Desigualdad de Cramer-Rao

Sea τ una función medible, $\tau: \mathbb{R}^m \rightarrow \mathbb{R}$ diferenciable y sea T una estadística real en L_2 tal que:

$$E_{\theta}(T) = \tau(\theta) = \int T f_{\theta} d\mu.$$

Se supone que la condición (1) se satisface, así como la siguiente condición:

$$\frac{\partial \tau}{\partial \theta_i} = \int T \frac{\partial \ln f_{\theta}}{\partial \theta_i} f_{\theta} d\mu, \quad (4)$$

entonces se tiene:

$$\frac{\left[\frac{\partial \tau}{\partial \theta_i} \right]^2}{E_{\theta} \left(\left[\frac{\partial \ln f_{\theta}}{\partial \theta_i} \right]^2 \right)} \leq \text{var}_{\theta}(T), \quad i = 1, \dots, m.$$

Demostración Si $U, V \in L_2$ sabemos que $|\text{cov}(U, V)|^2 \leq \text{var}(U) \text{var}(V)$, de donde

$$\left| \text{cov}_{\theta} \left(T, \frac{\partial \ln f_{\theta}}{\partial \theta_i} \right) \right|^2 \leq \text{var}_{\theta}(T) \text{var}_{\theta} \left(\frac{\partial \ln f_{\theta}}{\partial \theta_i} \right).$$

Por la condición (4) tenemos que $\text{cov}_{\theta} \left(T, \frac{\partial \ln f_{\theta}}{\partial \theta_i} \right) = E_{\theta} \left(T \frac{\partial \ln f_{\theta}}{\partial \theta_i} \right) = \frac{\partial \tau}{\partial \theta_i}$, es decir

$$\left[\frac{\partial \tau}{\partial \theta_i} \right]^2 \leq \text{var}_{\theta}(T) \text{var}_{\theta} \left(\frac{\partial \ln f_{\theta}}{\partial \theta_i} \right).$$

– La desigualdad de Cramer-Rao $\equiv m$ desigualdades.

– En el caso particular en que $\tau(\theta) = \theta_j$, caso en que T es un estimador insesgado de θ_j , la desigualdad se escribe:

$$\text{var}_{\theta}(T) \geq \frac{1}{E_{\theta} \left(\left[\frac{\partial \ln f_{\theta}}{\partial \theta_j} \right]^2 \right)} = \frac{1}{\text{var}_{\theta} \left(\frac{\partial \ln f_{\theta}}{\partial \theta_j} \right)} = \frac{1}{-E_{\theta} \left(\frac{\partial^2 \ln f_{\theta}}{\partial \theta_j^2} \right)}.$$

Práctica de la desigualdad

Si se desea estimar $\theta \in \mathbb{R}$ y si T estima sin sesgo θ , entonces la desigualdad da un minorante de la varianza de T y si se encuentra que $\text{var}_\theta(T) = \frac{1}{E_\theta\left(\left[\frac{d \ln f_\theta}{d\theta}\right]^2\right)}$, entonces T es el mejor estimador (en el sentido del error cuadrático medio).

Ejemplo 1 Se considera $(\{0, 1\}, \mathfrak{P}(\{0, 1\}), \{P_\theta : \theta \in]0, 1[]\}^n$, con $P_\theta(\{1\}) = \theta$, $P_\theta(\{0\}) = 1 - \theta$. Así tenemos $f_\theta(x) = \theta^{\sum_{i=1}^n x_i} (1 - \theta)^{n - \sum_{i=1}^n x_i}$, con $x \in \{0, 1\}^n$. Si tomamos la variable aleatoria $X_i(x) = x_i$, entonces $\sum_{i=1}^n X_i \sim B(n, \theta)$ y además:

$$\ln f_\theta(x) = \left(\sum_{i=1}^n x_i\right) \ln \theta + \left(n - \sum_{i=1}^n x_i\right) \ln(1 - \theta),$$

por lo que:

$$\frac{d \ln f_\theta(x)}{d\theta} = \frac{1}{\theta} \sum_{i=1}^n x_i - \frac{1}{1 - \theta} \left(n - \sum_{i=1}^n x_i\right).$$

Por otro lado, se verifica (1), pues $E_\theta\left(\frac{d \ln f_\theta}{d\theta}\right) = \frac{n\theta}{\theta^2} - \frac{n(1 - \theta)}{(1 - \theta)^2} = 0$ y además $\frac{d^2 \ln f_\theta}{d\theta^2}(x) = -\frac{1}{\theta^2} \sum_{i=1}^n x_i - \frac{1}{1 - \theta^2} \left(n - \sum_{i=1}^n x_i\right)$ cumple:

$$-E_\theta\left(\frac{d^2 \ln f_\theta}{d\theta^2}\right) = \frac{n\theta}{\theta^2} + \frac{n(1 - \theta)}{(1 - \theta)^2} = \frac{n}{\theta(1 - \theta)} = E_\theta\left(\left[\frac{d \ln f_\theta}{d\theta}\right]^2\right)$$

Si T estima sin sesgo θ , $E_\theta(T) = \theta$ y por Cramer-Rao:

$$\text{var}_\theta(T) \geq \frac{\theta(1 - \theta)}{n} = \frac{1}{E_\theta\left(\left[\frac{d \ln f_\theta}{d\theta}\right]^2\right)},$$

por lo tanto, si $T = \frac{1}{n} \sum_{i=1}^n X_i$, $E_\theta(T) = \theta$, es un estimador sin sesgo y $\text{var}_\theta(T) = \frac{\theta(1 - \theta)}{n}$, es decir T es el mejor estimador.

Ejemplo 2 Se considera $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}, \{N(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma > 0\}})^n$.

a) Demostrar que \bar{X} es el mejor estimador de μ .

b) Estudiar la estadística $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$.

Solución

a) Sea $f_\mu(x) = \frac{1}{(\sqrt{2\pi}\sigma)^n} e^{-\frac{1}{2} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma}\right)^2}$, $\frac{\partial \ln f}{\partial \mu} = \sum_{i=1}^n \frac{x_i - \mu}{\sigma^2}$, $E_\mu\left(\frac{\partial \ln f_\mu}{\partial \mu}\right) = 0$,

$$\frac{\partial^2 \ln f_\mu}{\partial \mu^2} = -\frac{n}{\sigma^2}, \text{ i.e. } E_\mu \left(-\frac{\partial^2 \ln f_\mu}{\partial \mu^2} \right) = (\sigma^2/n)^{-1} = \frac{1}{\text{var}_\mu(\bar{X})}.$$

$$\text{b) } \frac{\partial \ln f_\sigma}{\partial \sigma} = \frac{n}{\sigma} - \sum_{i=1}^n \frac{(x_i - \mu)^2}{\sigma^3}, \quad E_\sigma \left(\frac{\partial \ln f_\sigma}{\partial \sigma} \right) = \frac{n}{\sigma} - \frac{n\sigma^2}{\sigma^3} = 0,$$

$$\frac{\partial^2 \ln f_\sigma}{\partial \sigma^2} = -\frac{n}{\sigma^2} + 3 \sum_{i=1}^n \frac{(x_i - \mu)^2}{\sigma^4}, \quad E_\sigma \left(\frac{\partial^2 \ln f_\sigma}{\partial \sigma^2} \right) = -\frac{n}{\sigma^2} + \frac{3n\sigma^2}{\sigma^4} = \frac{2n}{\sigma^2}.$$

Así, $E(S^2) = \frac{n-1}{n-1}\sigma^2 = \sigma^2$, es un estimador sin sesgo.

$$\text{var}(S^2) = \frac{2\sigma^4}{n-1} \geq \frac{4\sigma^4}{2n} = \frac{2\sigma^4}{n}.$$

Caso en que el minorante de Cramer-Rao es esperado

Sea $\theta \in \mathbb{R}$, estudiaremos el caso en que existe T tal que:

$$(1) \quad \begin{cases} E_\theta(T) = \tau(\theta) \\ \text{var}_\theta(T) = \frac{\tau'(\theta)^2}{E_\theta \left(\left[\frac{d \ln f_\theta}{d\theta} \right]^2 \right)}. \end{cases}$$

La idea es ir a la demostración de Cramer-Rao pues $|\text{cov}(U, V)|^2 \leq \text{var}(U) \text{var}(V)$ y la igualdad se da si y sólo si $U = aV + b$.

Las condiciones (1) se satisfacen si:

$$(2) \quad \frac{d \ln f_\theta}{d\theta} = a(\theta)T + b(\theta).$$

Integrando obtenemos:

$$\ln f_\theta(\omega) = A(\theta)T(\omega) + B(\theta) + C(\omega)$$

$$f_\theta(\omega) = e^{A(\theta)T(\omega) + B(\theta) + C(\omega)},$$

es decir tenemos una estructura de tipo exponencial.

Si (2) sucede, tenemos que $A'(\theta)E_\theta(T) = -B'(\theta)$, lo que implica:

$$E_\theta(T) = -\frac{B'(\theta)}{A'(\theta)} = \tau(\theta),$$

con lo cual se ha demostrado el teorema siguiente.

Teorema 3.7.2 *La desigualdad de Cramer-Rao satisface (2) si y sólo si existe una estadística T , en una estructura de tipo exponencial que sea un estimador sin sesgo de $-\frac{B'(\theta)}{A'(\theta)}$.*

Definición 3.7.2 Eficacidad de un estimador

Sabemos que $\text{var}_\theta(T) \geq k(\theta) = \frac{\tau'(\theta)^2}{E_\theta\left(\left[\frac{d \ln f_\theta}{d\theta}\right]^2\right)}$, entonces llamaremos *eficacidad del estimador T* al valor:

$$\alpha = \frac{k(\theta)}{\text{var}_\theta(T)} \leq 1.$$

Si $\alpha = 1$ se dice que T es un *estimador eficaz* (i.e. cumple la desigualdad de Cramer-Rao). También se dice que T es *eficiente*.

3.8 Estudio de muestras empíricas de leyes gaussianas

En el estudio analizaremos muestras del tipo $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}, \{N(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma > 0\}})^n$.

3.8.1 Estudio de la estadística (\bar{X}, S^2)

Sean X_i las aplicaciones coordenadas $X_i: \mathbb{R}^n \rightarrow \mathbb{R}$, $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$, $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$.

Sabemos que (\bar{X}, S^2) es una estadística suficiente y completa y estima (μ, σ^2) .

Distribución de (\bar{X}, S^2)

Se sabe que los X_i son independientes de ley $N(\mu, \sigma^2)$. Sea $U_i = \frac{X_i - \mu}{\sigma}$, son independientes y de ley $N(0, 1)$. Si $U = (U_1, \dots, U_n)$, entonces $U \sim N(0, I_n)$.

Se efectúa el cambio de variable $V = AU$, $A = \begin{pmatrix} \frac{1}{\sqrt{n}} & \dots & \frac{1}{\sqrt{n}} \\ & & M \end{pmatrix}$, donde A es una matriz de orden n

ortogonal con la primera fila igual a $\frac{1}{\sqrt{n}}(1, \dots, 1)$. La ley de probabilidad de V es multinormal a n

dimensiones y $V \sim N(0, AIA') = N(0, I)$, por lo tanto $\sum_{i=1}^n U_i^2 = U'U = V'AA'V = V'V = \sum_{i=1}^n V_i^2$.

Sabemos que $\sum_{i=1}^n (X_i - \mu)^2 = \sum_{i=1}^n (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2$, o sea $\sum_{i=1}^n U_i^2 = \frac{n}{\sigma^2}(\bar{X} - \mu)^2 + \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma}\right)^2 = \frac{n}{\sigma^2}(\bar{X} - \mu)^2 + \frac{n-1}{\sigma^2}S^2$.

Como $V = AU$, se tiene que $V_1 = \frac{1}{\sqrt{n}} \sum_{i=1}^n U_i = \frac{n(\bar{X} - \mu)}{\sqrt{n}\sigma}$, $V_1^2 = \frac{n}{\sigma^2}(\bar{X} - \mu)^2$, entonces:

$$\frac{n-1}{\sigma^2}S^2 = V_2^2 + \dots + V_n^2.$$

– Dado que $\bar{X} \sim N(\mu, \sigma^2/n)$ se tiene $\frac{n-1}{\sigma^2}S^2 \sim \chi_{n-1}^2$, es decir $S^2 \sim \Gamma\left(\frac{n-1}{2}, \frac{n-1}{2\sigma^2}\right)$.

– Los V_i son independientes i.e. V_1 es independiente de $\sum_{i=2}^n V_i^2$, o sea \bar{X} y S^2 son independientes. El

resultado sólo es válido para leyes normales.

– La distribución de (\bar{X}, S^2) es el producto de una normal y una gama.

3.8.2 Aplicación al intervalo de confianza

Consideramos siempre la estructura $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}, \{N(\mu, \sigma^2): \mu \in \mathbb{R}, \sigma > 0\}})^n$ y se desea encontrar un intervalo de confianza para μ de nivel α .

Si conocemos σ el intervalo será $\bar{X} \pm \frac{\sigma}{\sqrt{n}} t_\alpha$, donde t_α es tal que $P(|N(0, 1)| < t_\alpha) = 1 - \alpha$.

Pero en general σ es desconocido por lo que es necesario encontrar otro método para poder obtener un intervalo de confianza para μ de nivel α .

Para esto consideramos la estadística:

$$T = \frac{\sqrt{\frac{n}{\sigma^2}}(\bar{X} - \mu)}{\sqrt{\frac{n-1}{\sigma^2} \frac{S^2}{n-1}}} = \frac{\bar{X} - \mu}{\sqrt{S^2/n}} \sim t_{n-1}.$$

Se puede calcular $P(|T| < t_{n-1, \frac{\alpha}{2}}) = 1 - \alpha$ y el intervalo será $\bar{X} \pm \sqrt{\frac{S^2}{n}} t_{n-1, \frac{\alpha}{2}}$.

3.8.3 Pruebas usuales para muestras normales

A – Una muestra: hipótesis sobre las medias

Estructura estadística $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}, \{N(\mu, \sigma^2): \mu \in \mathbb{R}, \sigma > 0\}})^n$

Hipótesis $H_0 = \{N(\mu_0, \sigma^2): \sigma > 0\}$, $H'_1 = \{N(\mu, \sigma^2): \mu > \mu_0, \sigma > 0\}$,

$H''_1 = \{N(\mu, \sigma^2): \mu < \mu_0, \sigma > 0\}$, $H_1 = \{N(\mu, \sigma^2): \mu \neq \mu_0, \sigma > 0\}$.

Las pruebas usuales, llamadas pruebas de Student, son las siguientes, definiendo $\nu = n - 1$:

a) Prueba de H_0 contra H'_1 (prueba unilateral): $\mu = \mu_0$ contra $\mu > \mu_0$.

$\Phi(\mathbf{x}) = 1 \iff T = \frac{\sqrt{n}(\bar{X} - \mu_0)}{S} > t_{\nu, \alpha}$, o sea se rechaza H_0 si $T(\mathbf{x}) > t_{\nu, \alpha}$, donde $T(\mathbf{x})$ es una realización de la variable aleatoria T de Student.

b) Prueba de H_0 contra H''_1 (prueba unilateral): $\mu = \mu_0$ contra $\mu < \mu_0$.

$\Phi(\mathbf{x}) = 1 \iff T = \frac{\sqrt{n}(\bar{X} - \mu_0)}{S} < -t_{\nu, \alpha}$, i.e. se rechaza H_0 si $T(\mathbf{x}) < -t_{\nu, \alpha}$, donde $T(\mathbf{x})$ es una realización de la variable aleatoria T de Student.

c) Prueba de H_0 contra H_1 (prueba bilateral): $\mu = \mu_0$ contra $\mu \neq \mu_0$.

$$\Phi(x) = 1 \iff \frac{\sqrt{n}(\bar{X} - \mu_0)}{S} < -t_{\nu, \frac{\alpha}{2}} \text{ o } \frac{\sqrt{n}(\bar{X} - \mu_0)}{S} > t_{\nu, \frac{\alpha}{2}},$$

$$\text{i.e. se rechaza } H_0 \iff \frac{\sqrt{n}|\bar{X} - \mu_0|}{S} > t_{\nu, \frac{\alpha}{2}}.$$

Propiedades

1) En los tres casos, por la definición de $t_{\nu, \alpha}$, el nivel de Φ es α . Además la imagen de estas pruebas es constante (igual a α) sobre H_0 es decir $\forall \sigma, \sigma > 0$. Estas pruebas son libres sobre H_0 . Observe que Φ está en función de la estadística suficiente (\bar{X}, S^2) .

Las tablas permiten encontrar los valores $t_{\nu, \alpha}$ para pequeños valores de ν . Para valores grandes de ν se utiliza el hecho de que la ley de Student se aproxima a la ley normal.

2) La potencia depende a la vez de μ y σ ; así $\frac{\sqrt{n}(\bar{X} - \mu_0)}{S}$ sigue una distribución de Student no centrada con $\nu = n - 1$ g.d.l. y de parámetro de descentraje $\delta = \frac{\sqrt{n}(\mu - \mu_0)}{\sigma}$. Se obtiene una ley de Student no centrada de parámetros ν y δ como el cociente de una variable aleatoria $N(\delta, 1)$ y la raíz cuadrada de un χ_ν^2 dividido por ν . Estas variables son independientes. Las tablas dan la potencia de las pruebas, así por ejemplo para la primera prueba unilateral, dan la probabilidad de que una variable aleatoria de Student no centrada de parámetros ν y δ sobrepase el valor $t_{\nu, \alpha}$.

3) Se prueba fácilmente que una prueba es convergente para todo $\alpha \in]0, 1[$, es decir que la potencia $\beta_\Phi(P) \rightarrow 1$, cuando $n \rightarrow \infty, \forall P \in \bar{H}_0$.

4) En los casos a) y b) las pruebas son U.M.P. y en el caso c) es U.M.P.I.

B – Dos muestras

B.1 Hipótesis sobre las varianzas

Estructura estadística

$$(\mathbb{R}, \mathfrak{B}_{\mathbb{R}, \{N(\mu_1, \sigma_1): \mu_1 \in \mathbb{R}, \sigma_1 > 0\}})^{n_1} \otimes (\mathbb{R}, \mathfrak{B}_{\mathbb{R}, \{N(\mu_2, \sigma_2): \mu_2 \in \mathbb{R}, \sigma_2 > 0\}})^{n_2}$$

Hipótesis $H_0 = \{N(\mu_1, \sigma^2) \otimes N(\mu_2, \sigma^2): \mu_1, \mu_2 \in \mathbb{R}, \sigma > 0\} \equiv \sigma_1 = \sigma_2, H'_1: \sigma_1 > \sigma_2, H''_1: \sigma_1 < \sigma_2,$

$H_1: \sigma_1 \neq \sigma_2.$

Denotemos $X_{11}, X_{12}, \dots, X_{1n_1}; X_{21}, X_{22}, \dots, X_{2n_2}$ las aplicaciones coordenadas y $\bar{X}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} X_{1i},$

$$S_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (X_{1i} - \bar{X}_1)^2, \bar{X}_2 = \frac{1}{n_2} \sum_{i=1}^{n_2} X_{2i}, S_2^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (X_{2i} - \bar{X}_2)^2.$$

Dado que S_1^2 y S_2^2 son independientes, entonces el cociente $F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(\nu_1, \nu_2)$, donde $\nu_1 = n_1 - 1$, y $\nu_2 = n_2 - 1$. Se denota $f_{\nu_1, \nu_2, \alpha}$ el valor que sobrepasa, con una probabilidad α , una variable aleatoria de Fisher de parámetros ν_1 y ν_2 .

Las pruebas usuales, llamadas pruebas de Fisher, son las siguientes (con $\nu_1 = n_1 - 1$ y $\nu_2 = n_2 - 1$):

a) Prueba de H_0 contra H_1' i.e. $\sigma_1 = \sigma_2$ contra $\sigma_1 > \sigma_2$.

$$\Phi(\mathbf{x}) = 1 \iff S_1^2/S_2^2 > f_{\nu_1, \nu_2, \alpha}.$$

Se rechaza H_0 si $(S_1^2/S_2^2)(\mathbf{x}) > f_{\nu_1, \nu_2, \alpha}$, donde $(S_1^2/S_2^2)(\mathbf{x})$ es una realización de la variable aleatoria F bajo H_0 .

b) Prueba de H_0 contra H_1'' i.e. $\sigma_1 = \sigma_2$ contra $\sigma_1 < \sigma_2$.

$$\Phi(\mathbf{x}) = 1 \iff S_2^2/S_1^2 > f_{\nu_2, \nu_1, \alpha}.$$

Se rechaza H_0 si $F(\mathbf{x}) > f_{\nu_2, \nu_1, \alpha}$ bajo la hipótesis H_0 .

c) Prueba de H_0 contra H_1 i.e. $\sigma_1 = \sigma_2$ contra $\sigma_1 \neq \sigma_2$.

$$\Phi(\mathbf{x}) = 1 \iff \frac{S_1^2}{S_2^2} > f_{\nu_1, \nu_2, \frac{\alpha}{2}} \text{ o } \frac{S_2^2}{S_1^2} > f_{\nu_2, \nu_1, \frac{\alpha}{2}}.$$

Se rechaza H_0 si $F(\mathbf{x})$ o $1/F(\mathbf{x})$ es mayor que $f_{\nu_1, \nu_2, \frac{\alpha}{2}}$ o $f_{\nu_2, \nu_1, \frac{\alpha}{2}}$.

En la práctica se considera $\sup(S_2^2, S_1^2)/\inf(S_2^2, S_1^2)$ y se compara con $f_{\nu_1', \nu_2', \frac{\alpha}{2}}$, donde ν_1' son los grados de libertad del numerador y ν_2' son los grados de libertad del denominador.

Propiedades

1) Las pruebas son de nivel α y libres sobre H_0 .

2) La potencia es fácil de calcular; no depende de μ_1 ni μ_2 y depende solamente del cociente $\rho = \sigma_1/\sigma_2$. Para la prueba de H_0 contra H_1' por ejemplo, para un nivel α , la potencia es igual a la probabilidad de que una Fisher $F(\nu_1, \nu_2)$ sobrepase el valor $\left(\frac{\sigma_2}{\sigma_1}\right)^2 f_{\nu_1, \nu_2, \alpha}$.

B.2 Hipótesis sobre las medias

Estructura estadística

$$(\mathbb{R}, \mathfrak{B}_{\mathbb{R}, \{N(\mu_1, \sigma): \mu_1 \in \mathbb{R}, \sigma > 0\}})^{n_1} \otimes (\mathbb{R}, \mathfrak{B}_{\mathbb{R}, \{N(\mu_2, \sigma): \mu_2 \in \mathbb{R}, \sigma > 0\}})^{n_2}$$

Notemos que las varianzas de las muestras son iguales a priori.

Hipótesis $H_0: \mu_1 = \mu_2$, $H'_1: \mu_1 > \mu_2$, $H''_1: \mu_1 < \mu_2$, $H_1: \mu_1 \neq \mu_2$.

Adoptamos las mismas notaciones que en B.1. Por otro lado:

$$S^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}.$$

Observemos que S^2 es una estimación “común” de la varianza para las dos muestras.

Las pruebas usuales son las siguientes (pruebas de Student) con $\nu = n_1 + n_2 - 2$ g.d.l.

a) Prueba de H_0 contra H'_1 ($\mu_1 = \mu_2$ contra $\mu_1 > \mu_2$).

$$\text{Se rechaza } H_0 \text{ si } \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} > -t_{\nu, \alpha}.$$

b) Prueba de H_0 contra H''_1 ($\mu_1 = \mu_2$ contra $\mu_1 < \mu_2$).

$$\text{Se rechaza } H_0 \text{ si } \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} < -t_{\nu, \alpha}.$$

c) Prueba de H_0 contra H_1 ($\mu_1 = \mu_2$ contra $\mu_1 \neq \mu_2$).

$$\text{Se rechaza } H_0 \text{ si } \frac{|\bar{X}_1 - \bar{X}_2|}{\sqrt{S^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} > t_{\nu, \frac{\alpha}{2}}.$$

Propiedades Tenemos siempre pruebas libres sobre H_0 , de nivel α . La potencia se obtiene a través de una ley de Student descentrada con $\nu = n_1 + n_2 - 2$ g.d.l. y de parámetro de descentrage $\delta =$

$$\frac{\mu_1 - \mu_2}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}.$$

Las pruebas son convergentes cuando a la vez n_1 y n_2 tienden a infinito.

Caso en que las varianzas son diferentes**Estructura estadística**

$$(\mathbb{R}, \mathfrak{B}_{\mathbb{R}, \{N(\mu_1, \sigma_1^2): \mu_1 \in \mathbb{R}, \sigma_1 > 0\}})^{n_1} \otimes (\mathbb{R}, \mathfrak{B}_{\mathbb{R}, \{N(\mu_2, \sigma_2^2): \mu_2 \in \mathbb{R}, \sigma_2 > 0\}})^{n_2}.$$

Hipótesis $H_0: \mu_1 = \mu_2$, $H'_1: \mu_1 > \mu_2$, $H_1: \mu_1 \neq \mu_2$.

Aquí indicamos la prueba utilizable para grandes valores de n_1 y n_2 . Con las notaciones precedentes:

$$\frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1), \quad \text{si } \mu_1 = \mu_2.$$

Por otro lado S_1^2 y S_2^2 convergen en probabilidad respectivamente a σ_1^2 y σ_2^2 , si n_1 y n_2 tienden a infinito.

$$\text{Si } \mu_1 = \mu_2, \quad \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \xrightarrow{\text{ley}} N(0, 1), \quad \text{cuando } n_1, n_2 \rightarrow \infty.$$

Sea t_α el valor tal que $P(N(0, 1) > t_\alpha) = \alpha$.

Las siguientes pruebas son *asintóticamente de nivel α* . Para valores grandes de n_1 y n_2 tenemos:

a) Prueba de H_0 contra H'_1 ($\mu_1 = \mu_2$ contra $\mu_1 > \mu_2$).

$$\text{Se rechaza } H_0 \text{ si } \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} > t_\alpha.$$

b) Prueba de H_0 contra H_1 ($\mu_1 = \mu_2$ contra $\mu_1 \neq \mu_2$).

$$\text{Se rechaza } H_0 \text{ si } \frac{|\bar{X}_1 - \bar{X}_2|}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} > t_{\frac{\alpha}{2}}.$$

3.9 La prueba χ^2 (chi-cuadrado)

Nos interesamos aquí en pruebas que permiten verificar la hipótesis H_0 : homogeneidad de muestras, independencia de caracteres, ajuste de una ley a una función de probabilidad determinada, dada a priori, etc., contra la hipótesis alternativa H_1 : no H_0 .

Para esto estudiaremos la ley límite de una distribución multinomial.

Teorema 3.9.1 Se considera $X = (X_1, \dots, X_k)$ una variable aleatoria multinomial $M(n, p_1, \dots, p_k)$.

Sea $Y_j = \frac{X_j - np_j}{\sqrt{np_j}}$, entonces $\sum_{j=1}^k Y_j^2 \sim$ asintóticamente como una χ_{k-1}^2 (es decir converge en ley cuando $n \rightarrow \infty$).

Demostración Sea $Y_j = \frac{X_j - np_j}{\sqrt{np_j}}$ y como $\sum_{j=1}^k X_j = n$, entonces $\sum_{i=j}^k Y_j \sqrt{p_j} = 0$. Además sabemos que $E(X_j) = np_j$, $\text{var}(X_j) = np_j(1 - p_j)$, por lo que:

$$Y_j = \sqrt{n} \frac{\frac{X_j}{n} - p_j}{\sqrt{p_j}} \xrightarrow{\text{ley}} N(0, 1 - p_j),$$

por el teorema central del límite.

Sea $K_n = \sum_{j=1}^k Y_j^2$ y sea S una matriz ortogonal tal que $S = \begin{pmatrix} M \\ \sqrt{p_1} \cdots \sqrt{p_k} \end{pmatrix}$.

Se define $Z = SY$, entonces $Z_k = \sum_{j=1}^k Y_j \sqrt{p_j} = 0$, por lo tanto:

$$K_n = \sum_{i=j}^k Y_j^2 = Z'SS'Z = \sum_{j=1}^k Z_j^2 = \sum_{j=1}^{k-1} Z_j^2.$$

La función característica de X es:

$$\begin{aligned} \Phi_X(t_1, \dots, t_k) &= (p_1 e^{it_1} + \cdots + p_k e^{it_k})^n \\ \Phi_Y(t_1, \dots, t_k) &= E(e^{it'Y}) = e^{-i \sum_{j=1}^k \sqrt{np_j} t_j} \left[\sum_{j=1}^k p_j e^{i \frac{t_j}{\sqrt{np_j}}} \right]^n. \end{aligned}$$

Pero, $\Phi_Z(\mathbf{u}) = \Phi_Y(S'\mathbf{u}) = \Phi_Y(\mathbf{t})$, donde $\mathbf{t} = S'\mathbf{u}$.

Sea $\mathbf{u} = S\mathbf{t}$, entonces $u_k = \sum_{j=1}^k \sqrt{p_j} t_j$, de donde $\Phi_Z(\mathbf{u}) = e^{-i \sqrt{nu_k}} \left[\sum_{j=1}^k p_j e^{i \frac{t_j}{\sqrt{np_j}}} \right]^n$ y además:

$$\Phi_{z_1, \dots, z_{k-1}}(u_1, \dots, u_{k-1}) = \Phi_z(u_1, \dots, u_{k-1}, 0) = \left[\sum_{j=1}^k p_j e^{i \frac{t_j}{\sqrt{np_j}}} \right]^n,$$

con $\mathbf{t}' = (u_1, \dots, u_{k-1}, 0)S = \mathbf{u}^* S$.

Por otro lado, $e^{i \frac{t_j}{\sqrt{np_j}}} = 1 + \frac{i}{\sqrt{n}} \frac{t_j}{\sqrt{p_j}} - \frac{1}{2n} \frac{t_j^2}{p_j} + \frac{1}{n} o(1)$, con $o(1) \rightarrow 0$, cuando $n \rightarrow \infty$. Así:

$$\sum_{j=1}^k p_j e^{i \frac{t_j}{\sqrt{np_j}}} = 1 + \frac{i}{\sqrt{n}} \sum_{j=1}^k t_j \sqrt{p_j} - \frac{1}{2n} \sum_{j=1}^k t_j^2 + \frac{1}{n} o(1),$$

donde $\sum_{j=1}^k t_j \sqrt{p_j} = \mathbf{s}'_k \mathbf{t} = \mathbf{s}'_k S' \mathbf{u}$, con $S' = (\mathbf{s}_1, \dots, \mathbf{s}_k)$. Pero como S es ortogonal $\mathbf{s}'_k S' = (0, \dots, 0, 1)$

y $\mathbf{s}'_k S' \mathbf{u} = 0$. Además:

$$\log \Phi_{z_1, \dots, z_{k-1}}(u_1, \dots, u_{k-1}) = n \log \left[1 - \frac{1}{2n} \sum_{j=1}^k t_j^2 + \frac{1}{n} o(n) \right] = -\frac{1}{2} \sum_{j=1}^k t_j^2 + o(1).$$

Dado que $\sum_{j=1}^k t_j^2 = \mathbf{t}'\mathbf{t} = (S'\mathbf{u}^*)'S'\mathbf{u}^* = \mathbf{u}^* \mathbf{u}^* = \sum_{j=1}^{k-1} u_j^2$, se concluye que:

$$\Phi_{z_1, \dots, z_{k-1}}(u_1, \dots, u_{k-1}) \rightarrow e^{-\frac{1}{2} \sum_{j=1}^{k-1} u_j^2}, \quad \text{cuando } n \rightarrow \infty.$$

Las variables aleatorias Z_1, \dots, Z_{k-1} son asintóticamente normales, reducidas e independientes y la suma de los cuadrados se distribuyen asintóticamente como un χ_{k-1}^2 .

– Para n grande, $K_n = \sum_{j=1}^k Y_j^2$ tiene una ley próxima a la χ^2 y se utiliza esta ley aproximada para construir una prueba de H_0 : “la ley considerada es la ley multinomial de parámetros (n, p_1, \dots, p_k) ”, contra H_1 : no H_0 de la forma siguiente:

– Si $K_n > \ell_{k-1, \alpha}$ se rechaza H_0 .

– Si $K_n \leq \ell_{k-1, \alpha}$ se acepta H_0 ,

con $P(\chi_{k-1}^2 > \ell_{k-1, \alpha}) = \alpha$.

– La prueba tiene nivel asintótico α y se demuestra que es convergente.

3.9.1 La prueba χ^2 como una prueba de ajuste

a) Caso de parámetros conocidos

Prueba de una proporción Si $X \sim B(n, p)$, $0 < p < 1$ y si $H_0 : p = p_0$, entonces se rechaza la hipótesis H_0 si:

$$K_n = \frac{(X - np_0)^2}{np_0(1 - p_0)} > \ell_{k-1, \alpha}, \quad k = 2 \text{ i.e } k - 1 = 1.$$

Ejemplo Se efectúan 1000 lanzamientos de una moneda y se obtiene 464 veces cara y 536 cruz. Probar con un nivel $\alpha = 0.05$ si la moneda es perfecta.

Consideramos que la probabilidad de obtener cara o cruz es la misma $H_0 : p = \frac{1}{2}$. Al aplicar la prueba χ^2 tenemos:

$$K_n = \frac{(464 - 500)^2}{500} + \frac{(536 - 500)^2}{500} = 5.184, \quad \ell_{1, 0.05} = 3.841,$$

es decir se rechaza H_0 y se admite que la moneda está mal equilibrada.

– Observemos que $K_n = \frac{(X_1 - np)^2}{np} + \frac{(X_2 - nq)^2}{nq} = \frac{(X_1 - np)^2}{npq}$.

– $\left[\frac{X}{n} \pm \frac{1}{n} \sqrt{\ell_{1, \alpha} X \left(1 - \frac{X}{n} \right)} \right]$ es un intervalo de confianza asintótico para p , de coeficiente de seguridad $1 - \alpha$.

Ajuste a una ley discreta

Estudiando 53680 familias con 8 hijos, se han observado los resultados siguientes: (k es el n° de

varones, n_k es el n° de familias con k varones)

k	0	1	2	3	4	5	6	7	8
n_k	215	1485	5331	10649	14959	11929	6678	2092	342

H_0 : El número de varones en una familia es distribuido como una ley binomial de parámetro un $\frac{1}{2}$.

Así $P_k = \binom{k}{8}(\frac{1}{2})^8 = \binom{k}{8}/2^8$ y tenemos la tabla siguiente:

k	0	1	2	3	4	5	6	7	8
p_k	$\frac{1}{256}$	$\frac{8}{256}$	$\frac{28}{256}$	$\frac{56}{256}$	$\frac{70}{256}$	$\frac{56}{256}$	$\frac{28}{256}$	$\frac{8}{256}$	$\frac{1}{256}$
np_k	210	1680	5580	11760	14700	11760	5880	1680	210
n_k	215	1485	5331	10649	14959	11929	6678	2192	342
$\frac{(n_k - np_k)^2}{np_k}$	0.12	22.0	49.8	101.7	5.38	2.98	117.6	104.7	82.97

Tenemos que $K_n = 478.34$ y hay 9 categorías, es decir 8 g.d.l. La cota al nivel 1% es $\ell_{8,0.01} = 20.09$, por lo que la probabilidad dentro de cada familia de tener un varón en cada nacimiento no es $\frac{1}{2}$, o bien la ley no es binomial (lo que supone que los nacimientos no constituyen en cuanto al sexo, eventos independientes en cada familia).

Ajuste a una ley continua

Sea X la variable aleatoria de función de repartición F completamente determinada. Se agrupan las observaciones en clases C_i y se calcula p_i , es decir la igualdad de la probabilidad observada a la probabilidad teórica de pertenecer a una clase determinada:

$$X_i = n^\circ \text{ observaciones de la clase } C_i,$$

$$k = n^\circ \text{ de clases,}$$

$$n = n^\circ \text{ de observaciones.}$$

$$\text{La estadística } K_n = \sum_{i=1}^k \frac{(X_i - np_{i_0})^2}{np_{i_0}} \xrightarrow{\text{ley}} \chi_{k-1}^2 \text{ y se denotará } K_n = \sum \frac{(o-t)^2}{t}.$$

– En la práctica el efectivo de cada clase debe ser mayor o igual a 8.

– Diremos que la muestra no se ajusta a la ley de probabilidad F , si $K_n > \ell_{k-1,\alpha}$.

– Si los efectivos por clase son pequeños, se agrupan en varias clases. Sin embargo se tiene interés en tener el máximo de clases para perder el mínimo de información.

– La distribución de K_n es independiente de F .

b) Caso de parámetros desconocidos

Sea $F(x, \theta)$ la función de repartición de una variable aleatoria, donde θ es un parámetro en \mathbb{R}^s que se debe estimar. Utilizando la muestra, θ es estimado por $\hat{\theta}_n$. Se tendrá entonces $K_n(\hat{\theta}_n) = \sum_{i=1}^k \frac{(X_i - np_i(\hat{\theta}_n))^2}{np_i(\hat{\theta}_n)}$.

Teorema 3.9.2 Teorema de Cramer

Cuando F pertenece a la familias de leyes exponenciales $K_n(\hat{\theta}_n)$ es asintóticamente distribuido como un χ^2_ν , con $\nu = k - s - 1$ g.d.l., donde s es el número de parámetros desconocidos.

– Las estimaciones se obtienen a partir del método de máxima verosimilitud o un método equivalente. Nosotros aceptaremos el teorema sin demostración.

Ejemplo En los 400 cuadros de un hematímetro, se han contado 1872 organismos con la repartición siguiente:

nº de organismos	0 – 1	2	3	4	5	6	7	8	9	10 >
nº de cuadros	20	43	53	86	70	54	37	18	10	9

¿Se puede suponer que el número de organismos por cuadro sigue una ley de Poisson?

El parámetro de esta ley de Poisson es desconocido y se estima por $\hat{\lambda} = \frac{1872}{400} = 4.68$.

Los efectivos teóricos son $400p_k$, con $p_k = e^{-\hat{\lambda}} \frac{\hat{\lambda}^k}{k!}$.

	0 – 1	2	3	4	5	6	7	8	9	10	total
np_k	21.0	40.6	63.4	74.1	69.4	54.2	36.2	21.2	11.0	8.66	400
x_k	20	43	53	86	70	54	37	18	10	9	400
$x_k - np_k$	1.08	2.35	10.4	11.8	0.56	0.16	0.79	3.18	1.02	0.34	
$(x_k - np_k)^2$	1.16	5.52	108.0	139.0	0.31	0.026	0.62	10.1	1.04	0.11	
$\frac{(x_k - np_k)^2}{np_k}$	0.055	0.136	1.71	1.88	0.005	0.000	0.017	0.477	0.094	0.013	4.38

Hay 10 clases y un parámetro estimado, $\nu = 8$ g.d.l. La cota al 5% es 15.507, por lo tanto la hipótesis H_0 “la ley es de Poisson de parámetro $\lambda = 4.68$ ” se acepta.

Ajuste a una ley continua

Las notas obtenidas en un examen por 200 estudiantes se reparte de la manera siguiente:

Nota n_i	10	11	12	13	14	15	16	17	18	19	20	21	22	23
nº estudiantes de nota n_i	1	5	10	15	20	27	33	30	20	15	12	7	4	1

H_0 : Las calificaciones se distribuyen normalmente.

La estimación de los parámetros desconocidos es $\bar{x} = 16.29$, $s = 2.611$.

Considerando las clases $] - \infty, 9.5[$, $]9.5, 10.5[$, \dots , $]22.5, 23.5[$, $]23.5, \infty[$, calculamos las probabilidades p_i de cada una de estas clases, bajo la hipótesis H_0 , o sea $N(16.29, (2.611)^2)$.

x_i	9.5	10.5	11.5	12.5	13.5	14.5	15.5	16.5	17.5	18.5	19.5	20.5	21.5	22.5	23.5
$(x_i - \bar{x})$	-6.79	-5.79	-4.79	-3.79	-2.79	-1.79	-0.79	0.21	1.21	2.21	3.21	4.21	5.21	6.21	7.21
$(x_i - \bar{x})/s$	-2.60	-2.22	-1.84	-1.45	-1.07	-0.69	-0.30	0.08	0.46	0.85	1.23	1.61	2.00	2.38	2.76
$F(t_i)$	0.005	0.013	0.033	0.073	0.142	0.245	0.382	0.532	0.677	0.802	0.891	0.946	0.977	0.991	0.997
p_i	0.005	0.008	0.020	0.040	0.069	0.103	0.137	0.150	0.145	0.125	0.089	0.055	0.031	0.014	0.006
$200p_i$	1	1.6	4	8	13.8	20.6	27.4	30	29	25	17.8	11	6.2	2.8	1.2

Es recomendable agrupar las dos primeras clases, así como las tres últimas. En la tabla $t_i = (x_i - \bar{x})/s$.

En la tabla siguiente se da el efectivo teórico así como todos los elementos para el cálculo de K_n .

Tenemos aquí once clases y dos parámetros estimados a partir de la muestra, i.e. $\nu = 8$ g.d.l.

	x_i	np_i	$ x_i - np_i $	$(x_i - np_i)^2$	$\frac{(x_i - np_i)^2}{np_i}$
0, 11.5	6.0	6.6	0.6	0.36	0.055
11.5, 12.5	10	8.0	2	4.0	0.500
12.5, 13.5	15	13.8	1.2	1.44	0.104
13.5, 14.5	20	20.6	0.6	0.36	0.017
14.5, 15.5	27	27.4	0.4	0.16	0.006
15.5, 16.5	33	30.0	3.0	9.0	0.300
16.5, 17.5	30	29.0	1.0	1.0	0.034
17.5, 18.5	20	25.0	5.0	25.0	1.000
18.5, 19.5	15	17.8	2.8	7.84	0.440
19.5, 20.5	12	11.0	1.0	1.0	0.090
20.5, 30	12	10.8	1.2	1.44	0.133

$$K_n = 2.679$$

La cota para la prueba es 15.507, con $\alpha = 0.05$. Se acepta la hipótesis H_0 .

3.9.2 Prueba de homogeneidad

Comparación de dos proporciones observadas: tabla 2×2

Se tienen dos muestras independientes de tamaño N_1 y N_2 . Se separan los individuos: los que tienen el carácter C y los que poseen el carácter complementario C' . Así tenemos la tabla:

	carácter C	carácter C'	Total
1ª muestra	n_1	n'_1	N_1
2ª muestra	n_2	n'_2	N_2

$$N$$

Hipótesis Las dos muestras son extraídas de una misma población donde cada individuo tiene la probabilidad p de tener el caracter C , i.e. p se estima por $\hat{p} = \frac{n_1 + n_2}{N}$.

Para la 1ª muestra el ajuste se hace con respecto a la binomial $B(N_1, \hat{p})$ y la segunda muestra se ajusta a $B(N_2, \hat{p})$, entonces:

$$\chi^2 = \frac{(n_1 - N_1\hat{p})^2}{N_1\hat{p}\hat{q}} + \frac{(n_2 - N_2\hat{p})^2}{N_2\hat{p}\hat{q}},$$

con $q = 1 - p$. Cada miembro de la suma es un χ_1^2 y son independientes. Hay un parámetro estimado, por lo tanto $\nu = 2 - 1 = 1$.

La hipótesis se rechaza si la cantidad $\chi^2 > \ell_{1,\alpha}$.

3.9.3 Comparación de k proporciones observadas

Se consideran k muestras y dos caracteres incompatibles C y C' .

	caracter C	caracter C'	Total
1ª muestra	n_1	n'_1	N_1
\vdots	\vdots	\vdots	\vdots
i^a muestra	n_i	n'_i	N_i
\vdots	\vdots	\vdots	\vdots
k^a muestra	n_k	n'_k	N_k
Total	M	M'	N

Hipótesis Las muestras son extraídas de una misma población de probabilidad p de tener el caracter C .

El valor de p es estimado por $\hat{p} = \frac{\sum_{i=1}^k n_i}{N} = \frac{M}{N}$. Así tenemos que $\chi^2 = \sum_{i=1}^k \frac{(n_i - N_i\hat{p})^2}{N_i\hat{p}\hat{q}}$, se distribuye asintóticamente como un χ_{k-1}^2 y se rechaza H_0 si $\chi^2 > \ell_{k-1,\alpha}$.

3.9.4 Caso general

Se consideran k muestras y C_1, \dots, C_r caracteres incompatibles.

	C_1	\dots	C_j	\dots	C_r	Total
1ª muestra	n_{11}	\dots	n_{1j}	\dots	n_{1r}	N_1
\vdots	\vdots		\vdots		\vdots	\vdots
i^a muestra	n_{i1}	\dots	n_{ij}	\dots	n_{ir}	N_i
\vdots	\vdots		\vdots		\vdots	\vdots
k^a muestra	n_{k1}	\dots	n_{kj}	\dots	n_{kr}	N_k
Total	M_1	\dots	M_j	\dots	M_r	N

Hipótesis Todas las muestras son extraídas de la misma población, de parámetros p_1, \dots, p_r .

Los parámetros son estimados por $\hat{p}_j = \frac{M_j}{N}$. El ajuste de cada muestra es una ley multinomial de parámetros $(N_i; \hat{p}_1, \dots, \hat{p}_r)$, $i = 1, \dots, k$.

Para cada muestra se establece un χ^2 . Por ejemplo para la i^{ava} muestra es:

$$\chi_i^2 = \sum_{j=1}^r \frac{(n_{ij} - N_i \hat{p}_j)^2}{N_i \hat{p}_j}.$$

Así tenemos $\chi^2 = \sum_{i=1}^k \chi_i^2 = \sum_{i=1}^k \sum_{j=1}^r \frac{(n_{ij} - N_i \hat{p}_j)^2}{N_i \hat{p}_j}$, con $k(r-1) - (r-1) = (k-1)(r-1)$ g.d.l. y se rechaza H_0 si:

$$\chi^2 > \ell_{(k-1)(r-1), \alpha}.$$

3.9.5 Prueba de independencia de caracteres

Caso 2×2 Se consideran dos caracteres R y C y una muestra de tamaño N . Se clasifican los individuos según si poseen o no los caracteres R o C .

	C	no C	Total
R	n_1	n'_1	N_1
no R	n_2	n'_2	N_2
Total	M_1	M_2	N

Hipótesis Los caracteres R y C son independientes.

La probabilidad de tener el caracter C , se estima por $\hat{p}_c = \frac{M_1}{N}$ y la probabilidad de tener el caracter R , se estima por $\hat{p}_r = \frac{N_1}{N}$.

La distribución de la muestra se ajusta a una ley multinomial a 4 categorías. Las probabilidades de

pertenecer a cada categoría son:

	C	$\text{no } C$
R	$\frac{M_1 N_1}{N^2}$	$\frac{M_2 N_1}{N^2}$
$\text{no } R$	$\frac{M_1 N_2}{N^2}$	$\frac{M_2 N_2}{N^2}$

Se obtiene así un χ^2 , con $4 - 2 - 1 = 1$ g.d.l., pues se estiman dos parámetros. Calculando $K_n = \sum \frac{(o - t)^2}{t}$ se rechaza H_0 , si $K_n > \ell_{1,\alpha}$.

Caso $k \times r$ En este caso tenemos:

– un caracter con r modalidades 2 a 2 incompatibles C_1, \dots, C_r

– un caracter con k modalidades 2 a 2 incompatibles R_1, \dots, R_k .

Hipótesis Los caracteres R y C son independientes.

	C_1	\dots	C_i	\dots	C_r	Total
R_1	n_{11}	\dots	n_{i1}	\dots	n_{r1}	N_1
\vdots	\vdots		\vdots		\vdots	\vdots
R_j	n_{1j}	\dots	n_{ij}	\dots	n_{rj}	N_j
\vdots	\vdots		\vdots		\vdots	\vdots
R_k	n_{1k}	\dots	n_{ik}	\dots	n_{rk}	N_k
Total	M_1	\dots	M_j	\dots	M_r	N

Se estima la probabilidad de pertenecer a la modalidad C_i : $\frac{M_i}{N}$

Se estima la probabilidad de pertenecer a la modalidad R_j : $\frac{N_j}{N}$

Los efectivos teóricos de cada casilla se calculan (bajo la hipótesis H_0) por:

$$N \frac{M_i}{N} \frac{N_j}{N} = \frac{M_i N_j}{N}.$$

El ajuste se hace por una ley multinomial a kr categorías:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^k \frac{\left(n_{ij} - \frac{M_i N_j}{N}\right)^2}{\frac{M_i N_j}{N}},$$

con $\nu = kr - 1 - (r - 1) - (k - 1) = (r - 1)(k - 1)$ g.d.l.

Ejemplo La tabla siguiente da la clasificación de 6800 personas sobre las cuales se ha observado a

la vez el color de ojos y del cabello:

		cabello				Total
		rubio	café	negro	rojo	
ojos	azul	1768	807	189	47	2811
	gris verde	946	1387	746	53	3132
	café	115	438	228	16	857
	Total	2829	2632	1223	116	6800

¿Se puede afirmar con un 5% de error que el color de ojos y el color del cabello son independientes?

La tabla de efectivos teóricos, bajo la hipótesis de independencia es:

		cabello			
		rubio	café	negro	rojo
ojos	azul	*1169	1088	506	48
	gris verde	1303	1212	503	54
	café	357	332	154	14

Por ejemplo $1169 = \frac{2811 \times 2829}{6800}$.

Se forma así un $\chi^2 = \sum \frac{(o-t)^2}{t}$, con $\nu = 6$ g.d.l. El valor χ^2 es obtenido por:

$$\chi^2 = \frac{(1768 - 1169)^2}{1169} + \frac{(1088 - 807)^2}{807} + \dots + \frac{(16 - 14)^2}{14} = 1075,$$

por lo tanto (para $\alpha = 0.05$, $\ell_{6,\alpha} = 12.52$) se rechaza la hipótesis y se concluye que hay dependencia.

Observaciones

– Hemos visto que en los dos casos (independencia-homogeneidad) la utilización de la prueba χ^2 es parecida y que las fórmulas son prácticamente las mismas.

– No debe confundirse un modelo con dos clases y varias muestras con un modelo de dos muestras y varias clases. Las hipótesis no son las mismas y la prueba no tiene el mismo sentido.

3.9.6 Método donde se minimiza aproximadamente el χ^2 de una muestra

Para una muestra de tamaño n , el χ^2 se define por $\chi^2 = \sum_{j=1}^k \frac{(x_j - np_j)^2}{np_j}$, donde el evento j -ésimo se produce X_j veces, con $\sum_{j=1}^k X_j = n$ y $\sum_{j=1}^k p_j = 1$.

Si los p_1, \dots, p_k dependen de ℓ parámetros independientes $\theta_1, \dots, \theta_\ell$, se toma como estimación de

estos parámetros, los valores que minimizan aproximadamente a χ^2 , por lo que para $r = 1, \dots, \ell$:

$$\frac{\partial \chi^2}{\partial \theta_r} = -2 \sum_{j=1}^k \left[\frac{X_j - np_j}{p_j} + \frac{(X_j - np_j)^2}{2np_j^2} \right] \frac{\partial p_j}{\partial \theta_r} = 0.$$

La aproximación consiste en despreciar el segundo término en los paréntesis cuadrados.

Así tenemos, $-\sum_{j=1}^k \frac{np_j}{p_j} \frac{\partial p_j}{\partial \theta_r} = -n \frac{\partial}{\partial \theta_r} \left(\sum_{j=1}^k p_j \right) = 0$ y tomamos los valores de los parámetros θ_r tales que:

$$\sum_{j=1}^k \frac{(X_j - np_j)}{p_j} \frac{\partial p_j}{\partial \theta_r} = \sum_{j=1}^k \frac{X_j}{p_j} \frac{\partial p_j}{\partial \theta_r} = 0, \quad r = 1, \dots, \ell. \quad (1)$$

Ejemplo 1 Se puede estimar p_1, \dots, p_k , a partir de una muestra de orden $n = \sum_{j=1}^k X_j$. Tomamos los parámetros p_1, \dots, p_{k-1} y las ecuaciones (1) se escriben:

$$\frac{X_i}{p_i} = \frac{X_k}{p_k}, \quad i = 1, \dots, k-1,$$

de donde se tiene $\frac{X_1}{p_1} = \dots = \frac{X_k}{p_k} = \frac{n}{1}$ y $\hat{p}_i = \frac{X_i}{n}$, $i = 1, \dots, k$.

Ejemplo 2 Se da una partición de Ω en k eventos E_1, \dots, E_k , la probabilidad p_i de E_i es desconocida. Se quiere estimar p_i , con la ayuda de s muestras de orden N_1, \dots, N_s . El evento E_j se produce X_{ij} veces en la muestra i , o sea $\sum_{j=1}^k X_{ij} = N_i$.

Así tenemos:

$$\chi^2 = \sum_{i=1}^s \sum_{j=1}^k \frac{(X_{ij} - N_i \theta_j)^2}{N_i \theta_j},$$

con $\theta_j = p_j$, ligadas por la relación $\sum_{j=1}^k \theta_j = 1$. Las ecuaciones (1), con $r = 1, \dots, k-1$, se escriben:

$$\sum_{i=1}^s \left(\sum_{j=1}^k \frac{X_{ij}}{\theta_j} \frac{\partial \theta_j}{\partial \theta_r} \right) = 0.$$

Como $\theta_k = 1 - \theta_1 - \dots - \theta_{k-1}$, definiendo $M_j = \sum_{i=1}^s X_{ij}$, $N = \sum_{j=1}^k M_j = \sum_{i=1}^s N_i$, la ecuación anterior se escribe:

$$\sum_{i=1}^s \left(\frac{X_{ir}}{\theta_r} - \frac{X_{ik}}{\theta_k} \right) = \frac{M_r}{\theta_r} - \frac{M_k}{\theta_k} = 0 \implies \frac{M_1}{\theta_1} = \dots = \frac{M_k}{\theta_k} = N \text{ y } \hat{p}_i = \frac{M_i}{N}.$$

Ejemplo 3 Se dan dos particiones de $\Omega = \bigcup_{i=1}^s U_i = \bigcup_{j=1}^k V_j$ y consideramos que U_i es independiente de V_j . Denotamos α_i, β_j las probabilidades desconocidas de estos eventos; la probabilidad de U_i y

V_j es $\alpha_i \cdot \beta_j$ y queremos estimar las probabilidades α_i y β_j , con una muestra de tamaño N , donde U_i y V_j se producen X_{ij} veces.

Los $s + k - 2$ parámetros son $\alpha_1, \dots, \alpha_{s-1}$ y $\beta_1, \dots, \beta_{k-1}$.

Las ecuaciones se escriben:

$$\sum_{i=1}^s \sum_{j=1}^k \frac{X_{ij} - N\alpha_i\beta_j}{\alpha_i\beta_j} \frac{\partial(\alpha_i\beta_j)}{\partial\alpha_r} = \sum_{j=1}^k \frac{X_{rj} - n\alpha_r\beta_j}{\alpha_r} = 0, \quad r = 1, 2, \dots, s-1,$$

$$\sum_{i=1}^s \sum_{j=1}^k \frac{X_{ij} - N\alpha_i\beta_j}{\alpha_i\beta_j} \frac{\partial(\alpha_i\beta_j)}{\partial\beta_t} = \sum_{i=1}^s \frac{X_{it} - n\alpha_i\beta_t}{\beta_t} = 0, \quad t = 1, 2, \dots, k-1,$$

por lo que $\hat{\alpha}_r = \frac{M_r}{N}$ $\hat{\beta}_t = \frac{N_t}{N}$.

Ejemplo 4 Buscamos estimar (μ, θ) de una ley normal $N(\mu, \theta)$ i.e. $\sigma^2 = \theta$. Se considera una partición de \mathbb{R} en k intervalos $I_1 =]-\infty, x_1[$, $I_2 = [x_1, x_2[$, \dots , $I_k = [x_{k-1}, +\infty[$ y una muestra de orden n , donde la variable toma X_j veces sus valores en el intervalo I_j . La probabilidad del evento “la variable toma sus valores en I_j ” es:

$$p_j = \int_{x_{j-1}}^{x_j} f(x) dx,$$

donde $f(x) = \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{1}{2\theta}(x-\mu)^2}$.

Calculando $\frac{\partial p_i}{\partial \mu}$, $\frac{\partial p_j}{\partial \theta}$ y sustituyendo en (1), se tiene que:

$$\hat{\mu} = \frac{1}{n} \sum_{j=1}^k X_j \frac{\int_{x_{j-1}}^{x_j} x f(x) dx}{\int_{x_{j-1}}^{x_j} f(x) dx}, \quad \hat{\theta} = \frac{1}{n} \sum_{j=1}^k X_j \frac{\int_{x_{j-1}}^{x_j} (x - \hat{\mu})^2 f(x) dx}{\int_{x_{j-1}}^{x_j} f(x) dx}.$$

Para simplificar el cálculo numérico, se hace un desarrollo limitado de estas expresiones:

– suponiendo que $X_1 = X_k = 0$,

– suponiendo que I_2, \dots, I_{k-1} tienen el mismo largo Δ y por centro $\zeta_j = \frac{1}{2}(x_j + x_{j-1})$ se encuentra:

$$\hat{\mu} = \frac{1}{n} \sum_{j=2}^{k-1} X_j \zeta_j + o(\Delta), \quad \text{si } \Delta \rightarrow 0$$

$$\hat{\theta} = \hat{\sigma}^2 = \frac{1}{n} \sum_{j=2}^{k-1} X_j (\zeta_j - a_1) - \frac{\Delta^2}{12} + \dots, \quad a_1 = \frac{1}{n} \sum_{j=2}^{k-1} X_j \zeta_j.$$

3.9.7 Utilización de la estadística χ^2

Sea E_1, \dots, E_k una partición de Ω , con probabilidades respectivas p_1, \dots, p_k .

Caso 1 Si se conocen los valores de los p_i , $i = 1, \dots, k$ y una muestra de orden n , entonces se tiene que el χ^2 de una muestra:

$$\chi^2 = \sum_{j=1}^k \frac{(X_j - np_j)^2}{np_j},$$

sigue una ley que tiende “vagamente”, cuando $n \rightarrow \infty$, hacia χ_{k-1}^2 , entonces dado un nivel α se encuentra un intervalo, donde χ^2 se encuentra con una probabilidad $1 - \alpha$.

Caso 2 Consideremos una muestra de orden n y se estima las probabilidades p_j por el método aproximado y queremos encontrar un intervalo, donde el χ^2 se encuentra con una probabilidad $1 - \alpha$, entonces es falso que la variable aleatoria $\chi^2 = \sum_{j=1}^k \frac{(X_j - n\hat{p}_j)^2}{n\hat{p}_j}$, se distribuya asintóticamente como una χ_{k-1}^2 . En el ejemplo 1, se tiene que $\chi^2 = 0$.

Se utiliza pues el teorema de Cramer y se tiene “Se consideran s muestras de orden N_j y se supone que E_i se produce X_{ij} veces en la muestra j ”.

La probabilidad p_i de E_i depende de ℓ parámetros y se estima \hat{p}_i , por las ecuaciones (1). Se supone

$N = \sum_{j=1}^s N_j$ grande, entonces:

$$\chi^2 = \sum_{i=1}^k \sum_{j=1}^s \frac{(X_{ij} - N_j \hat{p}_i)^2}{N_j \hat{p}_i},$$

se distribuye asintóticamente como un $\chi_{ks-s-\ell}^2$.

– En el ejemplo 2 tenemos que $\chi^2 = \sum_{i=1}^s \sum_{j=1}^k \frac{\left(X_{ij} - \frac{N_i M_j}{N}\right)^2}{\frac{N_i M_j}{N}}$ se distribuye aproximadamente como

$$\chi_{ks-(k-1)-s}^2 = \chi_{(k-1)(s-1)}^2.$$

– En el ejemplo 3 tenemos que $\chi^2 = \sum_{i=1}^s \sum_{j=1}^k \frac{(X_{ij} - N\hat{\alpha}_j \hat{\beta}_j)^2}{N\hat{\alpha}_j \hat{\beta}_j}$ se distribuye asintóticamente como

$$\chi_{ks-(k+s-2)-1}^2 = \chi_{(k-1)(s-1)}^2.$$

3.10 Ejercicios

1. Sea X_1, \dots, X_n una muestra de densidad $f(x)$ positiva sobre \mathbb{N} . Se desea probar la hipótesis $H_0: f(x) = e^{-1}/x!$, $x = 0, 1, \dots$, contra la hipótesis $H_1: f(x) = \left(\frac{1}{2}\right)^{x+1}$, $x = 0, 1, \dots$. Utilizando el lema de Neyman–Pearson encontrar una prueba Φ para el caso $k = 1$, $n = 1$. Grafique la potencia.
2. Sea X_1, \dots, X_n una muestra de distribución $N(\theta, 1)$. Se quiere probar la hipótesis $H_0: \theta = 0$ contra

- $H_1: \theta = -1$. Probar que la prueba U.M.P. de H_0 contra H_1 utiliza la estadística \bar{X} y que si $n = 25$ y $\alpha = 0.05$, la potencia de la prueba es 0.999 cuando H_1 es cierta.
3. Sea X_1, \dots, X_{10} una muestra de tamaño 10 de distribución $N(0, \sigma^2)$. Determinar una prueba U.M.P. de nivel $\alpha = 0.05$, para probar $H_0: \sigma = 1$ contra $H_1: \sigma_1 > 1$. ¿Es ésta la mejor prueba de H_0 contra $H_1: \sigma > 1$?
4. Sea X_1, \dots, X_n una muestra de tamaño n , de densidad $f(x, \theta) = \theta x^{\theta-1}$, $0 < x < 1$. Probar que la prueba U.M.P. de $H_0: \theta = 1$ contra $H_1: \theta = 2$ depende de $\prod_{i=1}^n X_i$. Explícite la prueba Φ .
5. Sea X_1, \dots, X_{10} una muestra de densidad $f(x, \theta) = \theta^x(1 - \theta)^{1-x}$, $x = 0, 1$ y sea la hipótesis $H_0: \theta = \frac{1}{4}$ contra $H_1: \theta < \frac{1}{4}$. Se rechaza H_0 si $\sum_{i=1}^{10} x_i \leq 1$. Determinar la potencia de la prueba para $0 < \theta \leq \frac{1}{4}$.
6. Sea X una variable aleatoria con densidad $f(x, \theta) = \frac{1}{\theta}$, $0 < x < \theta$. Sea $Y_1 < Y_2 < Y_3 < Y_4$ las estadísticas de orden de una muestra de tamaño 4, de distribución igual a X . Sea y_4 el valor observado de Y_4 . Se rechaza $H_0: \theta = 1$ contra $H_1: \theta \neq 1$, si $y_4 \leq \frac{1}{2}$ o $y_4 \geq 1$. Determinar la potencia de la prueba.
7. Considere una distribución normal $N(\theta, 4)$, la hipótesis $H_0: \theta = 0$ se rechaza y se acepta la hipótesis alternativa $H_1: \theta > 0$, si el valor observado de la muestra \bar{x} , de tamaño $n = 25$ es mayor o igual que $\frac{3}{5}$. Determinar la potencia de la prueba.
8. Se consideran dos distribuciones normales independientes $N(\mu_1, 400)$ y $N(\mu_2, 225)$. Sea $\theta = \mu_1 - \mu_2$ y sean \bar{X} y \bar{Y} las medias muestrales. Se considera que las muestras son independientes de tamaño n . Se rechaza $H_0: \theta = 0$ y se acepta $H_1: \theta > 0$ si $\bar{X} - \bar{Y} \geq c$. Si $\beta(\theta)$ es la potencia de la prueba, determinar n y c tal que $\beta(0) = 0.05$ y $\beta(10) = 0.90$ aproximadamente.
9. Sea X_1, \dots, X_n una muestra de distribución $N(\theta, 16)$. Determinar el tamaño de la muestra n y una prueba U.M.P. de $H_0: \theta = 25$ contra $H_1: \theta < 25$, con potencia $\beta(\theta)$ tal que $\beta(25) = 0.10$ y $\beta(23) = 0.90$.
10. Se consideran las variables aleatorias $X_{11}, X_{12}, \dots, X_{1n}$ independientes uniformes sobre $[0, \theta_1]$, con $\theta_1 > 0$ y $X_{21}, X_{22}, \dots, X_{2m}$ independientes uniformes sobre $[0, \theta_2]$, con $\theta_2 > 0$. Las $n + m$ variables

son independientes. Se consideran las variables aleatorias ordenadas:

$$M_1 = X_{1(n)}, \quad M_2 = X_{2(m)}.$$

a) Determinar la ley conjunta de M_1 y M_2 .

b) Determinar la función de repartición y de densidad de $Y = \frac{M_2}{M_1}$.

c) Se propone probar la hipótesis $H_0: \frac{\theta_2}{\theta_1} = 1$ contra $H_1: \frac{\theta_2}{\theta_1} > 1$. Se conviene en aceptar H_0 si $Y \leq y_\alpha$, con un error de primera especie α . Determinar y_α para $\alpha = 0.05$ y $n = m = 20$. Estudiar la función potencia en función de $u = \frac{\theta_2}{\theta_1}$.

11. Se considera la estructura $(\mathbb{N}, \mathcal{P}(\mathbb{N}), \{P_{p,\lambda} : p \in [0, 1], \lambda > 0\})^n$, donde $P_{p,\lambda}$ se define por:

$$P_{p,\lambda}(\{i\}) = \begin{cases} 1 - p + pe^{-\lambda} & \text{si } i = 0 \\ pe^{-\lambda} \frac{\lambda^i}{i!} & \text{si } i \neq 0. \end{cases}$$

Sean X_i las aplicaciones coordenadas y se define $N_0 = \sum_{i=1}^n \mathbb{1}_{[X_i=0]}$.

a.1) Determinar la ley de N_0 y $E(N_0)$.

a.2) Dar la función de verosimilitud de la muestra (X_1, \dots, X_n) .

a.3) Demostrar que en los dos casos siguientes, existe una estadística suficiente que se determinará:

α) λ fijo, p arbitrario,

β) p fijo, λ arbitrario.

a.4) En el caso α) demostrar que la estadística es suficiente y completa.

b) Supongamos λ fijo.

b.1) Determinar un estimador sin sesgo y convergente de p , función de N_0 .

b.2) Dar otro estimador sin sesgo de p , función de la media muestral.

b.3) Comparar los estimadores.

c) Supongamos siempre λ fijo, y se consideran las hipótesis $H_0: p = 1$, $H_1: p < 1$.

c.1) Construir una prueba U.M.P. de H_0 contra H_1 de nivel α .

c.2) Aplicación numérica: $n = 10$, $\lambda = \log 2$, $\alpha = 0.02$.

12. Sea la estructura estadística $(\mathbb{R}^2, \mathfrak{B}_{\mathbb{R}^2}, \{P_\theta: \theta > 0\})$, donde P_θ es una ley de probabilidad sobre $(\mathbb{R}^2, \mathfrak{B}_{\mathbb{R}^2})$ de densidad con respecto a la medida de Borel:

$$f_\theta(x, y) = \begin{cases} \frac{20}{\theta^5}(y-x)^3 & \text{si } 0 \leq x \leq y \leq \theta \\ 0 & \text{si no.} \end{cases}$$

Se denota X y Y las aplicaciones coordenadas de \mathbb{R}^2 en \mathbb{R} .

- a.1) Verificar que f_θ es una densidad.
- a.2) Se define $X' = \theta - Y$, $Y' = \theta - X$. Determinar la ley de (X', Y') . Comparar las leyes de probabilidad de X y $\theta - Y$.
- a.3) Determinar la densidad, función de repartición, esperanza y varianza de cada una de las variables aleatorias X y Y . ¿Son independientes? Determinar el coeficiente de correlación.
- a.4) Demostrar que Y es una estadística suficiente y completa. Determinar $E^Y(X)$.
- a.5) Determinar el estimador $\hat{\theta}$ de θ por el método de máxima verosimilitud. Deducir de $\hat{\theta}$ un estimador corregido θ^* que sea sin sesgo y de varianza mínima.
- b) Se considera la estructura estadística $(\mathbb{R}^2, \mathfrak{B}_{\mathbb{R}^2}, \{P_\theta: \theta \in \mathbb{R}_+^*\})^n$ y se tienen n pares (X_i, Y_i) independientes, cada uno de ley P_θ . Se define $U_n = \inf_{1 \leq i \leq n} X_i$, $V_n = \sup_{1 \leq i \leq n} Y_i$.
- b.1) Determinar la ley de V_n .
- b.2) Estudiar la convergencia de V_n , cuando $n \rightarrow \infty$. Se examinará todos los modos de convergencia.
- b.3) Demostrar que $n(V_n - \theta)$ converge en ley, cuando $n \rightarrow \infty$ y precise su límite.
- b.4) Determinar la ley conjunta de (U_n, V_n) . Se podrá comenzar por el cálculo de la probabilidad del evento $[X \leq U_n \leq V_n \leq Y]$.
- b.5) Demostrar que V_n es suficiente.

13. Se considera la estructura estadística $(\mathbb{R}^2, \mathfrak{B}_{\mathbb{R}^2}, \{P_\rho: \rho > 0\})^n$, donde P_ρ es la ley de probabilidad definida por:

$$f(x, y, \rho) = \begin{cases} \frac{1}{\pi\rho^2} & \text{si } x^2 + y^2 \leq \rho^2 \\ 0 & \text{si } x^2 + y^2 > \rho^2. \end{cases}$$

Sea $Z_i = (X_i, Y_i)$, la i^{ava} aplicación coordenada de $(\mathbb{R}^2)^n$ en \mathbb{R}^2 .

Sea $H_0 = \{P_1\} : \rho = 1$, $H_1 = \{P_\rho : \rho > 1\}$, $H'_1 = \{P_\rho : \rho < 1\}$, $H''_1 = \{P_\rho : \rho \neq 1\}$.

- a) Demostrar que la estadística $T = \sup(X_i^2 + Y_i^2)$ es suficiente.
- b) Construir una prueba U.M.P. de nivel α de H_0 contra H_1 . Dar la potencia de la prueba.
- c) Determinar una prueba estocástica de nivel α , la más potente de H_0 contra $\{P_{\rho_1}\}$, donde $\alpha^{1/2n} \leq \rho_1 \leq 1$. Precise la potencia y deducir una prueba determinista U.M.P. de nivel α de H_0 contra H'_1 .
- d) Utilizando los resultados anteriores construir una prueba de nivel α , de H_0 contra H''_1 .
- e) ¿Las pruebas obtenidas son admisibles?
14. a) El par $Z = (X, Y)$ es distribuido según una ley bi-normal de media nula, varianza unidad, y coeficiente de correlación ρ .
- a.1) Sea $U = X - \rho Y$, $V = \sqrt{1 - \rho^2} Y$, dar ley de U y de V . Demostrar que U y V son independientes.
- a.2) Calcular $E(U^2 V^2)$, $E(UV^3)$, $E(Y^4)$ y $E(V^4)$. Deducir que $E(X^2 Y^2) = 1 + 2\rho^2$.
- b.1) Dar la función de verosimilitud de una muestra Z_1, \dots, Z_n de variables aleatorias independientes de igual distribución igual que Z .
- b.2) ¿Existe una estadística suficiente con valores en \mathbb{R}^2 ? ¿Con valores en \mathbb{R} ? ¿Existe un estimador eficaz de ρ ?
- c) Sea $R = \frac{1}{n} \sum_{i=1}^n X_i Y_i$.
- c.1) Demostrar que R es un estimador sin sesgo y convergente de ρ . Calcular $\text{var}(R)$.
- c.2) Determinar la ley asintótica de $\sqrt{n}(R - \rho)$ y deducir la ley de distribución de $\frac{n(R - \rho)^2}{1 + \rho}$.
- c.3) Sea $T \sim \chi^2$ a un g.d.l., se define χ_α , donde α es tal que $0 < \alpha < 1$, por $P(T \geq \chi_\alpha) = \alpha$. Determinar un intervalo de confianza para ρ , de coeficiente de seguridad asintótico $1 - \alpha$.
- c.4) Sea $\hat{\rho}$ el estimador de ρ obtenido por el máximo de verosimilitud.
- c.4.i) Demostrar que $\hat{\rho}$ es una raíz de la ecuación de tercer grado de la forma $\lambda^3 + p\lambda + q = 0$, definiendo $\lambda = \rho - \frac{1}{3n} \sum_{i=1}^n X_i Y_i$. Demostrar que para n grande, ésta ecuación admite una única solución real. Recuerde que la ecuación $\lambda^3 + p\lambda + q = 0$ admite una sola raíz real si $4p^3 + 27q^2 \geq 0$, que se verifica en

particular si $p > 0$.

c.4.ii) Demostrar que cuando $n \rightarrow \infty$, $\text{var } \hat{\rho} \approx \frac{(1-\rho^2)^2}{n(1+\rho^2)}$. ¿Cuál es la eficacia asintótica de $\hat{\rho}$ y la de R ? Comparar los dos estimadores.

15. a) Se consideran n variables aleatorias continuas Y_1, \dots, Y_n de densidad conjunta g definida por:

$$g(\mathbf{y}, \theta) = g(y_1, \dots, y_n, \theta) = \begin{cases} \theta^n n! \exp(-\theta \sum_{i=1}^n y_i) & \text{si } 0 < y_1 < \dots < y_n \\ 0 & \text{si no,} \end{cases}$$

con $\theta > 0$. Sea $Z_1 = Y_1, Z_i = Y_i - Y_{i-1}, i = 2, \dots, n$.

a.1) Dar la densidad conjunta de Z_1, \dots, Z_n . Demostrar que son independientes.

a.2) Sea $U_i = (n-i+1)Z_i, i = 1, \dots, n$. Demostrar que las variables aleatorias U_i son independientes de ley $\Gamma(1, \theta)$.

Deducir la ley de probabilidad de la variable aleatoria $T_k = \sum_{i=1}^{k-1} Y_i + (n-k+1)Y_k$, para $1 \leq k \leq n$. Exprese T_k en función de los U_i .

a.3) Calcular la función característica o generatriz de momentos de Y_1, \dots, Y_n . Deducir la función característica de T_k y encontrar el resultado anterior.

b) Se considera una muestra de n variables aleatorias independientes de igual distribución, X_1, \dots, X_n , cuya densidad común es f_θ definida por:

$$f(x, \theta) = \begin{cases} \theta e^{-\theta x} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0, \end{cases}$$

con $\theta > 0$ i.e. $X_i \sim \Gamma(1, \theta)$.

Se consideran las hipótesis $H_0: \theta = \theta_0, H_1: \theta = \theta_1$, con $\theta_1 > \theta_0, H'_1: \theta > \theta_0$.

b.1) Construir la mejor prueba de nivel α de la hipótesis H_0 contra H_1 . Demostrar que la prueba es U.M.P. de H_0 contra H'_1 . Se denotará Φ_n la prueba basada en la n -muestra.

b.2) Se consideran las estadísticas de orden relativas a la muestra anterior, es decir $X_{(1/n)}, \dots, X_{(n/n)}$ y denotamos $Y_i = X_{(i/n)}$. Se pretende buscar una prueba de nivel α de H_0 contra H_1 , conociendo únicamente los valores Y_1, \dots, Y_k , con $1 \leq k \leq n$. ¿Cuál es la ley conjunta de Y_1, \dots, Y_k ? Deducir

la mejor prueba de nivel α de H_0 contra H_1 , conociendo sólo las primeros k estadísticas de orden.

Demostrar que la prueba es U.M.P. de H_0 contra H'_1 . Se denotará $\Phi_{k/n}$ la prueba basada en las k primeras estadísticas de orden de la n -muestra.

b.3) Comparar las pruebas Φ_k y $\Phi_{k/n}$ y demostrar que a nivel igual, tienen potencia igual.

b.4) Se puede por ejemplo, aplicar éstos resultados al estudio de la vida media de un material dado.

Si se supone que la duración media de vida es $\frac{1}{\theta}$, ¿a qué corresponden las diferentes hipótesis H_0 , H_1 , H'_1 ? Discutir en qué es equivalente concretamente a someter k objetos a una experiencia y observar todas las duraciones de vida o al contrario observar n objetos en experiencia y retener las k más cortas duraciones de vida. ¿En qué puede ser el segundo tipo de experiencia preferible? Demostrar en qué y cómo esto se precisa por los cálculos de $E(X_{(k/n)})$ y $E(X_{(k/k)})$.

16. Sea $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}}, \{P_\theta: \theta \in \mathbb{R}_+\})^n$, con $n > 1$, una estructura estadística donde P_θ es la ley de probabilidad de densidad respecto a la medida de Lebesgue de \mathbb{R} , f_θ definida por:

$$f_\theta(x) = \frac{1}{2\theta} e^{-\left|\frac{x}{\theta}\right|}, \quad x \in \mathbb{R}.$$

Sean X_1, \dots, X_n las aplicaciones coordenadas de esta estructura.

a) ¿Cuáles son la esperanza matemática y la varianza de P_θ ?

b) Sea $S^2 = \frac{1}{n} \sum_{i=1}^n X_i^2$, encontrar una función U de S^2 que sea un estimador sin sesgo y convergente de θ^2 .

c) Determinar una estadística suficiente, así como su ley. Deducir una función V de la estadística, que sea un estimador insesgado y convergente de θ .

d) Determinar si los estimadores U , V son eficaces de θ^2 y θ respectivamente.

e) Construir una puerta U.M.P. de nivel α , ($0 < \alpha < 1$) de la hipótesis $H_0 = \{\theta_0\}$ contra $H =]\theta_0, +\infty[$, $\theta_0 > 0$.

Aplicación numérica Determinar completamente la prueba para $\theta_0 = 4$, $n = 12$, $\alpha = 0.05$.

17. Sea la estructura estadística $(\mathbb{N}, \mathcal{P}(\mathbb{N}), \{P_\theta: \theta \in]0, 1[\})^n$, $n > 1$, donde P_θ es la ley geométrica de parámetro θ . Se denotan X_1, \dots, X_n las aplicaciones coordenadas de esta estructura.

a) Demostrar que existe una estadística suficiente real T y dar su ley.

b) Demostrar que T es completa.

c) Sea $U = \frac{n}{T+n}$, demostrar que U es un estadística convergente de θ .

d) Sea $Z = (\sqrt{n}(T/n + 1)\theta - 1) / \sqrt{1 - \theta}$, demostrar que la ley de Z converge a la ley normal reducida, cuando $n \rightarrow \infty$. Deducir la ley límite de $\sqrt{n}(U - \theta)$.

e) Sean las hipótesis: $H_0 = \{\frac{1}{2}\}$, $H_1 =]\frac{1}{2}, 1[$, $H_2 =]0, \frac{1}{2}[\cup]\frac{1}{2}, 1[$.

Construir una prueba Φ U.M.P. de nivel α , ($0 < \alpha < 1$), de H_0 contra H_1 . ¿Existe una prueba U.M.P. de nivel α de H_0 contra H_2 ?

f) Utilizando la ley límite de Z , construir a partir de Φ una prueba Ψ de H_0 contra H , de nivel asintóticamente igual a α . Explicitar la prueba para $n = 100$, $\alpha = 0.01$.

18. a) Se denota $E(\theta, \sigma)$ la ley de probabilidad sobre $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}})$ absolutamente continua con respecto a la medida de Lebesgue de densidad $x \mapsto \frac{1}{\sigma} e^{-\frac{x-\theta}{\sigma}} \mathbb{1}_{[\theta, +\infty)}(x)$, donde $\theta \in \mathbb{R}$ y $\sigma > 0$ son parámetros.

Sobre un espacio de probabilidad $(\Omega, \mathfrak{A}, P)$, se definen $n > 1$ variables aleatorias independientes reales $(X_i)_{i=1, \dots, n}$, de misma ley $E(\theta, \sigma)$. Se denota $X_{(1)}, \dots, X_{(n)}$ las variables ordenadas correspondientes.

a.1) Demostrar que $Y = X_{(1)}$ sigue una ley E , donde los parámetros se deben determinar.

a.2) Se define $Y = X_{(1)}$, $X_{(i)} - X_{(1)} = Y_i$, $i = 2, \dots, n$ y se denota $Z = \sum_{i=2}^n Y_i$. Justicar el hecho de que $Z = \sum_{i=1}^n [X_{(i)} - X_{(1)}]$.

Dar la densidad conjunta de $X_{(1)}, \dots, X_{(n)}$ y la de Y, Y_2, \dots, Y_n .

Deducir que las variables aleatorias Y y (Y_2, \dots, Y_n) son independientes y que las variables aleatorias Z y Y son independientes, de leyes respectivas $\Gamma\left(n-1, \frac{1}{\sigma}\right)$ y $E(\theta, \sigma/n)$.

a.3) Si las variables aleatorias reales $X \sim E(\theta, \sigma)$, ¿cuál es la ley de $U = X - \theta$? Utilizando los resultados clásicos dar $E(X_i)$, $\text{var}(X_i)$, $E(Y)$, $\text{var}(Y)$, $E(Z)$, $\text{var}(Z)$.

b) Se considera la estructura estadística $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}, \{E(\theta, \sigma) : \theta \in \mathbb{R}, \sigma > 0\}})^n$, $n > 1$, sean X_1, \dots, X_n las aplicaciones coordenadas.

b.1) Demostrar que (Y, Z) con valores en $(\mathbb{R}^2, \mathfrak{B}_{\mathbb{R}^2})$ es suficiente y completa. Demostrar que para σ conocida, Y es suficiente y completa y Z es libre.

b.2) **Estimación de θ y σ**

b.2. α) Demostrar que existen reales a, b, c, d tales que las estadísticas $S = aY + bZ$ y $T = cY + dZ$ estiman sin sesgo σ y θ respectivamente. ¿Son los estimadores encontrados, dentro de los estimadores insesgados aquellos los de más pequeña varianza?

Estudiar, cuando $n \rightarrow \infty$ la convergencia de S y T .

b.2. β) Determinar estimadores de (θ, σ) por el máximo de verosimilitud, $(\hat{\theta}, \hat{\sigma})$.

Comparar los errores cuadráticos $E_{\theta, \sigma}((\hat{\theta} - \theta)^2)$ y $E_{\theta, \sigma}((T - \theta)^2)$, así como $E_{\theta, \sigma}((\hat{\sigma} - \sigma)^2)$ y $E_{\theta, \sigma}((S - \sigma)^2)$.

Discutir los resultados encontrados.

b.3) **Prueba sobre σ**

Sean las hipótesis $H_0: \{E^n(\theta, \sigma): \theta \in \mathbb{R}, \sigma = \sigma_0\}$, $H_1 = \{E^n(\theta, \sigma): \theta \in \mathbb{R}, \sigma = \sigma_1\}$, $\sigma_1 > \sigma_0$,

$H = \{E^n(\theta, \sigma): \theta \in \mathbb{R}, \sigma > \sigma_0\}$. Se fija $\alpha \in]0, 1[$.

b.3. α) Se supone θ conocido.

Demostrar que existe una prueba H_0 contra H , U.M.P. de nivel α .

Determinar la prueba, utilizando una tabla estadística.

Explique por qué la prueba se puede utilizar si θ es desconocido.

b.3. β) θ es no conocido.

Se nota \mathcal{T}_α la familia de pruebas Φ de H_0 contra H , que son tales que $E_{\theta, \sigma_0}(\Phi) = \alpha$, $\forall \theta \in \mathbb{R}$. Demostrar que $\Phi \in \mathcal{T}_\alpha \iff E_{\sigma_0}^Y(\Phi) = \alpha$, donde $E_{\sigma_0}^Y$ designa la esperanza condicionada a Y para el valor σ_0 de σ . ¿Por qué se omitió el parámetro θ ?

Se considera la prueba $\Phi = \mathbb{1}_{[Z > \ell(\alpha)]}$, donde $\ell(\alpha)$ está determinado por la condición que Φ es de nivel α .

Indicar cómo se puede calcular $\ell(\alpha)$ a partir de las tablas.

Demostrar que Φ es U.M.P. de nivel α , para probar H_0 contra H , cuando Y es conocida (es decir, condicionalmente a Y). Deducir que para probar H_0 contra H , Φ es U.M.P. en la familia \mathcal{T}_α .

19. Sea $(\mathbb{R}^n, \mathfrak{B}_{\mathbb{R}^n}, \{P_\theta: \theta \in \Theta\})$, se dice que es una estructura de tipo exponencial sii existe una medida σ -finita μ , tal que $\forall \theta, P_\theta \ll \mu$ y su densidad es $f_\theta(x) = e^{A(\theta) \cdot T(x) + b(\theta) + c(x)}$, $\Theta \subset \mathbb{R}^k$, $A(\theta) \in \mathbb{R}^k$, $T(x) \in \mathbb{R}^k$, $A(\theta) \cdot T(x)$ es el producto escalar.

Demostrar que si se tiene una estructura exponencial, existe una estadística suficiente.

Aplicación Demostrar que las siguientes estructuras son de tipo exponencial y determinar sus estadísticas suficientes:

- a) $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}, N(m, \sigma^2)}; m \in \mathbb{R}, \sigma > 0)^n$
- b) $(\mathbb{N}, \mathfrak{P}(\mathbb{N}), P_\lambda; \lambda > 0)^n$
- c) $(\mathbb{N}, \mathfrak{P}(\mathbb{N}), B(n, p); n \in \mathbb{N}, 0 < p < 1)$
- d) $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}, \Gamma(p, a)}; a > 0, p > 0)^n$.
20. a) Sea la estructura estadística $(\mathbb{N}, \mathcal{P}(\mathbb{N}), P(\lambda), \lambda \geq \lambda_0 > 0)^n$, donde P_λ es la ley de Poisson de parámetro λ .
- Sean X_1, \dots, X_n las aplicaciones coordenadas.
- a.1) Encontrar un estimador T sin sesgo de λ que sea suficiente y completo.
- a.2) Construir una prueba de $H_0: \lambda = \lambda_0$ contra $H_1: \lambda > \lambda_0$, de nivel α . ¿Es esta prueba la mejor?
- b) Sean X_1, \dots, X_n variables aleatorias de distribución exponencial de parámetro $\frac{1}{\theta}$. Encontrar una prueba U.M.P.
21. Sea la estructura estadística $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}, (P_\theta: \theta > 0)})^n$, con P_θ ley uniforme en $[0, \theta]$. Sean X_1, \dots, X_n las aplicaciones coordenadas. Demostrar que $U = \sup_{1 \leq i \leq n} X_i$ y $V = \frac{\sum_{i=1}^n X_i}{U}$ son independientes.
22. Sean X, Y variables aleatorias independientes tales que $X \sim \Gamma(n_1, \theta_1)$ y $Y \sim \Gamma(n_2, \theta_2)$. Sea $U = X + Y$, $V = \frac{X}{X + Y}$.
- a) Determinar la densidad de (U, V) . Estudiar el caso $\theta_1 = \theta_2$.

- b) Determinar la densidad marginal de U (no calcular la integral).
- c) Determinar la densidad de V condicionada a $U = u$. Estudiar el caso $\theta_1 = \theta_2$.
- d) Sea $\theta_2 = \theta$ y $\theta_1 = \theta + \delta$, demostrar que U es suficiente para θ .
- e.1) Sea $H_0: \delta = 0$ y $H_1: \delta > 0$, encontrar una prueba U.M.P. de nivel α de H_0 contra H_1 .

Sugerencia Note que H_0 no es una hipótesis simple pues todos los pares $(\theta, 0) \in H_0$. Debe buscarse una estructura independiente de θ , es decir debe tomarse una estructura suficiente para θ .

Empezar por una prueba $H_0: \delta = 0$ contra $H_1: \delta = \delta_1 > 0$ y probar que es independiente de δ_1 .

e.2) Efectuar la prueba si $n_1 = 10$, $n_2 = 1$, $\alpha = 0,05$.

e.3) Si consideramos $W = \frac{n_1}{n_2} \frac{1-V}{V}$, probar que W es de ley $F(2n_1, 2n_2)$ bajo H_0 , F de Fisher.

23. Sea la estructura estadística $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}, \{P_\theta: \theta > 0\}})^n$, con $P_\theta = \Gamma(1, \theta)$. Sean X_1, \dots, X_n las aplicaciones coordenadas.

a) Determinar la ley de $\sum_{i=1}^n X_i$ y demostrar que $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ es un estimador sin sesgo y convergente de θ .

b) Construir una prueba Φ determinista de $H_0: \theta = 1$, contra $H_1: \theta > 1$.

Aplicación numérica $n = 10$, $\alpha = 0.05$

c) Determinar la potencia para $\theta = 1.5$.

d) Sea $L(X_1, \dots, X_n, \theta)$ la función de verosimilitud, verifique que $T = \prod_{i=1}^n X_i$ es suficiente.

24. Se considera un espacio de probabilidad. Sean E_1 y E_2 dos eventos y sea $p = P(E_1)$, $q = P(E_2)$ y $r = P(E_1 \cap E_2)$, ($0 < p, q < 1$, $0 < r < \inf(p, q)$). Se denota \bar{E}_1 y \bar{E}_2 los eventos contrarios de E_1 y E_2 . Se realizan n pruebas independientes y se designan por A, B, C y D las variables aleatorias cuyos valores observados son respectivamente el número de realizaciones de $E_1 \cap E_2$, $E_1 \cap \bar{E}_2$, $\bar{E}_1 \cap E_2$ y $\bar{E}_1 \cap \bar{E}_2$. Además se define $T = B + C$.

a.1) Determinar la ley de probabilidad de (A, B, C, D) .

a.2) Determinar la ley de probabilidad de (A, T, D) .

a.3) Determinar la ley de probabilidad de T .

a.4) Determinar la ley de probabilidad de (A, B) , condicionada a $T = t$.

a.5) Determinar la ley de probabilidad B condicionada a $T = t$.

Demostrar que cuando p y q son iguales esta ley a.5) no depende de los parámetros p, q, r .

b) Se supone que p y q son iguales.

b.1) Demostrar que la estadística (A, T) es suficiente para (p, r) .

b.2) ¿Cuáles son los estimadores de máxima verosimilitud de p y r ? Calcular la media y la varianza.

c) Nos proponemos construir una prueba de la hipótesis de $H_0: p = q$ contra $H_1: p < q$.

Se efectúan los cambios de parámetros siguientes $\theta = \frac{p+q}{2}, \delta = \frac{p-q}{2}$.

c.1) Si nos basamos sobre la distribución condicionada a $T = t$ de la muestra (A, B, C, D) , demostrar que existe una prueba de nivel α (denotada $\Phi_{\alpha,t}$) de la hipótesis $\delta = 0$ contra la hipótesis $\delta = \delta_1$, ($\delta_1 < 0$), que sea la más potente para cualquiera que sean los valores de los parámetros especificados θ y r por un lado y para cualquier δ_1 ($\delta_1 < 0$) por otro lado.

Precisar la ley de probabilidad de la estadística de la prueba. ¿Sabido que $T = t$, existe una estadística suficiente para δ ?

c.2) Deducir de $\Phi_{\alpha,t}$ una prueba de H_0 contra H_1 de nivel constante α . Discutir sus propiedades.

c.3) Se define $V = \frac{B-C}{\sqrt{B+C}} = \frac{2B-T}{\sqrt{T}}$.

c.3.i) Estudiar bajo H_0 , la distribución condicional (sabiendo que $T = t$) asintótica de V , cuando $t \rightarrow \infty$. Indicar cómo este resultado puede ser utilizado para determinar aproximadamente la prueba $\Phi_{\alpha,t}$.

c.3.ii) Estudiar bajo H_0 la distribución asintótica de V , cuando $n \rightarrow \infty$. Se podrá estudiar el comportamiento de las variables aleatorias $\frac{B-C}{\sqrt{n}}$ y $\frac{B+C}{n}$. Explicar como se puede deducir para n grande, una prueba de H_0 contra H_1 de nivel aproximado α , cuya región crítica es del tipo: $V < \ell$. Comparar este resultado a los precedentes.

d) **Aplicación** Dos pruebas de aptitud Q_1 y Q_2 son cada una puestas a n individuos recogidos al azar e independientes en una población dada. Se anota a cada individuo si da las respuestas exactas a Q_1 y Q_2 . Indicar como se puede utilizar los resultados anteriores para probar la hipótesis según la cual las dos pruebas son de igual dificultad contra la hipótesis Q_1 es más difícil que Q_2 .

Precisar las conclusiones para cada una de las dos situaciones siguientes, con un nivel $\alpha = 0.05$ (con $n = 20$ en el primer caso y $n = 200$ en el segundo caso). Las tablas indican las frecuencias de

aprobación y fracaso.

Q_1	Q_2	aprobación	fracaso	Q_1	Q_2	aprobación	fracaso
aprobación		7	1	aprobación		62	35
fracaso		8	4	fracaso		65	38

25. Sea X una variable aleatoria de función de repartición F , dada por $F(x) = 1 - e^{-\theta x}$, si $x > 0$; $F(x) = 0$, si no, donde $\theta > 0$. Sean X_1, \dots, X_n una muestra de n variables aleatorias independientes distribuidas como X . El verdadero valor del parámetro θ es desconocido y lo denotamos θ_0 .

a) Escribir la verosimilitud L de esta muestra. Determinar el estimador de máxima verosimilitud de θ , que lo denotamos S . Demostrar que S es una estadística suficiente de θ . Calcular la media y la varianza de S . Determinar un estimador sin sesgo S' en función de S . Determinar la distribución límite de $\sqrt{n}(S - \theta_0)$ y de $\sqrt{n}(S' - \theta_0)$, cuando $n \rightarrow \infty$.

b) El intervalo $[0, +\infty[$ lo dividimos en $k + 1$ intervalos $[0, h[$, $[h, 2h[$, $[2h, 3h[$, \dots , $[(k - 1)h, kh[$, $[kh, +\infty[$. Suponemos ahora que el valor tomado por cada variable aleatoria X_i , $i = 1, 2, \dots, n$, no puede ser conocido exactamente. Se sabe solamente a cuál de los $(k + 1)$ intervalos definidos la variable pertenece. En otros términos, a cada muestra X_1, \dots, X_n corresponden $k + 1$ variables aleatorias N_j , $j = 1, 2, \dots, k + 1$, tales que el valor observado de N_j es el número de realizaciones de los X_i pertenecientes al $j^{\text{ésimo}}$ intervalo. Las únicas variables utilizables para estimar θ , son ahora N_1, N_2, \dots, N_{k+1} .

a) Se denota $p_j = P((j - 1)h \leq X \leq jh)$, para $j = 1, \dots, k$ y $P(X \geq kh) = p_{k+1}$.

Expresar los P_j , ($j = 1, 2, \dots, k + 1$), en función de θ y de h .

b) Dar la ley de probabilidad conjunta de las $(k + 1)$ variables aleatorias N_1, N_2, \dots, N_{k+1} .

c) Determinar el estimador T de máxima verosimilitud de θ en estas nuevas condiciones y la distribución límite de $\sqrt{n}(T - \theta_0)$, cuando $n \rightarrow \infty$.

d) ¿Existe un par de estadísticas reales suficientes para θ ?

e) Demostrar que la varianza asintótica de $\sqrt{n}(T - \theta_0)$ es superior a la de $\sqrt{n}(S - \theta_0)$, pero tiende a esta última si $h \rightarrow 0$ y $hk \rightarrow +\infty$.

f) Demostrar que si $hk \rightarrow +\infty$, $P(N_{k+1} = 0) \rightarrow 1$ y que:

$$P\left(h \sum_{j=1}^{k+1} jN_j - hn \leq \sum_{i=1}^n X_i \leq h \sum_{j=1}^{k+1} jN_j\right) \rightarrow 1.$$

Deducir que si $kh \rightarrow +\infty$ y $h \rightarrow 0$, $h \sum_{j=1}^{k+1} jN_j \xrightarrow{\text{Pr}} \sum_{i=1}^n X_i$ y de aquí que $T \xrightarrow{\text{Pr}} S$.

Capítulo 4

Estimación paramétrica

4.1 Estimación paramétrica en \mathbb{R}^r

En muchos problemas de estimación, nos encontramos con la situación de tener una muestra de función de densidad $f(x, \theta)$, donde $\theta \in \mathbb{R}$ es un parámetro (desconocido), cuyo verdadero valor θ_0 es estimado por los elementos de la muestra.

Es nuestro interés considerar dos tipos de estimadores paramétricos: estimadores puntuales y estimadores por intervalos.

En estimación puntual, el estimador del verdadero parámetro θ_0 , es una variable aleatoria T , la cual está en función de los elementos de la muestra.

En estimación por intervalo se considera dos variables aleatorias T_1 y T_2 tales que $T_1 < T_2$ c.p.d., las cuales determinan una probabilidad específica para que θ pertenezca al intervalo aleatorio $[T_1, T_2]$.

Estas ideas referentes a la estimación paramétrica, serán extendidas al caso en que θ es un vector.

Consideramos una estructura estadística $(\Omega, \mathfrak{A}, \{P_\theta : \theta \in \Theta\})$ que está dominada por una medida μ , σ -finita i.e. la función densidad es $f_\theta(x) = f(x, \theta) = L(x, \theta)$ y además suponemos que $[f_\theta > 0]$ no depende de θ . El espacio $\Omega = \mathbb{R}^m$ y la tribu $\mathfrak{A} = B_{\mathbb{R}^m}$ tribu de Borel, $m \geq 1$ y la variable aleatoria $X: (\Omega', \mathfrak{A}', P') \longrightarrow (\Omega, \mathfrak{A}, P)$, $X(\omega') = \omega$.

4.1.1 Caso $\theta \in \mathbb{R}$

Sea X una variable aleatoria con función de densidad $f(x, \theta)$, $\theta \in \Theta \subset \mathbb{R}$. Nos interesamos en un intervalo abierto $\Theta_0 \subset \Theta$ tal que $\theta \in \overset{\circ}{\Theta}_0$, donde θ_0 es el verdadero valor de θ . Supongamos que:

$$\int f_{\theta} d\mu = 1$$

es derivable dos veces con respecto a θ , bajo el signo de integral. Así tenemos que:

$$\frac{d}{d\theta} \int f_{\theta} d\mu = \int \left(\frac{\partial}{\partial \theta} \log f_{\theta} \right) f_{\theta} d\mu = E_{\theta} \left(\frac{\partial}{\partial \theta} \log f_{\theta} \right) = 0 \quad (1)$$

$$\frac{d^2}{d\theta^2} \int f_{\theta} d\mu = \int \left(\frac{\partial^2}{\partial \theta^2} \log f_{\theta} \right) f_{\theta} d\mu + \int \left(\frac{\partial}{\partial \theta} \log f_{\theta} \right)^2 f_{\theta} d\mu = 0. \quad (2)$$

En el análisis de estas condiciones vamos a definir:

$$S(x, \theta) = \frac{\partial}{\partial \theta} \log f_{\theta}(x)$$

$$S'(x, \theta) = \frac{\partial}{\partial \theta} S(x, \theta) = \frac{\partial^2}{\partial \theta^2} \log f_{\theta}(x)$$

$$H(\theta, \theta') = \int (\log f_{\theta'}) f_{\theta} d\mu$$

$$A(\theta, \theta') = \int S(\cdot, \theta') f_{\theta} d\mu$$

$$B^2(\theta, \theta') = \int (S(\cdot, \theta'))^2 f_{\theta} d\mu$$

$$D(\theta, \theta') = \int S'(\cdot, \theta') f_{\theta} d\mu,$$

donde $\theta \in \Theta_0$ y $(\theta, \theta') \in \Theta_0 \times \Theta_0$.

Si f_{θ} es regular respecto a las primeras derivadas, se puede probar que $H(\theta, \theta')$ tiene derivadas parciales con respecto a θ y θ' en $\Theta_0 \times \Theta_0$. Por otro lado $H(\theta, \theta')$ como función de θ' , para θ fijo, tiene un máximo en $\theta' = \theta$.

La función $H(\theta, \theta')$ es de importancia básica en relación a la teoría de información y entropía en mecánica estadística.

Consideremos la función $H(\theta, \theta')$ y supongamos que $x \in \Omega = \mathbb{R}$. Se divide \mathbb{R} en intervalos I_{α} ($\alpha \in \mathbb{Z}$) disjuntos con $I_{\alpha} =]x_{\alpha}, x_{\alpha+1}]$, es decir:

$$P_{\theta}(I_{\alpha}) = F_{\theta}(x_{\alpha+1}) - F_{\theta}(x_{\alpha}).$$

Sea $\Delta = \max_{\alpha} |I_{\alpha}|$, se define:

$$H_{\Delta}(\theta, \theta') = \sum_{\alpha=-\infty}^{+\infty} (\log P_{\theta'}(I_{\alpha})) P_{\theta}(I_{\alpha}) = \log \left[\prod_{\alpha=-\infty}^{+\infty} (P_{\theta'}(I_{\alpha}))^{P_{\theta}(I_{\alpha})} \right]. \quad (3)$$

Dado que $[f_{\theta} > 0]$ es independiente de θ , entonces no existe I_{α} tal que:

$$P_{\theta'}(I_{\alpha}) = 0 \quad \text{y} \quad P_{\theta}(I_{\alpha}) > 0,$$

es decir se tiene $P_{\theta'} \ll P_{\theta}$, $\forall \theta \in \Theta$, $\forall \theta' \in \Theta$.

Usando el hecho que si p_1, \dots, p_r y q_1, \dots, q_r son dos conjuntos de números positivos:

$$p_1^{q_1} \cdots p_r^{q_r} \leq (p_1 + \cdots + p_r)^{q_1 + \cdots + q_r},$$

tenemos que el término entre paréntesis cuadrados de (3) no puede aumentar por la división de intervalos $\{I_{\alpha}\}$, entonces si existe una división $\{I_{\alpha}\}$ tal que la cantidad negativa $H_{\Delta}(\theta, \theta')$ exista y sea finita, $H_{\Delta}(\theta, \theta')$ no disminuye, cuando $\Delta \rightarrow 0$. Así, $\lim_{\Delta \rightarrow 0} H_{\Delta}(\theta, \theta') = H(\theta, \theta')$.

En el caso $\mathfrak{X} \in \Omega = \mathbb{R}^r$ se utilizan intervalos del tipo $I_{\alpha} = \prod_{i=1}^r [x_{i\alpha-1}, x_{i\alpha}]$ y H_{Δ} se define de manera similar.

Para la convergencia de las integrales (1) y (2) es suficiente determinar funciones P_{θ} -integrales que dominen a las funciones $\frac{\partial}{\partial \theta} \log f_{\theta}$, $\frac{\partial^2}{\partial \theta^2} \log f_{\theta}$ y $\left(\frac{\partial}{\partial \theta} \log f_{\theta}\right)^2$.

Las condiciones (1) y (2) las llamaremos condiciones de regularidad. Así diremos que f es regular con respecto a la primera derivada en Θ_0 si:

$$E_{\theta}(S(\cdot, \theta)) = \frac{d}{d\theta} \int f_{\theta} d\mu = 0$$

y regular con respecto a la segunda derivada en Θ_0 , si $B^2(\theta, \theta) < \infty$ y

$$E_{\theta}(S'(\cdot, \theta)) + E_{\theta}((S(\cdot, \theta))^2) = \frac{d^2}{d\theta^2} \int f_{\theta} d\mu = 0.$$

4.1.2 Caso $\theta \in \mathbb{R}^r$

Sea X una variable aleatoria con función de densidad $f(x, \theta)$, $\theta \in \mathbb{R}^r$ de componentes $(\theta_1, \dots, \theta_r) \in \Theta \subset \mathbb{R}^r$ funcionalmente independientes. Consideremos un intervalo abierto $\Theta_0 \subset \Theta$. Si suponemos

que la igualdad:

$$\int f_{\theta} d\mu = 1,$$

es dos veces diferenciable obtenemos que:

$$\begin{aligned} \frac{\partial}{\partial \theta_p} \int f_{\theta} d\mu &= \int \left(\frac{\partial}{\partial \theta_p} \log f_{\theta} \right) f_{\theta} d\mu = E_{\theta} \left(\frac{\partial}{\partial \theta_p} \log f_{\theta} \right) = 0, \\ \frac{\partial^2}{\partial \theta_p \partial \theta_q} \int f_{\theta} d\mu &= \int \left(\frac{\partial^2}{\partial \theta_p \partial \theta_q} \log f_{\theta} \right) f_{\theta} d\mu + \int \left(\frac{\partial}{\partial \theta_p} \log f_{\theta} \right) \left(\frac{\partial}{\partial \theta_q} \log f_{\theta} \right) f_{\theta} d\mu = 0, \end{aligned}$$

$$\text{o sea } E_{\theta} \left(\frac{\partial^2}{\partial \theta_p \partial \theta_q} \log f_{\theta} \right) = -\text{cov}_{\theta} \left(\frac{\partial}{\partial \theta_p} \log f_{\theta}, \frac{\partial}{\partial \theta_q} \log f_{\theta} \right).$$

Diremos que $f(x, \theta)$ es regular con respecto a la primera derivada parcial en θ_0 si:

$$E_{\theta} \left(\frac{\partial}{\partial \theta_p} \log f_{\theta} \right) = \frac{\partial}{\partial \theta_p} \int f_{\theta} d\mu = 0, \quad p = 1, \dots, r$$

y regular respecto a la segunda derivada parcial en Θ_0 , si la matriz

$$I_{\theta} = \left\| \text{cov}_{\theta} \left(\frac{\partial}{\partial \theta_p} \log f_{\theta}, \frac{\partial}{\partial \theta_q} \log f_{\theta} \right) \right\|$$

es finita y si:

$$E_{\theta} \left(\frac{\partial}{\partial \theta_p} \log f_{\theta} \cdot \frac{\partial}{\partial \theta_q} \log f_{\theta} \right) = -E_{\theta} \left(\frac{\partial^2}{\partial \theta_p \partial \theta_q} \log f_{\theta} \right), \quad p, q = 1, \dots, r.$$

Como en el caso $\theta \in \mathbb{R}$, definiremos:

$$S_p(x, \theta) = \frac{\partial}{\partial \theta_p} \log f_{\theta}(x)$$

$$S_{pq}(x, \theta) = \frac{\partial^2}{\partial \theta_p \partial \theta_q} \log f_{\theta}(x)$$

$$H(\theta, \theta') = \int (\log f_{\theta'}) f_{\theta} d\mu = E_{\theta}(\log f_{\theta'})$$

$$A_p(\theta, \theta') = \int S_p(\cdot, \theta') f_{\theta} d\mu = E_{\theta}(S_p(\cdot, \theta'))$$

$$B_{pq}(\theta, \theta') = \int S_p(\cdot, \theta') S_q(\cdot, \theta') f_{\theta} d\mu$$

$$D_{pq}(\theta, \theta') = \int S_{pq}(\cdot, \theta') f_{\theta} d\mu,$$

donde $p, q = 1, \dots, r$, $\theta \in \Theta_0$ y $(\theta, \theta') \in \Theta_0 \times \Theta_0$.

Dado que las componentes de $\theta = (\theta_1, \dots, \theta_r)$ son funcionalmente independientes, las componentes

de la variable aleatoria $(S_1(x, \theta), \dots, S_r(x, \theta))$ son linealmente independientes y la matriz $(B_{pq}(\theta, \theta'))$ es definida positiva para $(\theta, \theta') \in \Theta_0 \times \Theta_0$.

Un razonamiento análogo al caso $\theta \in \mathbb{R}$ se efectúa para $H(\theta, \theta')$.

4.2 Estimación puntual unidimensional

Sea $X = (X_1, \dots, X_n)$ una variable aleatoria de función de densidad $f(x_1, \dots, x_n, \theta)$, con $\theta \in \Theta \subset \mathbb{R}$ que satisface las condiciones de regularidad respecto a la primera derivada en $\Theta_0 \subset \Theta$.

Si $T(x_1, \dots, x_n)$ es un estimador insesgado de θ , entonces:

$$\text{var}_\theta(T) \geq \frac{1}{E_\theta(S_n^2)} = \frac{1}{B^2(\theta, \theta)},$$

con $S_n(x_1, \dots, x_n, \theta) = \frac{\partial}{\partial \theta} \log f_\theta(x_1, \dots, x_n)$.

La igualdad se obtiene si $K(T(x_1, \dots, x_n) - \theta) \equiv \frac{\partial}{\partial \theta} \log f_\theta(x_1, \dots, x_n)$, salvo en un conjunto de medida cero (en general K depende de θ).

4.2.1 Caso de una muestra aleatoria

Tomando en cuenta lo anterior, si la función de verosimilitud $L(x_1, \dots, x_n, \theta)$ es derivable con respecto a θ , para una muestra de tamaño n , se tiene:

$$S_n(x_1, \dots, x_n, \theta) = \frac{\partial}{\partial \theta} \log L(x_1, \dots, x_n, \theta) = \sum_{i=1}^n \frac{\partial}{\partial \theta} \log f_\theta(x_i) = \sum_{i=1}^n S(x_i, \theta),$$

por lo que $E_\theta(S_n^2) = nE_\theta(S^2) = nB^2(\theta, \theta)$.

Si T es un estimador de θ insesgado (i.e. $T(x_1, \dots, x_n)$ está en función de la muestra):

$$\text{var}_\theta(T) \geq \frac{1}{nB^2(\theta, \theta)}.$$

La igualdad se establece si y sólo si $K(T - \theta) \equiv \sum_{i=1}^n S(X_i, \theta) = \sum_{i=1}^n \frac{\partial}{\partial \theta} \log f_\theta(X_i)$.

Así hemos establecido el siguiente teorema.

Teorema 4.2.1 Si (X_1, \dots, X_n) es una muestra de tamaño n de función de densidad $f(x, \theta)$, la cual es regular con respecto a la primera derivada sobre $\Theta_0 \subset \Theta$ (intervalo abierto) y si T es un estimador insesgado de θ_0 , entonces $\text{var}_\theta(T) \geq \frac{1}{nB^2(\theta, \theta)}$.

La igualdad se da si $K(T - \Theta_0) \equiv \sum_{i=1}^n S(X_i, \theta_0)$ i.e. $\text{var}_{\theta_0}(T) = \frac{1}{nB^2(\theta_0, \theta_0)}$.

Fisher denomina al término $nB^2(\theta_0, \theta_0)$ la cantidad de información contenida en la muestra concerniente a θ_0 ; $B^2(\theta_0, \theta_0)$ es la cantidad de información sobre θ_0 , por observación. Desde este punto de vista la eficiencia de un estimador insesgado de θ_0 , es el porcentaje de información contenida en T relativa a un estimador eficiente de θ_0 .

Ejemplo Si (X_1, \dots, X_n) es una muestra de una distribución de Poisson de parámetro λ , la estadística \bar{X} es un estimador sin sesgo y eficiente de λ .

4.2.2 Valor mínimo de la varianza de un estimador sesgado

Una generalización del teorema anterior es posible, considerando el sesgo de T para θ como sigue.

Si $E_\theta(T) = \theta + b_n(\theta)$, donde $b_n(\theta)$ es el sesgo de T y si $b_n(\theta)$ es derivable, entonces:

$$\text{var}_\theta(T) \geq \frac{(1 + b'_n(\theta))^2}{nB^2(\theta, \theta)}.$$

La igualdad se establece $\iff K(T - \theta - b_n(\theta)) \equiv \sum_{i=1}^n S(X_i, \theta)$ salvo en un conjunto de medida cero.

Sea T una estadística suficiente y U un estimador sin sesgo de θ . Sea $E_\theta^T(U) = g \circ T$, entonces $g(T)$ es un estimador sin sesgo de θ y $\text{var}_\theta(E_\theta^T(U)) \leq \text{var}_\theta(U)$.

4.2.3 Estimación puntual para grandes muestras

Un estimador eficiente de un parámetro existe para muestras de tamaño n , sólo en ciertos casos especiales. En tales casos $S_n(x_1, \dots, x_n, \theta)$ toma la forma $K(\hat{\theta}_n - \theta) \equiv \sum_{i=1}^n S(X_i, \theta)$ y el estimador eficiente $\hat{\theta}_n(X_1, \dots, X_n)$ es por tanto dado por la solución de la ecuación $S_n(x_1, \dots, x, \theta) = 0$, para θ .

Si $S_n(x_1, \dots, x_n, \theta)$ no tiene esta forma especial, sin embargo de la ecuación se obtiene una solución para θ , que denotaremos sin ambigüedad $\hat{\theta}_n$, nos preguntamos por las propiedades del estimador de θ_0 . La respuesta a esta pregunta es que bajo ciertas condiciones que veremos, $\hat{\theta}_n$ es un estimador eficiente para θ_0 , en un sentido asintótico para grandes muestras.

Distribución asintótica de $S_n(x_1, \dots, x_n, \theta)$

Teorema 4.2.2 Sea (X_1, \dots, X_n) una muestra con función de densidad $f(x, \theta)$, donde $f(x, \theta)$ es regular respecto a la primera derivada de θ en Θ_0 , entonces si $B^2(\theta, \theta)$ existe y es finita, $S_n(X_1, \dots, X_n, \theta_0)$ es asintóticamente distribuido como $N(0, nB^2(\theta_0, \theta_0))$.

Demostración Sea $S_n(X_1, \dots, X_n, \theta_0)$ la variable aleatoria definida por $S_n \circ (X_1, \dots, X_n)$, es claro que $E_{\theta_0}(S_n) = 0$ y además $\text{var}_{\theta_0}(S_n) = nB^2(\theta_0, \theta_0)$, entonces por el teorema central del límite, cuando $n \rightarrow \infty$:

$$\frac{S_n}{\sqrt{nB^2(\theta_0, \theta_0)}} \rightarrow N(0, 1).$$

Convergencia de los estimadores de máxima verosimilitud

Hemos visto que $\frac{1}{n}S_n \xrightarrow{\text{Pr}} 0$, lo que sugiere que la ecuación:

$$S_n(x_1, \dots, x_n, \theta) = 0,$$

aporta una sucesión de soluciones $\{\hat{\theta}_n : n \geq 1\}$, la cual converge a θ con probabilidad 1. Esta aseveración se formaliza en el teorema siguiente.

Teorema 4.2.3 Sea (X_1, \dots, X_n) una muestra con función de densidad $f(x, \theta)$, donde $f(x, \theta)$ es regular respecto a la segunda derivada en Θ_0 . Así $S'(x, \theta_0)$ es continua en todo punto $\theta \in \Theta_0$, para todo $x \in \Omega$, salvo en un conjunto de medida cero, entonces existe una sucesión de soluciones de la ecuación $S_n(x_1, \dots, x_n, \theta) = 0$, la cual converge seguramente a θ_0 . En particular si se tiene una única solución $\hat{\theta}(x_1, \dots, x_n, \theta)$, para todo $n \geq n_0$, entonces la sucesión $\{\hat{\theta}(X_1, \dots, X_n) : n \geq n_0\}$ converge casi seguro a θ_0 .

Demostración Dado que $S(x, \theta)$ es continua en Θ_0 , entonces $A(\theta_0, \theta)$ es continua y estrictamente decreciente en θ para algún intervalo $\Theta'_0 \subset \Theta_0$ que contenga a θ_0 . Sin pérdida de generalidad se puede escoger $\Theta'_0 =]\theta_0 - \delta, \theta_0 + \delta[$, con $\delta > 0$ tal que $\Theta'_0 \subset \Theta_0$ y como $A(\theta_0, \theta_0) \downarrow$ sobre Θ'_0 y $A(\theta_0, \theta_0) = 0$, entonces:

$$A(\theta_0, \theta_0 - \delta) > 0, \quad A(\theta_0, \theta_0 + \delta) < 0.$$

Por otro lado $\frac{1}{n}S_n(X_1, \dots, X_n, \theta)$ es la media de una muestra proveniente de una población de esperanza $A(\theta_0, \theta_0)$, si θ_0 es el verdadero valor de θ y por la ley fuerte de grandes números:

$$\frac{1}{n}S_n(X_1, \dots, X_n, \theta) \xrightarrow{\text{c.s.}} A(\theta_0, \theta),$$

es decir $\forall \epsilon > 0, \forall \eta > 0$, existen $n_{\epsilon, \eta} \in \mathbb{N}$ tal que si $n \geq n_{\epsilon, \eta}$:

$$P_{\theta_0} \left(\sup_{k \geq n} \left| \frac{1}{k} S_k(X_1, \dots, X_k, \theta) - A(\theta_0, \theta) \right| < \eta \right) > 1 - \epsilon. \quad (3)$$

Sea $E_n = \{(x_1, \dots, x_n) \in \Omega^n : \left| \frac{1}{n} S_n(x_1, \dots, x_n, \theta) - A(\theta_0, \theta) \right| < \eta\}$, $\forall n \geq n_{\epsilon, \eta}$.

Observemos que en realidad tenemos que $\frac{1}{n}S_n(X_1, \dots, X_n, \theta) \xrightarrow{\text{c.s.}} A(\theta_0, \theta)$ y no como se escribió en (3), pero como $X: \Omega' \rightarrow \Omega = \mathbb{R}^m$ escribimos E_n en vez de:

$$\omega' \mapsto X(\omega') = \mathbf{x}$$

$$E'_n = \{(\omega'_1, \dots, \omega'_n) \in \Omega'^n : \left| \frac{1}{n} S_n(X_1, \dots, X_n, \theta)(\omega'_1, \dots, \omega'_n) - A(\theta_0, \theta) \right| < \eta\},$$

pues $P(X \in G) = P_X(G)$. Así, $\forall n \geq n_{\epsilon, \eta}$ sobre E_n se tiene:

$$A(\theta_0, \theta) - \eta < \frac{1}{n} S_n(x_1, \dots, x_n, \theta) < A(\theta_0, \theta) + \eta$$

Consideremos $\eta < \inf\{A(\theta_0, \theta_0 - \delta), -A(\theta_0, \theta_0 + \delta)\}$:

a) Si $\eta < A(\theta_0, \theta_0 - \delta) \implies \frac{1}{n} S_n(x_1, \dots, x_n, \theta_0 - \delta) > 0$, $\forall n \geq n_{\epsilon, \eta(\delta)}$, $\forall (x_1, \dots, x_n) \in B_n$,

b) Si $\eta < -A(\theta_0, \theta_0 + \delta) \implies \frac{1}{n} S_n(x_1, \dots, x_n, \theta_0 + \delta) < 0$, $\forall n \geq n'_{\epsilon, \eta(\delta)}$, $\forall (x_1, \dots, x_n) \in B'_n$,

entonces el conjunto B_n (resp. B'_n) donde se cumple a) (resp. b)), contiene a E_n , $\forall n \geq n_{\epsilon, \eta(\delta)}$ (resp. $n'_{\epsilon, \eta(\delta)}$).

Sea $n_{\epsilon\delta} = \sup\{n_{\epsilon, \eta(\delta)}, n'_{\epsilon, \eta(\delta)}\}$, entonces $E_n \subset B_n \cap B'_n$, $\forall n \geq n_{\epsilon\delta}$ y además:

$$P_{\theta_0} \left(\frac{1}{n} S_n(x_1, \dots, x_n, \theta_0 - \delta) > 0, \frac{1}{n} S_n(x_1, \dots, x_n, \theta_0 + \delta) < 0 : \forall n \geq n_{\epsilon\delta} \right) > 1 - \epsilon.$$

Como $\frac{1}{n}S_n(x_1, \dots, x_n, \theta)$ es continua en θ , $\forall (x_1, \dots, x_n)$ salvo en un conjunto de medida cero, se tiene que existe una solución $\hat{\theta}_n$ de la ecuación:

$$\frac{1}{n} S_n(x_1, \dots, x_n, \theta) = 0,$$

con $\hat{\theta}_n \in (\theta_0 \pm \delta)$, $\forall (x_1, \dots, x_n)$, salvo un conjunto de medida cero.

Sea $C_n = \{(x_1, \dots, x_n, \theta) \in \Omega^n : \frac{1}{n}S_n(x_1, \dots, x_n, \theta) = 0, \theta \in (\theta_0 \pm \delta)\}, \forall n \geq n_{\epsilon, \delta}$ y sea $(x_1^*, \dots, x_n^*) \in E_n$, entonces $\frac{1}{n}S_n(x_1^*, \dots, x_n^*, \theta) \leq 0$, si $\theta = \theta_0 \pm \delta, \forall n \geq n_{\epsilon, \delta}$ y como $\frac{1}{n}S_n(x_1, \dots, x_n, \theta)$ es continua en $\theta \in (\theta_0 \pm \delta)$, salvo un conjunto de medida cero, tenemos que:

$$\frac{1}{n}S_n(x_1^*, \dots, x_n^*, \hat{\theta}_n^*) = 0,$$

para algún $\hat{\theta}_n^* \in (\theta_0 \pm \delta)$, salvo un conjunto de medida cero, $\forall n \geq n_{\epsilon, \delta}$. Por la definición de C_n se tiene que:

$$(x_1^*, \dots, x_n^*) \in C_n, \quad \forall n \geq n_{\epsilon, \delta},$$

es decir $E_n \subset C_n$ y $P_{\theta_0}(\frac{1}{n}S_n(x_1, \dots, x_n, \theta) = 0, \text{ para algún } \theta \in]\theta_0 \pm \delta[, \forall n \geq n_{\epsilon, \delta}) > 1 - \epsilon$ o equivalentemente $\forall \epsilon > 0, \forall \delta > 0$, existe $n_{\epsilon, \delta}$ tal que si $n \geq n_{\epsilon, \delta}$:

$$P_{\theta_0}(|\hat{\theta}_n - \theta_0| < \delta, \forall n \geq n_{\epsilon, \delta}) > 1 - \epsilon.$$

En particular, si se tiene una única solución para $n \geq n_0$, se tiene que:

$$\hat{\theta}_n \xrightarrow{\text{c.s.}} \theta_0.$$

Distribución asintótica de los estimadores de máxima verosimilitud

Teorema 4.2.4 Sea (X_1, \dots, X_n) una muestra con función de densidad $f(x, \theta_0)$, donde $f(x, \theta)$ es regular con respecto a las dos primeras derivadas en Θ_0 .

Si el estimador de máxima verosimilitud $\hat{\theta}(x_1, \dots, x_n)$ de θ_0 es el único para $n \geq n_0$ y medible con respecto a μ^n , entonces tenemos que $\hat{\theta}(X_1, \dots, X_n)$ es asintóticamente normal de distribución $N(\theta_0, \frac{1}{nB^2(\theta_0, \theta_0)})$, cuando $n \rightarrow \infty$.

Demostración Dado que $\hat{\theta}$ es único a partir de un n_0 , por el Teorema 4.2.3 se tiene que $\hat{\theta}_n \xrightarrow{\text{c.s.}} \theta_0$.

Como $\hat{\theta}$ es una variable aleatoria, entonces $\forall \epsilon > 0, \forall \delta > 0$, existe $n(\epsilon, \delta, n_0)$ tal que:

$$P_{\theta_0}(E_n, \forall n \geq n(\epsilon, \delta, n_0)) > 1 - \epsilon,$$

donde $E_n = \{(x_1, \dots, x_n) \in \Omega^n : |\hat{\theta}(x_1, \dots, x_n) - \theta_0| < \delta\}$.

Por otro lado, sabemos que $S(x, \theta)$ es derivable i.e. $S_n(x_1, \dots, x_n, \theta)$ es derivable. Así tenemos que:

$$\frac{1}{\sqrt{n}}S_n(x_1, \dots, x_n, \theta) = \frac{1}{\sqrt{n}}S_n(x_1, \dots, x_n, \hat{\theta}) + \frac{1}{n}S'_n(x_1, \dots, x_n, \theta^*)(\theta_0 - \hat{\theta}), \quad (4)$$

donde $\theta^*(X_1, \dots, X_n)$ es una variable aleatoria tal que $|\theta_0 - \theta^*| \leq |\theta_0 - \hat{\theta}|$.

La ecuación (4) es válida en el conjunto E_n , por lo que:

$$P_{\theta_0} \left(\frac{1}{\sqrt{n}}S_n(X_1, \dots, X_n, \theta_0) = \frac{1}{n}S'_n(X_1, \dots, X_n, \theta^*)[\sqrt{n}(\theta_0 - \hat{\theta})], \forall n \geq n(\epsilon, \delta, n_0) \right) > 1 - \epsilon.$$

Sabemos que $\frac{1}{n}S'_n(X_1, \dots, X_n, \theta_0) \xrightarrow{\text{Pr}} -B^2(\theta_0, \theta_0)$ y como $\hat{\theta} \xrightarrow{\text{Pr}} \theta_0$, se tiene que $\theta^* \xrightarrow{\text{Pr}} \theta_0$, por lo que $\frac{1}{n}S'_n(X_1, \dots, X_n, \theta^*) \xrightarrow{\text{Pr}} -B^2(\theta_0, \theta_0)$.

Así las sucesiones:

$$\begin{aligned} & \frac{1}{\sqrt{n}}S_n(X_1, \dots, X_n, \theta_0), \\ & \frac{1}{n}S'_n(X_1, \dots, X_n, \theta^*)[\sqrt{n}(\theta_0 - \hat{\theta})], \\ & \frac{1}{n}S'_n(X_1, \dots, X_n, \theta_0)[\sqrt{n}(\theta_0 - \hat{\theta})], \\ & \sqrt{n}(\hat{\theta} - \theta_0)B^2(\theta_0, \theta_0), \end{aligned}$$

convergen en la ley si una de ellas converge. Pero por el teorema 4.2.2 se tiene:

$$\sqrt{n}(\hat{\theta} - \theta_0)B^2(\theta_0, \theta_0) \xrightarrow{\text{ley}} N(0, nB^2(\theta_0, \theta_0)),$$

o sea que $\hat{\theta} \sim N\left(\theta_0, \frac{1}{nB^2(\theta_0, \theta_0)}\right)$ asintóticamente, cuando $n \rightarrow \infty$.

Definición 4.2.1 Sea T un estimador de θ_0 , si $T \xrightarrow{\text{Pr}} \theta_0$, cuando $n \rightarrow \infty$, se dice que T es un estimador consistente de θ_0 .

Teorema 4.2.5 Sea (X_1, \dots, X_n) es una muestra con función de densidad $f(x, \theta)$, donde $f(x, \theta)$ es regular con respecto a la primera derivada en Θ_0 y $B^2(\theta, \theta)$ es finita para todo $\theta \in \Theta_0$. Si T es un estimador consistente de θ_0 y si la derivada del sesgo de T tiende a cero, cuando $n \rightarrow \infty$, entonces:

$$\liminf_n \text{var}_{\theta_0}(\sqrt{n}T) \geq \frac{1}{B^2(\theta_0, \theta_0)}.$$

Demostración Sea T estimador de θ_0 tal que $E_{\theta_0}(T) = \theta_0 + b_n(\theta_0)$, entonces:

$$\text{var}_{\theta_0}(\sqrt{n}T) \geq \frac{(1 + b'_n(\theta_0))^2}{B^2(\theta_0, \theta_0)}$$

y tomando el liminf a ambos lados se obtiene el resultado.

Corolario 4.2.1 *Bajo las hipótesis del teorema 4.2.4, el estimador de máxima verosimilitud $\hat{\theta}$ para θ_0 es consistente y asintóticamente eficiente.*

Definición 4.2.2 *Se llama eficiencia asintótica de un estimador T de θ_0 al cociente:*

$$\alpha = \frac{B^{*2}(\theta_0, \theta_0)}{B^2(\theta_0, \theta_0)} \leq 1,$$

donde $\frac{1}{B^{*2}(\theta_0, \theta_0)}$ es la varianza de la distribución límite de $\sqrt{n}(T - \theta_0)$.

Si $\alpha = 1$ se dice que T es asintóticamente eficiente.

Funciones del estimador de máxima verosimilitud de varianza asintótica $\frac{1}{n}$

En el teorema 4.2.4 hemos visto que bajo ciertas condiciones de regularidad, el estimador $\hat{\theta}$ de máxima verosimilitud tiene por varianza asintótica $\frac{1}{nB^2(\theta_0, \theta_0)}$, donde θ_0 es el verdadero valor de θ . Un problema de interés teórico se presenta cuando se tiene un nuevo parámetro ξ función de θ , tal que la varianza asintótica de $\xi(\hat{\theta})$ sea $\frac{1}{n}$.

Sea $\xi(\theta)$ una función de θ , teniendo una derivada $\xi'(\theta)$ y una inversa $\theta(\xi)$ en algún intervalo Θ_0 conteniendo a θ , tal que $\theta(\xi)$ tiene una derivada $\theta'(\xi)$ con respecto a ξ .

Sea ξ_0 el verdadero valor de ξ , es decir $\theta(\xi_0) = \theta_0$. Así:

$$S(x, \theta(\xi)) = \frac{d}{d\xi} \log f(x, \theta(\xi)) \xi'(\theta),$$

$$B^2(\theta(\xi), \theta(\xi)) = (\xi'(\theta))^2 \int \left[\frac{\partial}{\partial \xi} \log f(\cdot, \theta(\xi)) \right]^2 f(\cdot, \theta(\xi)) d\mu.$$

El estimador $\hat{\xi}$ para ξ_0 tiene una varianza asintótica $\frac{1}{n}$ sii la integral de la ecuación anterior es igual a 1, es decir que:

$$\xi(\theta) = \int_{\theta_0}^{\theta} B(\theta, \theta) d\theta + \xi_0,$$

con $B(\theta, \theta) = +\sqrt{B^2(\theta, \theta)}$.

Teorema 4.2.6 *Sea (X_1, \dots, X_n) una muestra con función de densidad $f(x, \theta_0)$, donde $f(x, \theta)$ es regular con respecto a su segunda derivada. Si $\hat{\theta}$ es el estimador de máxima verosimilitud para*

θ_0 , entonces $\xi(\hat{\theta})$, donde $\xi(\theta) = \int_{\theta_0}^{\theta} B(\theta, \theta) d\theta + \xi_0$ tiene una derivada $\xi'(\theta)$ y una inversa en Θ_0 es distribuido asintóticamente como $N(\xi_0, \frac{1}{n})$.

Ejercicio Sea (X_1, \dots, X_n) una muestra de una distribución de Poisson $P(\mu_0)$. Determine la estadística función del estimador de máxima verosimilitud de varianza asintótica $\frac{1}{n}$.

Estimación por intervalo

Definición 4.2.3 Sea (X_1, \dots, X_n) una variable aleatoria con función de densidad $f_n(\cdot, \theta)$ y sean $T_1(x_1, \dots, x_n), T_2(x_1, \dots, x_n)$ dos funciones medibles reales tales que se tiene $T_1 < T_2$ c.p.d. Si T_1 y T_2 pueden ser determinadas para que dado γ :

$$P_{\theta}(T_1 < \theta < T_2) = \gamma,$$

entonces el intervalo aleatorio $]T_1, T_2[$ es llamado intervalo de confianza para θ de coeficiente de seguridad γ .

Los procedimientos para determinar los intervalos de confianza son variados. Aquí veremos una forma de construir intervalos de confianza, con una muestra de función de densidad $f(x, \theta)$.

Teorema 4.2.7 Sea (X_1, \dots, X_n) una muestra con función de densidad $f(x, \theta_0)$. Supongamos que la función $g(x_1, \dots, x_n, \theta)$ satisface:

- a) g está definida en todo punto θ de un intervalo $]\theta_1, \theta_2[$ conteniendo a θ_0 , en todo punto de Ω^n , excepto en un conjunto de medida cero.
- b) g es monótona y continua en θ .
- c) la función de repartición de $g(x_1, \dots, x_n, \theta)$ no depende de θ .

Sea $]g_1, g_2[$ un intervalo para el cual $P_{\theta}(g_1 < g < g_2) = \gamma$, entonces si θ_0 es el verdadero valor de θ , las soluciones T_1, T_2 de la ecuación $g(x_1, \dots, x_n, \theta) = g_1, g_2$ existen y $]T_1, T_2[$ es un intervalo de confianza para θ_0 .

Demostración Dado que la función de repartición de $g(x_1, \dots, x_n, \theta)$ no depende de θ , para $\gamma, 0 < \gamma < 1$ y para $\theta = \theta_0$, se pueden escoger g_1 y g_2 tales que:

$$P_{\theta}(g_1 < g(x_1, \dots, x_n, \theta_0) < g_2) = \gamma.$$

Sea $\omega = (x_1, \dots, x_n) \in \Omega^n$, la ecuación $g_1 = g(\omega, \theta)$ admite una solución θ_1 , pues $g(\omega, \theta)$ es continua en θ ; la solución es $\theta_1(\omega)$. Igualmente la ecuación $g_2 = g(\omega, \theta)$ admite una solución θ_2 y se denota $\theta_2(\omega)$.

Así tenemos que:

$$g(\omega, \theta_1(\omega)) < g(\omega, \theta) < g(\omega, \theta_2(\omega)), \quad \forall \omega \in E_n,$$

donde $E_n = \{\omega \in \Omega^n : g_1 < g(\omega, \theta) < g_2\}$. Como $g(\omega, \theta)$ es monótona en θ , la desigualdad anterior es equivalente a:

$$\theta_1(\omega) < \theta_0 < \theta_2(\omega), \quad \forall \omega \in E_n, \text{ por lo que } P_{\theta_0}(\theta_1 < \theta_0 < \theta_2) = \gamma.$$

Nos podemos preguntar si es posible obtener una función $g(x_1, \dots, x_n, \theta)$ cuya función de repartición sea independiente de θ , asumiendo que $F(x, \theta)$ es continua en todo x . Una función tal se puede encontrar. En efecto la función $F_n(x_1, \dots, x_n, \theta) = \prod_{i=1}^n F(x_i, \theta)$ nos sirve si $F(x, \theta)$ es continua y monótona en θ , para todos los elementos de Ω , salvo en un conjunto de medida cero, pues las funciones $F(X_i, \theta)$, $i = 1, \dots, n$, consideradas como variables aleatorias son independientes y cada una sigue una ley uniforme $U(0, 1)$. Así:

$$-\log F(X_i, \theta) \sim \Gamma(1, 1), \quad -\sum_{i=1}^n \log F(X_i, \theta) \sim \Gamma(n, 1)$$

y además:

$$P_{\theta}(-\log b_2 < -\sum_{i=1}^n \log F(X_i, \theta) < -\log b_1) = \frac{1}{\Gamma(n)} \int_{\log b_1}^{\log b_2} y^{n-1} e^{-y} dy,$$

no depende de θ . Si b_1 y b_2 se escogen de modo que la integral sea igual a γ :

$$P_{\theta}(b_1 < \prod_{i=1}^n F(X_i, \theta) < b_2) = \gamma$$

y como $F(x, \theta)$ es continua y monótona en θ , $\forall x \in \Omega$ salvo en un conjunto de medida cero, también es válido para $\prod_{i=1}^n F(x_i, \theta)$ i.e.

$$P_{\theta_0}(\theta_1 < \theta_0 < \theta_2) = \gamma$$

nos determina un intervalo de confianza para θ , si θ_0 es el verdadero valor de θ . Observemos que si eliminamos la hipótesis de monotonía de g en θ , no obtenemos un intervalo, sino un conjunto (aleatorio) E dependiente de (X_1, \dots, X_n) y tal que $P_{\theta_0}(\theta_0 \in E) = \gamma$.

4.2.4 Intervalos de confianza para muestras de distribución discreta

En el caso en que $f(x, \theta)$ es una función de densidad de una variable aleatoria discreta, el resultado anterior no se aplica, pues no podemos determinar intervalos de confianza, con un coeficiente de seguridad exactamente igual a γ . Sin embargo bajo ciertas condiciones podemos tener un intervalo con probabilidad mayor o igual a γ .

Teorema 4.2.8 Sea (X_1, \dots, X_n) una muestra con función de densidad $f(x, \theta_0)$, donde X es una variable aleatoria discreta y $\theta \in]\theta_1, \theta_2[$. Sea T un estimador de θ_0 definido en todo punto masa de Ω^n y contenido en (θ_1, θ_2) y sea $G(t, \theta)$ su función de repartición la cual es continua y decreciente en θ en todo punto masa de T , tal que:

$$\lim_{\theta \rightarrow \theta_1} G(t, \theta) = 1, \quad \lim_{\theta \rightarrow \theta_2} G(t, \theta) = 0, \quad \forall t \in]\theta_1, \theta_2[.$$

Sean t_2 y t_1 los valores de t para los cuales $G(t, \theta) = \gamma_1$ y $G^*(t, \theta) = 1 - G(t, \theta) = \gamma_1^*$ respectivamente, con γ_1, γ_1^* positivos tales que $0 < \gamma = 1 - \gamma_1 - \gamma_1^* < 1$, entonces el intervalo $[T_1, T_2]$ es un intervalo de confianza para θ_0 , con coeficientes de seguridad mayor o igual a γ .

Demostración Sea $\theta_1 = \sup\{t/G(t, \theta_0) \leq \gamma_1\}$ y $\theta_2 = \inf\{t/G^*(t, \theta_0) \leq \gamma_1^*\}$.

Sea $V = \{\omega/\theta_1 < T(\omega) < \theta_2\}$, entonces:

$$P_{\theta_0}(\theta_1 < T < \theta_2) \geq 1 - \gamma_1^* - \gamma_1 = \gamma,$$

$\forall \omega \in V, \theta_1 < T(\omega) < \theta_2 \implies G(t, \theta_0) > \gamma_1$ y $G^*(t, \theta_0) > \gamma_1^*$, con $t = T(\omega)$.

Sea T_2 la variable aleatoria solución de $G(t, \theta) = \gamma_1$ i.e. $T_2(\omega)$ es solución de $G(T_2(\omega), \theta) = \gamma_1$, es decir $G(t, T_2(\omega)) = \gamma_1$, con $t = T(\omega)$. Igualmente $G^*(t, T_1(\omega)) = \gamma_1^*$, con $t = T(\omega)$.

Así, $\forall \omega \in V, t = T(\omega), G(t, T_2(\omega)) \leq G(t, \theta_0), G^*(t, T_1(\omega)) \leq G^*(t, \theta_0)$, es decir:

$$T_1(\omega) \leq \theta_0 \leq T_2(\omega), \quad \forall \omega \in V,$$

pues G es decreciente en $\theta, \forall t \in (\theta_1, \theta_2)$ y en particular para $t = T(\omega)$, con $\omega \in V$, o sea:

$$P_{\theta_0}(T_1 \leq \theta_0 \leq T_2) \geq \gamma.$$

4.2.5 Estimación por intervalo para grandes muestras

Sea (X_1, \dots, X_n) una muestra con función de densidad $f(x, \theta)$ y sea la función de estimación de verosimilitud $h_n(x_1, \dots, x_n, \theta)$ definida por

$$h_n(x_1, \dots, x_n, \theta) = \frac{S_n(x_1, \dots, x_n, \theta)}{nB(\theta_0, \theta)}.$$

Sabemos que $\sqrt{n}h_n(X_1, \dots, X_n, \theta_0)$ converge en ley a $N(0, 1)$, si θ_0 es el verdadero valor de θ y bajo las hipótesis del teorema 4.2.4 tenemos que si $n \geq n_0$:

$$\sqrt{nh_n}(X_1, \dots, X_n, \theta_0) = h'_n(X_1, \dots, X_n, \theta^*)[\sqrt{n}(\theta_0 - \hat{\theta})],$$

$\forall (x_1, \dots, x_n) \in E_n$, $\hat{\theta}$ y θ^* son tales que $|\theta_0 - \theta^*| \leq |\theta_0 - \hat{\theta}|$. Las dos sucesiones de variables aleatorias $\sqrt{nh_n}(X_1, \dots, X_n, \theta_0)$, $h'_n(X_1, \dots, X_n, \theta^*)[\sqrt{n}(\theta_0 - \hat{\theta})]$, convergen en distribución a $N(0, 1)$. Además si $\frac{\partial}{\partial \theta} B^2(\theta_0, \theta)$ existe y es acotada, entonces $h'_n(X_1, \dots, X_n, \theta^*)$ y $h'_n(X_1, \dots, X_n, \hat{\theta})$ convergen en probabilidad, cuando $n \rightarrow \infty$, a $B(\theta_0, \theta_0)$. Así podemos escribir:

$$\lim_{n \rightarrow \infty} P_{\theta_0}(-\lambda_\gamma < h'_n(X_1, \dots, X_n, \hat{\theta})[\sqrt{n}(\theta_0 - \hat{\theta}_0)] < \lambda_\gamma) = \gamma,$$

donde λ_γ y γ son tales que $P(|N(0, 1)| < \lambda_\gamma) = \gamma$, es decir:

$$\lim_{n \rightarrow \infty} P_{\theta_0} \left(\hat{\theta} - \frac{\lambda_\gamma}{\sqrt{nh'_n}(X_1, \dots, X_n, \hat{\theta})} < \theta_0 < \hat{\theta} + \frac{\lambda_\gamma}{\sqrt{nh'_n}(X_1, \dots, X_n, \hat{\theta})} \right) = \gamma.$$

Así tenemos un intervalo de confianza $\hat{\theta} \pm \frac{\lambda_\gamma}{\sqrt{nh'_n}(X_1, \dots, X_n, \hat{\theta})}$, para θ_0 de coeficiente de seguridad asintótico γ .

4.2.6 Estimación por intervalos para funciones de estimación generales

Teorema 4.2.9 Sea (X_1, \dots, X_n) una muestra de función de densidad $f(x, \theta)$, la cual es regular respecto a las dos primeros θ derivadas y sea $g_n(X_1, \dots, X_n, \theta)$ una función medible tal que:

a) $\int (\sqrt{n}g_n(x_1, \dots, x_n, \theta))^j f(x_1, \dots, x_n, \theta) d\mu^n = 0, 1; j = 1, 2$, para $\theta \in \Theta_0$.

b) $\sqrt{n}g_n(X_1, \dots, X_n, \theta) \xrightarrow{\text{ley}} N(0, 1)$, si θ_0 es el verdadero valor de θ .

c) $g_n(\cdot, \theta)$ tiene una θ derivada $g'_n(\cdot, \theta)$ continua para $\theta \in \Theta_0$, para todo punto de Ω^n salvo un conjunto de medida cero.

d) Si θ_0 es el verdadero valor de θ , $g'_n(X, \theta) \xrightarrow{\text{Pr}} B^*(\theta_0, \theta)$ uniformemente con respecto a $\theta \in \Theta_0$, donde $B^*(\theta_0, \theta_0)$ es acotada en Θ_0 y $B^*(\theta_0, \theta_0) \neq 0$ y $X = (X_1, \dots, X_n)$.

e) Si $A_n^*(\theta, \theta') = \int g_n(\cdot, \theta')L(\cdot, \theta)d\mu^n$, $B_n^*(\theta, \theta') = \int g'_n(\cdot, \theta')L(\cdot, \theta)d\mu^n$, entonces:

$$\frac{\partial}{\partial \theta'} A_n^*(\theta, \theta') = B_n^*(\theta, \theta'), \quad \forall (\theta, \theta') \in \Theta_0 \times \Theta_0,$$

$$A_n^*(\theta, \theta) = 0, \quad \forall \theta \in \Theta_0,$$

$$\lim_{n \rightarrow \infty} B_n^*(\theta, \theta) = B^*(\theta, \theta), \quad \forall \theta \in \Theta_0.$$

Definición 4.2.4 Toda función $g_n(x_1, \dots, x_n, \theta)$ con las propiedades enunciadas anteriormente es llamada función de estimación regular para θ_0 .

Se verifica usando un argumento similar al teorema 4.2.3 que si $\forall n \geq n_0$, para algún n_0 , la ecuación $g_n(x_1, \dots, x_n, \theta) = 0$ tiene una raíz, entonces las ecuaciones:

$$g_n(x_1, \dots, x_n, \theta) = 0, \quad n \geq n_0,$$

proveen una sucesión de ceros $\tilde{\theta}(x_1, \dots, x_n)$, $n \geq n_0$, que converge en probabilidad a θ_0 .

Además las dos sucesiones de variables aleatorias:

$$\sqrt{n}g_n(x_1, \dots, x_n, \theta_0), \quad g'_n(x_1, \dots, x_n, \tilde{\theta}^*)[\sqrt{n}(\theta_0 - \tilde{\theta})], \quad n \geq n_0,$$

donde $\tilde{\theta}^*$ satisface $|\theta_0 - \tilde{\theta}^*| < |\theta_0 - \tilde{\theta}|$, convergen juntas en ley a una distribución $N(0, 1)$.

Pero $g'_n(x_1, \dots, x_n, \tilde{\theta}^*)$ y $g'_n(x_1, \dots, x_n, \tilde{\theta})$ convergen en probabilidad, cuando $n \rightarrow \infty$, a $B^*(\theta_0, \theta_0)$. Así tenemos que:

$$\lim_{n \rightarrow \infty} P_{\theta_0} \left(\tilde{\theta} - \frac{\lambda_\gamma}{\sqrt{n}g'_n(x_1, \dots, x_n, \tilde{\theta})} < \theta_0 < \tilde{\theta} + \frac{\lambda_\gamma}{\sqrt{n}g'_n(x_1, \dots, x_n, \tilde{\theta})} \right) = \gamma \quad (5)$$

i.e. $\tilde{\theta} \pm \frac{\lambda_\gamma}{\sqrt{n}g'_n(x_1, \dots, x_n, \tilde{\theta})}$ es un intervalo de confianza para θ_0 , de coeficiente de seguridad asintótico γ .

4.2.7 Intervalo de confianza asintótico más pequeño

Vamos a probar que el coeficiente del cuadrado de la longitud del intervalo de verosimilitud, con un intervalo similar obtenido a través de una función general $g_n(x_1, \dots, x_n, \theta)$ satisfaciendo ciertas

condiciones, converge en probabilidad a un número que no puede exceder a 1.

El cociente r_n del cuadrado de la longitud de los intervalos se escribe:

$$r_n = c_n^2 d_n^2,$$

$$\text{donde } c_n = \frac{g'_n(X_1, \dots, X_n, \theta_0)}{h'_n(X_1, \dots, X_n, \hat{\theta})} \text{ y } d_n = \frac{g'_n(X_1, \dots, X_n, \tilde{\theta})}{g'_n(X_1, \dots, X_n, \theta_0)}.$$

Sabemos que el numerador y el denominador de d_n^2 convergen a $B^{*2}(\theta_0, \theta_0)$ i.e. $d_n^2 \xrightarrow{\text{Pr}} 1$. Similarmente c_n^2 y r_n convergen en probabilidad a $\frac{B^{*2}(\theta_0, \theta_0)}{B^2(\theta_0, \theta_0)} \leq 1$.

En efecto, por la condición i) del teorema 4.2.9 tenemos que derivando para $j = 1$, con respecto a θ en $\theta = \theta_0$:

$$E_{\theta_0}(g'_n(X_1, \dots, X_n, \theta_0)) = E_{\theta_0}\left(\sqrt{n}g_n(X_1, \dots, X_n, \theta_0) \frac{1}{\sqrt{n}}S_n(X_1, \dots, X_n, \theta_0)\right).$$

Si elevamos al cuadrado y aplicamos la desigualdad de Schwarz tenemos:

$$B^{*2}(\theta_0, \theta_0) \leq E_{\theta_0}\left(\left[\frac{1}{\sqrt{n}}S_n(X_1, \dots, X_n)\right]^2\right) = B^2(\theta_0, \theta_0),$$

usando la condición i) del teorema 4.2.9 para $j = 2$ y tomando el límite obtenemos:

$$\frac{B^{*2}(\theta_0, \theta_0)}{B^2(\theta_0, \theta_0)} \leq 1.$$

Teorema 4.2.10 Sea (X_1, \dots, X_n) una muestra de función de densidad $f(x, \theta_0)$, donde $f(x, \theta)$ es regular con respecto a la segunda derivada para $\theta \in \Theta_0$, entonces si la función $g_n(x_1, \dots, x_n, \theta)$ es una función regular de estimación para θ_0 , un intervalo de confianza para θ_0 , cuando $n \rightarrow \infty$, está dado por (5). Además el intervalo de confianza para θ_0 , dado por la función de estimación de verosimilitud, es el más pequeño asintóticamente.

Es interesante notar la equivalencia existente entre el intervalo asintóticamente más pequeño y la eficiencia asintótica de un estimador.

4.3 Estimación puntual multidimensional

Se considera $\theta \in \Theta_r \subset \mathbb{R}^r$ y el verdadero valor $\theta_0 \in \Theta_{r0} \subset \Theta_r$. Así un elemento $\theta = (\theta_1, \dots, \theta_r)$ y un estimador T de θ_0 es $T = (T_1, \dots, T_r)$ una estadística vectorial.

Diremos que T es insesgado si T_i es insesgado para θ_i , $i = 1, \dots, r$ y T es inconsistente si sus componentes lo son.

Nuestro interés se centra en obtener una versión multidimensional de la eficiencia de un estimador.

Sea:

$$S_{pn}(x_1, \dots, x_n, \boldsymbol{\theta}) = \frac{\partial}{\partial \theta_p} \log L(x_1, \dots, x_n, \boldsymbol{\theta}) = \sum_{i=1}^n S_p(x_i, \boldsymbol{\theta}),$$

$$S_{pqn}(x_1, \dots, x_n, \boldsymbol{\theta}) = \frac{\partial}{\partial \theta_q} S_{pn}(x_1, \dots, x_n, \boldsymbol{\theta}) = \sum_{i=1}^n \frac{\partial^2}{\partial \theta_p \partial \theta_q} \log f(x_i, \boldsymbol{\theta}).$$

Sea T un estimador insesgado de $\boldsymbol{\theta}$ y sean a_1, \dots, a_r números reales, entonces:

$$\int \sum_{p=1}^r (T_p - \theta_p) a_p L(\cdot, \boldsymbol{\theta}) d\mu^n = 0$$

i.e. $\sum_{p=1}^r a_p T_p$ es un estimador insesgado de $\sum_{p=1}^r a_p \theta_p$. Suponemos que las componentes T_p son linealmente independientes.

Si $f(x, \boldsymbol{\theta})$ es regular con respecto a la primera derivada parcial, tenemos derivando con respecto a θ_q :

$$\int \sum_{p=1}^r (T_p - \theta_p) a_p S_{qn}(\cdot, \boldsymbol{\theta}) L(\cdot, \boldsymbol{\theta}) d\mu^n = a_q, \quad q = 1, \dots, r.$$

Multiplicando por b_q , sumando respecto a q y elevando al cuadrado, tenemos:

$$\left(\sum_{q=1}^r a_q b_q \right)^2 \leq \int \left[\sum_{p=1}^r (T_p - \theta_p) a_p \right]^2 L(\cdot, \boldsymbol{\theta}) d\mu^n \int \left[\sum_{q=1}^r S_{qn}(\cdot, \boldsymbol{\theta}) b_q \right]^2 L(\cdot, \boldsymbol{\theta}) d\mu^n.$$

Si evaluamos $\boldsymbol{\theta}$ en $\boldsymbol{\theta}_0$, se tiene:

$$\frac{\left(\sum_{q=1}^r a_q b_q \right)^2}{n \operatorname{var}_{\boldsymbol{\theta}_0} \left(\sum_{p=1}^r S_{pn}(\cdot, \boldsymbol{\theta}_0) b_p \right)} \leq \operatorname{var}_{\boldsymbol{\theta}_0} \left(\sum_{p=1}^r a_p T_p \right), \quad (6)$$

lo que nos da una cota inferior de la varianza, para el estimador $\sum_{p=1}^r a_p T_p$ del parámetro $\sum_{p=1}^r a_p \theta_{p0}$.

La igualdad ocurre cuando existe una relación lineal entre las variables i.e.

$$K(\boldsymbol{\theta}) \sum_{p=1}^r (T_p - \theta_p) a_p \equiv \sum_{p=1}^r S_{pn}(\cdot, \boldsymbol{\theta}) b_p, \quad \mu^n\text{-c.p.d.} \quad (7)$$

donde $K(\boldsymbol{\theta})$ no depende de (x_1, \dots, x_n) y ninguno de los conjuntos a_1, \dots, a_r y b_1, \dots, b_r son idénticamente nulos. Si un estimador T de $\boldsymbol{\theta}_0$ satisface la ecuación anterior diremos que el estimador es eficiente.

Sin embargo esta definición es difícil de trabajar. Vamos a obtener otro resultado más simple.

Sea $B = (B_{pq})$ la matriz de covarianza de T en $\theta = \theta_0$, entonces la matriz de covarianza de $\sqrt{n}T = \sqrt{n}(T_1, \dots, T_r)$ en $\theta = \theta_0$, es nB .

La matriz de covarianza en $\theta = \theta_0$ del vector $\frac{1}{\sqrt{n}}(S_{1n}, \dots, S_{rn})$ es la matriz de información de Fisher $I_{\theta_0} = (\text{cov}_{\theta_0}(S_p, S_q)) = (I_{pq})$. En efecto:

$$\frac{1}{\sqrt{n}}S_{pn}(x_1, \dots, x_n, \theta) - \frac{1}{\sqrt{n}}S_{qn}(x_1, \dots, x_n, \theta) = \sum_{i=1}^r \sum_{j=1}^r S_p(x_i, \theta) S_q(x_j, \theta),$$

por lo que:

$$E_{\theta_0} \left(\frac{1}{n} S_{pn} S_{qn} \right) = \frac{1}{n} \sum_{k=1}^n E_{\theta_0} (S_p \cdot S_q) = \text{cov}_{\theta_0}(S_p, S_q).$$

Además como las componentes T_1, \dots, T_p son linealmente independientes, se tiene que su matriz de covarianza es definida positiva, es decir $|B| > 0$. Igualmente I_{θ_0} es definida positiva y $|I_{\theta_0}| > 0$. De la ecuación (6) se deduce:

$$\left(\sum_{p=1}^r \sum_{q=1}^r \delta_{pq} a_q b_p \right)^2 \leq \left(\sum_{p=1}^r \sum_{q=1}^r B_{pq} a_p a_q \right) \left(\sum_{p=1}^r \sum_{q=1}^r I_{pq} b_q b_p \right). \quad (6')$$

Además la desigualdad (6') se tiene cuando la matriz de covarianza de $\sqrt{n}T_1, \dots, \sqrt{n}T_r, \frac{1}{\sqrt{n}}S_{1n}, \dots, \frac{1}{\sqrt{n}}S_{rn}$ es semidefinida positiva, lo que implica que:

$$1 = \det(\delta_{pq}) \leq \det(B_{pq}) \det(I_{pq}) = |B| \cdot |I_{\theta_0}| \quad (8)$$

y tenemos que la igualdad ocurre solamente si tenemos (7). Lo anterior sugiere definir la eficiencia de un estimador r -dimensional T de θ_0 por:

$$e_T = \frac{1}{|B| \cdot |I_{\theta_0}|}.$$

Un estimador T es eficiente si $e_T = 1$. En resumen tenemos el siguiente teorema.

Teorema 4.3.1 Sea (x_1, \dots, x_n) una muestra de función de densidad $f(x, \theta_0)$, con $\theta_0 \in \Theta_{r0}$ y se considera que $f(x, \theta)$ es regular respecto a las segundas derivadas en Θ_{r0} .

Si T es un estimador insesgado de θ_0 , cuyas componentes son linealmente independientes y su matriz de covarianza B existe y es definida positiva, entonces para cualquier par de conjuntos reales

a_1, \dots, a_r y b_1, \dots, b_r no idénticamente nulos, se tienen las desigualdades (6) y (8). Si T es un estimador eficiente de las igualdades de (6) y (8) se tienen, es decir si (7) ocurre.

Queda aún por demostrar (8). Sea V la matriz de covarianza del vector:

$$\left(\sqrt{n}T_1, \dots, \sqrt{n}T_r, \frac{1}{\sqrt{n}}S_{1n}, \dots, \frac{1}{\sqrt{n}}S_{rn} \right),$$

entonces:

$$V = \begin{pmatrix} B & D_\theta \\ D_\theta & I_\theta \end{pmatrix},$$

donde $D_\theta = \left(\text{cov}_\theta(\sqrt{n}T_q, \frac{1}{\sqrt{n}}S_{pn}) \right)$. Pero:

$$\text{cov}_\theta(T_q, S_{pn}) = \int (T_q - \theta_q) \left(\frac{\partial}{\partial \theta_p} \log L(\cdot, \theta) \right) L(\cdot, \theta) d\mu^n = \frac{\partial}{\partial \theta_p} \int T_p L(\cdot, \theta) d\mu^n = \delta_{pq}.$$

Por otro lado, dado que V es semidefinida positiva, I_θ es invertible y se tiene:

$$|V| = |I_\theta| \cdot |B - D_\theta I_\theta^{-1} D'_\theta| = |I_\theta| \cdot |B - I_\theta^{-1}| \geq 0.$$

Además $B - I_\theta^{-1}$ es semidefinida positiva (ver Anexo B.11.1, página 194), por lo que $|B| > |I_\theta^{-1}|$, o sea $|B| \cdot |I_\theta| \geq 1$ y la igualdad ocurre si y sólo si $|V| = 0$, es decir si existe una relación lineal entre T_1, \dots, T_r y S_{1n}, \dots, S_{rn} , o sea la ecuación (6) ocurre.

4.3.1 Estimación multidimensional para grandes muestras

Teorema 4.3.2 Sea (X_1, \dots, X_n) una muestra de función de densidad $f(x, \theta_0)$, con $\theta_0 \in \Theta_{r0}$, donde $f(x, \theta)$ es regular con respecto a su primera derivada en Θ_{r0} , entonces si I_θ es definida positiva en Θ_{r0} , tenemos que $(S_{1n}(\cdot, \theta_0), \dots, S_{rn}(\cdot, \theta_0))$ se distribuye asintóticamente como una multinormal $N(0, nI_{\theta_0})$.

Demostración Utilizando el teorema del límite central al caso multivariado obtenemos el resultado, es decir si Y_1, \dots, Y_n son vectores r -dimensionales, $Y'_i = (Y_{i1}, \dots, Y_{ir})$, con media μ y matriz de covarianza V definida positiva, entonces si $\bar{Y}_k = \frac{1}{n} \sum_{i=1}^n Y_{ik}$ (resp. $Z_k = n\bar{Y}_k$), $k = 1, \dots, r$, la variable aleatoria de medias muestrales definida por $(\bar{Y}_1, \dots, \bar{Y}_r)$ (resp. (Z_1, \dots, Z_r)) converge en ley a una distribución multinormal $N(\mu, \frac{1}{n}V)$ (resp. $N(n\mu, nV)$), cuando $n \rightarrow \infty$.

Sabemos que $S_{pn}(x_1, \dots, x_n, \theta_0) = \sum_{i=1}^n S_p(x_i, \theta_0)$ por lo que:

$$(S_{1n}, \dots, S_{rn}) \sim N(0, nI_{\theta_0}), \quad \text{cuando } x \rightarrow \infty.$$

Teorema 4.3.3 Sea (X_1, \dots, X_n) una muestra con función de densidad $f(x, \theta_0)$, donde $f(x, \theta)$ es regular respecto a sus segundas derivadas parciales en Θ_{r0} .

Así $S_{pq}(x, \theta)$, $p, q = 1, \dots, r$ es una función continua en θ , $\forall x \in \Omega$ salvo un conjunto de medida cero, entonces existe una sucesión de soluciones de la ecuación:

$$S_{pq}(x_1, \dots, x_n, \theta) = 0, \quad p = 1, \dots, r,$$

la cual converge c.s. a θ_0 , cuando $n \rightarrow \infty$. Si la solución es única para $n \geq n_0$, la sucesión converge c.s. a θ_0 .

Demostración Sabemos por el teorema 4.2.3 que existe una sucesión de soluciones de la ecuación $S_{pn}(x_1, \dots, x_n, \theta) = 0$, para $p = 1, \dots, r$, la cual converge c.s. a θ_{p0} , para $p = 1, \dots, r$. Al considerar el sistema de ecuaciones simultáneamente se obtiene el resultado. Si la solución es única para $n \geq n_0$:

$$\begin{pmatrix} \hat{\theta}_1(X_1, \dots, X_n) \\ \vdots \\ \hat{\theta}_r(X_1, \dots, X_n) \end{pmatrix}_{n \geq n_0} \xrightarrow{\text{c.s.}} \begin{pmatrix} \hat{\theta}_{10} \\ \vdots \\ \hat{\theta}_{r0} \end{pmatrix}.$$

Teorema 4.3.4 Si (X_1, \dots, X_n) es una muestra de función de densidad $f(x, \theta_0)$, $\theta_0 \in \Theta_{r0}$ y $f(x, \theta)$ es regular con respecto a sus dos primeras derivadas parciales en Θ_{r0} y si el estimador de máxima verosimilitud $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_r)$ que satisface $S_{pq}(\cdot, \theta) = 0$, para $p = 1, \dots, r$, es único a partir de algún n_0 y es medible, entonces $\hat{\theta}$ se distribuye asintóticamente como una multinormal $N(\theta_0, \frac{1}{n} I_{\theta_0}^{-1})$.

Demostración Como $\hat{\theta}$ converge c.s. a θ_0 , $\forall \epsilon > 0$, $\forall \delta > 0$, existe $n(\epsilon, \delta, n_0)$ y existe un conjunto E_n definido por $E_n = \{(x_1, \dots, x_n) \in \Omega^n / \|\hat{\theta}(x_1, \dots, x_n) - \theta_0\| < \delta\}$, tal que:

$$P_{\theta_0}(E_n/n \geq n(\epsilon, \delta, n_0)) > 1 - \epsilon.$$

Así tenemos que en E_n :

$$\frac{1}{\sqrt{n}}(S_{1n}(\cdot, \theta_0), \dots, S_{rn}(\cdot, \theta_0)) = \frac{1}{\sqrt{n}}(S_{1n}(\cdot, \hat{\theta}), \dots, S_{rn}(\cdot, \hat{\theta})) + \frac{1}{n} B_{\theta^*} \sqrt{n}(\theta_0 - \hat{\theta}),$$

donde B_{θ^*} es la matriz de término general $\frac{\partial^2}{\partial \theta_p \partial \theta_q} \log L(\cdot, \theta) \Big|_{\theta=\theta^*}$ y θ^* es una variable aleatoria tal que $\|\theta_0 - \theta^*\| < \|\theta_0 - \hat{\theta}\|$. Por lo tanto:

$$P_{\theta_0} \left(\frac{1}{\sqrt{n}} (S_{1n}(\cdot, \theta_0), \dots, S_{rn}(\cdot, \theta_0)) = \frac{1}{n} B_{\theta^*} \sqrt{n}(\theta_0 - \hat{\theta}) : n \geq n(\epsilon, \delta, n_0) \right) > 1 - \epsilon.$$

Las sucesiones de variables aleatorias:

- a) $\frac{1}{\sqrt{n}} (S_{1n}(\cdot, \theta_0), \dots, S_{rn}(\cdot, \theta_0))$,
- b) $\frac{1}{n} B_{\theta^*} [\sqrt{n}(\theta_0 - \hat{\theta})]$,

convergen en ley si una de ellas lo hace; pero $\frac{1}{n} B_{\theta^*} \xrightarrow{\text{Pr}} -I_{\theta_0}$, entonces las tres sucesiones a), b) y $I_{\theta_0} \sqrt{n}(\hat{\theta} - \theta_0)$ convergen en ley si una de ellas lo hace.

Por el teorema 4.3.2 tenemos:

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{\text{ley}} N(0, I_{\theta_0}^{-1}).$$

4.3.2 Eficiencia asintótica de los estimadores de máxima verosimilitud

– Sea $\theta_0 \in \Theta_{r0}$, se dice que un estimador T de θ_0 es consistente, si cada componente de $T = (T_1, \dots, T_r)$ es un estimador consistente de cada componente de θ_0 .

– Si T es un estimador consistente de θ_0 , cuyas componentes $T = (T_1, \dots, T_r)$ son tales que $\sqrt{n}((T_1 - \theta_{10}), \dots, (T_r - \theta_{r0}))$ tiene por límite una distribución normal $N(0, I_{\theta_0}^{*-1})$, cuando $n \rightarrow \infty$, se define la eficiencia asintótica por:

$$ef(T/\theta_0) = \frac{|I_{\theta_0}^*|}{|I_{\theta_0}|} \leq 1.$$

Corolario 4.3.1 Bajo las hipótesis del teorema 4.2.4, los estimadores de máxima verosimilitud $\hat{\theta}$ tienen eficiencia asintótica 1.

4.3.3 Regiones de confianza multidimensionales

Teorema 4.3.5 Sea (X_1, \dots, X_n) una muestra con función de densidad $f(x, \theta_0)$, $\theta_0 \in \mathbb{R}^r$ y sea $g(x_1, \dots, x_n, \theta)$ una función medible y definida en todo punto $\theta \in \Theta_r$, para todo elemento de Ω^n , salvo un conjunto de medida cero, de modo que cuando la función de repartición es $F(x, \theta)$, la función de repartición de g no depende de θ .

Sea $E_r = \{\mathbf{y} \in \Theta_r : g_1 < g(X_1, \dots, X_n, \mathbf{y}) < g_2\}$, donde g_1 y g_2 son tales que:

$$P_{\theta}(g_1 < g(X_1, \dots, X_n, \theta) < g_2) = \gamma,$$

entonces E_r es un conjunto aleatorio tal que si θ_0 es el verdadero valor de θ :

$$P_{\theta_0}(\theta_0 \in E_r) = \gamma.$$

Demostración La prueba es inmediata.

El conjunto E_r es una región de confianza para θ_0 . Más generalmente si $g_n(x_1, \dots, x_n, \theta)$ es un vector con componentes $(g_{1n}(\cdot, \theta), \dots, g_{sn}(\cdot, \theta))$, cuya función de repartición no depende de θ , cuando $F(x, \theta)$ es la función de repartición de la población.

Sea $E_s^* \subset \mathbb{R}^s$ el conjunto tal que:

$$P((g_{1n}, \dots, g_{sn}) \in E_s^*) = \gamma$$

y sea $E_r = \{\mathbf{y} \in \Theta_r : (g_{1n}(X_1, \dots, X_n, \mathbf{y}), \dots, g_{sn}(X_1, \dots, X_n, \mathbf{y})) \in E_s^*\}$, entonces E_r es una región de confianza para θ_0 , es decir $P(\theta_0 \in E_r) = \gamma$.

Generalmente las funciones $g_n(x_1, \dots, x_n, \theta)$ más interesantes son aquellas que proveen regiones de confianza convexas o al menos simplemente conectadas.

El caso $s = r = 1$ se analizó cuando $g_n(x_1, \dots, x_n, \theta)$ es continua y monótona en θ . Para valores de r y s mayores la situación es más compleja. Sin embargo en adición a las propiedades de convexidad se requiere que E_r sea el más pequeño en algún sentido; pero es difícil para valores pequeños de n . Para valores de n grandes, existen una solución satisfactoria, que veremos.

Ejemplo Para ilustrar estos conceptos supongamos que (X_1, \dots, X_n) es una muestra de distribución $N(\mu_0, \sigma_0^2)$. Sabemos que la función:

$$g(x_1, \dots, x_n, \mu_0, \sigma_0) = \frac{(n-1)s^2 + n(\bar{x} - \mu_0)^2}{\sigma_0^2}$$

es distribuida como χ_n^2 . Así para γ dado se determina χ_γ^2 tal que:

$$P(g < \chi_\gamma^2) = \gamma.$$

Sea $E_2 = \{(y_1, y_2) \in \mathbb{R}^2 / \frac{y_2^2}{n} - \frac{(y_1 - \bar{x})^2}{\chi_\gamma^2} < \frac{(n-1)s^2}{n\chi_\gamma^2}, y_2 > 0\}$.

Si (μ_0, σ_0) es el verdadero parámetro, $P((\mu_0, \sigma_0) \in E_2) = \gamma$.

Consideremos ahora $g_p(x_1, \dots, x_n, \mu_0, \sigma_0)$, $p = 1, 2$, donde $g_1 = \frac{n(\bar{x} - \mu_0)^2}{\sigma_0^2}$, $g_2 = \frac{n-1}{\sigma_0^2} s^2$ las cuales son independientes distribuidas como χ_{10}^2 y χ_{n-1}^2 respectivamente. Si escogemos χ_{10}^2 y χ_{20}^2 tales que:

$$P(g_1 < \chi_{10}^2, g_2 > \chi_{20}^2) = \gamma,$$

el conjunto $E_2^* = \{(y_1, y_2) \in \mathbb{R}^2 / \frac{n(y - \bar{x})^2}{y_2} < \chi_{10}^2, \frac{\sqrt{n-1}s}{y_2} > \chi_{20}^2, y_2 > 0\}$ y tenemos que:

$$P((\mu_0, \sigma_0) \in E_2^*) = \gamma.$$

Los conjuntos E_2 y E_2^* son regiones de confianza determinada por funciones de estimación de una y dos dimensiones respectivamente. Intuitivamente la región E_2^* es más satisfactoria que E_2 , dado que E_2^* es acotada y convexa y E_2 no tiene ninguna de estas propiedades.

Si se escoge a g como la función de densidad de (\bar{X}, S^2) , se tiene una región E_2^{**} la cual es mejor que E_2^* , con el criterio de ser más pequeña.

4.3.4 Regiones de confianza más pequeñas, para grandes muestras

1) Por el teorema 4.3.2 se sabe que si θ_0 es el verdadero parámetro θ , la sucesión de variables aleatorias:

$$U_n(X_1, \dots, X_n, \theta_0) = n\mathbf{h}'_n B^{-1} \mathbf{h}_n, \quad n \geq 1,$$

converge a una distribución χ_r^2 , donde $\mathbf{h}'_n = \frac{1}{n}(S_{1n}, \dots, S_{rn})(X_1, \dots, X_n, \theta_0)$ y $B = (B_{pq}) = B(\theta_0, \theta_0) = I_{\theta_0}$.

2) En las hipótesis del teorema 4.3.4 se tiene que:

$$U_n^* = U_n \quad \text{si } \hat{\theta} \in \Theta_r$$

$$U_n^* = 0 \quad \text{si } \hat{\theta} \notin \Theta_r,$$

para todos los puntos de Ω^n , salvo un conjunto de medida cero, donde:

$$U_n^* = n(\hat{\theta} - \theta_0)' B^{(n)} (\hat{\theta} - \theta_0), \quad B_{pq}^{(n)} = \left(\frac{\partial h_n}{\partial \theta_p} \right)' B^{-1} \left(\frac{\partial h_n}{\partial \theta_q} \right) \Big|_{\theta = \hat{\theta}^*},$$

con $\|\hat{\theta}^* - \theta_0\| < \|\hat{\theta} - \theta_0\|$. Ahora $B_{pq}^{(n)} \xrightarrow{\text{Pr}} B_{pq}$, pues:

$$B_{pq}^{(n)} = \sum_{i=1}^n \sum_{j=1}^n B^{ij} \left(\frac{\partial h_{in}}{\partial \theta_p} \right) \left(\frac{\partial h_{jn}}{\partial \theta_q} \right) \Big|_{\theta=\hat{\theta}^*} \xrightarrow{\text{Pr}} \sum_{i=1}^n \sum_{j=1}^n B^{ij} B_{ip} B_{jq} = B_{pq},$$

cuando $n \rightarrow \infty$, con $(B^{ij}) = B^{-1}$.

Así U_n y U_n^* son sucesiones de variables aleatorias que juntas convergen en ley a χ_r^2 .

Sea el valor χ_γ^2 tal que:

$$\lim_{n \rightarrow \infty} P(U_n < \chi_\gamma^2) = \lim_{n \rightarrow \infty} P(U_n^* < \chi_\gamma^2) = \gamma$$

y sean E_r, E_r^* los conjuntos de puntos $\mathbf{y} \in \Theta_{r0}$ tales que:

$$U_n(X_1, \dots, X_n, \mathbf{y}) < \chi_\gamma^2, \quad U_n^*(X_1, \dots, X_n, \mathbf{y}) < \chi_\gamma^2,$$

respectivamente.

Si θ_0 es el verdadero valor de θ , se tiene:

$$\lim_{n \rightarrow \infty} P(\theta_0 \in E_r) = \lim_{n \rightarrow \infty} P(\theta_0 \in E_r^*) = \gamma.$$

Dado que E_r^* es un elipsoide en Θ_{r0} centrado en θ_0 , su volumen aproximado para n grande es:

$$\frac{K(r)\chi_\gamma}{\sqrt{n|B^{(n)}|}},$$

donde $K(r)$ es una constante que depende sólo de r .

3) Sea $g_n(X_1, \dots, X_n, \theta)$ un vector aleatorio de componentes linealmente independientes $g_{pn}(x_1, \dots, x_n)$,

$p = 1, \dots, r$, con θ derivadas parciales:

$$g_{pq}(x_1, \dots, x_n, \theta) = \frac{\partial}{\partial \theta_q} g_{pn}(x_1, \dots, x_n, \theta), \quad p, q = 1, \dots, r.$$

Se considera que $g_{pq}(\cdot, \theta)$ y $g_{pqn}(\cdot, \theta)$ satisfacen las condiciones:

- $g_{pn}(\cdot, \theta)$ es continua para $\theta \in \Theta_{r0}$, en todo punto de Ω salvo en un conjunto de medida cero.
- Si θ_0 es el verdadero de θ , $g'_n(\cdot, \theta) \xrightarrow{\text{Pr}} B^*(\theta_0, \theta)$ uniformemente con respecto a $\theta \in \Theta_{r0}$, donde $B_{pq}^*(\theta_0, \theta)$ es acotada en Θ_{r0} y $B^*(\theta_0, \theta_0) = B^* \neq 0$.

c) Sea $A_n^*(\theta, \theta') = \int g_n(\cdot, \theta')L(\cdot, \theta)d\mu^n$ de componentes:

$$A_{pn}^*(\theta, \theta') = \int g_{pn}(\cdot, \theta')L(\cdot, \theta)d\mu^n, \quad p = 1, \dots, r$$

y sea $B_n^*(\theta, \theta') = \int g_n'(\cdot, \theta')L(\cdot, \theta)d\mu^n$ matriz de entrada p, q :

$$B_{pqn}^*(\theta, \theta') = \int g_{pqn}(\cdot, \theta')L(\cdot, \theta)d\mu^n, \quad p, q = 1, \dots, r,$$

entonces:

$$\frac{\partial}{\partial \theta'} A_n^*(\theta, \theta') = B_n^*(\theta, \theta'), \quad \forall (\theta, \theta') \in \Theta_{r0} \times \Theta_{r0}$$

$$\lim_{n \rightarrow \infty} B_n^*(\theta, \theta) = B^*(\theta, \theta), \quad \forall \theta \in \Theta_{r0}.$$

d) Consideremos que $\int \sqrt{n}g_{pn}(\cdot, \theta)L(\cdot, \theta)d\mu^n = 0$ (i.e. $A_n^*(\theta, \theta) = 0$) y que:

$$\int [\sqrt{n}g_{pn}(\cdot, \theta)][\sqrt{n}g_{qn}(\cdot, \theta)]L(\cdot, \theta)d\mu^n = C_{pqn}(\theta).$$

e) Si θ_0 es el verdadero valor de θ , entonces:

$$\sqrt{n}g_n(\cdot, \theta_0) = \sqrt{n}(g_{1n}(\cdot, \theta_0), \dots, g_{rn}(\cdot, \theta_0)) \xrightarrow{\text{ley}} N(0, C),$$

donde $C_{pq} = \lim_{n \rightarrow \infty} C_{pqn}(\theta_0)$ y C es definida positiva.

Definición 4.3.1 Un vector satisfaciendo las condiciones anteriores se llama función vector de estimación regular para θ_0 .

Observemos que el vector $\mathbf{h}_n(\theta) = \frac{1}{n}(S_{1n}(\cdot, \theta), \dots, S_{rn}(\cdot, \theta))$ es una función vector de estimación regular.

4) Se puede probar que para una función vector de estimación regular, si existe una única sucesión $(\tilde{\theta}_1, \dots, \tilde{\theta}_r)$, para $n \geq n_0$, para algún n_0 , de las ecuaciones:

$$g_{pn}(x_1, \dots, x_n, \theta) = 0, \quad p = 1, \dots, r, \quad n \geq n_0$$

y si θ_0 es el verdadero valor de θ , la sucesión de variables aleatorias:

$$(\tilde{\theta}_1 - \theta_{10}, \dots, \tilde{\theta}_r - \theta_{r0}) \xrightarrow{\text{ley}} N(0, D^{-1}),$$

cuando $n \rightarrow \infty$, donde $D_{pq} = \sum_{i=1}^n \sum_{j=1}^n C^{ij} B_{ip}^* B_{jq}^*$ i.e. $D = B^{*'} C^{-1} B^*$.

Por un argumento similar al utilizado en la sección 4.3.4.1) tenemos:

$$V_n = n \mathbf{h}_n^{*'} C^{-1} \mathbf{h}_n^* \xrightarrow{\text{ley}} \chi_r^2,$$

donde $\mathbf{h}_n^{*'} = \frac{1}{n} (g_{1n}(\cdot, \boldsymbol{\theta}_0), \dots, g_{rn}(\cdot, \boldsymbol{\theta}_0))$ y la variable aleatoria:

$$V_n^* = n(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})' D(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{\text{ley}} \chi_r^2,$$

donde $D_{pq}^{(n)} = \sum_{i=1}^n \sum_{j=1}^n C^{ij} g_{ipn}(\cdot, \tilde{\boldsymbol{\theta}}^*) g_{jqn}(\cdot, \tilde{\boldsymbol{\theta}}^*)$ y $\|\tilde{\boldsymbol{\theta}}^* - \boldsymbol{\theta}\| < \|\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|$.

Sean E_r' y $E_r^{*'}$ los conjuntos de puntos $\mathbf{y} \in \Theta_{r0}$ tales que:

$$V_n(X_1, \dots, X_n, \mathbf{y}) < \chi_\gamma^2, \quad V_n^*(X_1, \dots, X_n, \mathbf{y}) < \chi_\gamma^2,$$

entonces si $\boldsymbol{\theta}_0$ es el verdadero valor de $\boldsymbol{\theta}$:

$$\lim_{n \rightarrow \infty} P(\boldsymbol{\theta}_0 \in E_r') = \lim_{n \rightarrow \infty} P(\boldsymbol{\theta}_0 \in E_r^{*'}) = \gamma$$

Como $E_r^{*'}$ es un elipsoide en Θ_{r0} , centrado en $\boldsymbol{\theta}_0$, su volumen es:

$$\frac{K(r) \chi_\gamma}{\sqrt{n} |D^{(n)}|}.$$

El coeficiente de los volúmenes de E_r' y $E_r^{*'}$ converge en probabilidad, cuando $n \rightarrow \infty$, al número r dado por $r = \frac{\sqrt{|D|}}{\sqrt{|B|}}$. Pero $D = B^{*'} C^{-1} B^*$, por lo que:

$$r = \frac{|B^*|}{\sqrt{|C||B|}}.$$

Por la condición 4) se tiene que:

$$\int g_{pqn}(\cdot, \boldsymbol{\theta}) L(\cdot, \boldsymbol{\theta}) d\mu^n + \int g_{pn}(\cdot, \boldsymbol{\theta}) S_{qn}(\cdot, \boldsymbol{\theta}) L(\cdot, \boldsymbol{\theta}) d\mu^n = 0$$

y en $\boldsymbol{\theta} = \boldsymbol{\theta}_0$:

$$E_{\boldsymbol{\theta}_0}(g_{pqn}(\cdot, \boldsymbol{\theta}_0)) = -\text{cov}_{\boldsymbol{\theta}_0} \left(\sqrt{n} g_{pn}(\cdot, \boldsymbol{\theta}_0), \frac{S_{qn}(\cdot, \boldsymbol{\theta}_0)}{\sqrt{n}} \right).$$

Pero $E_{\boldsymbol{\theta}_0}(g_{pqn}(\cdot, \boldsymbol{\theta}_0)) = B_{pqn}^*(\boldsymbol{\theta}_0, \boldsymbol{\theta}_0) = B_{pqn}^*$, así que la variable aleatoria:

$$\left(\sqrt{n} g_{1n}(\cdot, \boldsymbol{\theta}_0), \dots, \sqrt{n} g_{rn}(\cdot, \boldsymbol{\theta}_0), \frac{1}{\sqrt{n}} S_{1n}(\cdot, \boldsymbol{\theta}_0), \dots, \frac{1}{\sqrt{n}} S_{rn}(\cdot, \boldsymbol{\theta}_0) \right),$$

tiene por matriz de covarianza:

$$\begin{pmatrix} C & B^* \\ B^* & B \end{pmatrix}$$

i.e. $|C||B| \geq |B^*|^2$, o sea $r = \frac{|B^*|^2}{|C||B|} \leq 1$.

Como E_r y E_r^* son asintóticamente equivalentes a E_r^* y $E_r^{*'}$ respectivamente, no existe una función vector de estimación regular para $\theta_0 g(x_1, \dots, x_n, \cdot)$ cuya región de confianza para θ_0 sea asintóticamente más pequeña que la región de confianza para θ_0 , de $h_n(x_1, \dots, x_n, \theta)$.

Teorema 4.3.6 Sea (X_1, \dots, X_n) una muestra con función de densidad $f(x, \theta_0)$, donde $\theta_0 \in \Theta_{r,0}$ y $f(x, \theta_0)$ es regular con respecto a las segundas derivadas parciales en $\Theta_{r,0}$.

Si $g_{pn}(x_1, \dots, x_n, \theta)$, $p = 1, \dots, r$, es una función vector de estimación regular para θ_0 , E_r^l es una región de confianza para θ_0 de coeficientes asintótico γ . La región de confianza para θ_0 , asintóticamente más pequeña de coeficientes asintótico γ , está dado por E_r .

4.4 Ejercicios

1. Probar que las dos distribuciones $f_1(x, \theta) = \frac{1}{\theta}$, $0 < x \leq \theta$ y $f_2(x, \theta') = \frac{1}{\theta'}$, $0 < x \leq \theta'$, no son absolutamente continuas una respecto a la otra, a menos que $\theta = \theta'$.
2. Sea X variable aleatoria de función de repartición $F(x, \theta) = (1 - e^{-\theta x}) \mathbb{1}_{[0, +\infty[}(x)$, $0 < \theta < \infty$. Probar que $E_\theta(S(x, \theta)) = 0$ y determine $\text{var}_\theta(S(X, \theta))$, así como $H(\theta, \theta')$ y probar que para θ fijo, tiene un máximo con respecto a θ' en $\theta' = \theta$.
3. Sea X una variable aleatoria discreta tal que $P_\theta(X = x) = (1 - \theta)\theta^{x-1}$, $x = 1, 2, \dots$, con $0 < \theta < 1$. Probar que $E_\theta(S(X, \theta)) = 0$ y calcule $\text{var}_\theta(S(X, \theta))$.
4. Si (X_1, \dots, X_n) es una muestra de una distribución con momento de orden k finito, probar que $\frac{1}{n} \sum_{i=1}^n X_i^k$ es un estimador consistente de $E(X^k)$.
5. Si \bar{X} es la medida muestral de una distribución $B(1, p)$, probar que \bar{X} es suficiente para p . Además probar que \bar{X} es suficiente como estimador para p .
6. a) Sea (X_1, \dots, X_n) una muestra con densidad $f(x, \theta) = \theta e^{-\theta x}$, $x > 0$, $\theta > 0$. Probar que \bar{X} es suficiente para θ y que $(n-1)/n\bar{X}$ es el único estimador insesgado para θ dependiendo de \bar{X} .

- b) Probar que \bar{X} es el estimador de máxima verosimilitud para $\frac{1}{\theta}$ y que es un estimador suficiente.
7. Si $\hat{\theta}_1$ y $\hat{\theta}_2$ son estimadores eficientes para θ , probar que el coeficiente de correlación entre ellos es 1.
8. a) Si $\hat{\theta}$ es un estimador eficiente de θ y $\tilde{\theta}$ es otro estimador insesgado de θ , de varianzas respectivas σ^2 y $k\sigma^2$, con $k > 1$, probar que su correlación es $\frac{1}{\sqrt{k}}$.
- b) Si $\hat{\theta}$ es eficiente para θ y $\hat{\theta}_1, \hat{\theta}_2$ son estimadores ineficientes pero insesgados, ambos con varianza $k\sigma^2$ ($k > 1$), probar que la correlación entre $\hat{\theta}_1$ y $\hat{\theta}_2$ es $\frac{2-k}{k}$.
9. Sean $(X_{(1)}, \dots, X_{(n)})$ las estadísticas de orden de una muestra de densidad $f(x, \theta) = \frac{1}{\theta}$, si $0 \leq x \leq \theta$. Probar que $X_{(n)}$ es una estadística suficiente para θ y que $\frac{n-1}{n}X_{(n)}$ es un estimador insesgado para θ . Probar que $(X_{(n)}, \frac{X_{(n)}}{\sqrt[1-\gamma]})$ es un intervalo de confianza para θ de coeficiente γ .
10. Si $X_{(1)}, \dots, X_{(n)}$ son las estadísticas de orden para una muestra de densidad $f(x, \theta) = \frac{1}{\theta_2 - \theta_1}$, si $\theta_1 < x < \theta_2$, con $0 < \theta_1 < \theta_2$. Probar que $(X_{(1)}, X_{(n)})$ es una estadística suficiente para (θ_1, θ_2) . Probar que $\frac{nX_{(1)}}{n-1} - \frac{X_{(n)}}{n-1}$ y $\frac{nX_{(n)}}{n-1} - \frac{X_{(1)}}{n-1}$ son estimadores de varianza mínima de θ_1 y θ_2 respectivamente. Además probar que $\frac{X_{(1)} + X_{(n)}}{2}$ y $\frac{n+1}{n-1}(X_{(n)} - X_{(1)})$ son estimadores de mínima varianza de $\frac{(\theta_1 + \theta_2)}{2}$ y $\theta_2 - \theta_1$.
11. Verificar que la varianza de todo estimador de σ^2 en una muestra de un tamaño n , de distribución $N(\mu, \sigma^2)$ es al menos $\frac{2\sigma^4}{n}$.
12. Si X_1, \dots, X_n es una muestra de distribución $N(\mu, \sigma^2)$, con μ conocida. Probar que $\frac{1}{n} \sqrt{\frac{\pi}{2}} \sum_{i=1}^n |x_i - \mu|$ es un estimador insesgado de σ , con eficiencia asintótica $\frac{1}{\pi - 2}$.
13. Sea X_1, \dots, X_n una muestra de distribución $C(1, \mu)$. Probar que la eficiencia asintótica de la mediana muestral para estimar μ es $\frac{8}{\pi^2}$.
14. a) Si X_1, \dots, X_n es una muestra de densidad $\Gamma(k+1, \lambda)$, donde $\lambda > 0$ y k es una conocida, probar que el estimador $\hat{\lambda}$ de máxima verosimilitud es $\frac{(k+1)}{\bar{X}}$. Verificar que este estimador es sesgado pero consistente y que la distribución asintótica es $N(\lambda, \frac{\lambda^2}{n(k+1)})$, cuando $n \rightarrow \infty$.
- b) Probar que $\sqrt{k+1} \log \hat{\lambda}$ es una función de $\hat{\lambda}$ de distribución asintótica normal con varianza $\frac{1}{n}$, cuando $n \rightarrow \infty$.

15. Sea X_1, \dots, X_n una muestra de distribución $B(1, p)$. Probar que el estimador de máxima verosimilitud \hat{p} de p es \bar{X} y que su distribución es normal $N(p, \frac{p(1-p)}{n})$, asintóticamente. Verificar que $\text{sen}^{-1}(2\hat{p}-1)$ tiene una distribución asintótica normal con varianza $\frac{1}{n}$, cuando $n \rightarrow \infty$.
16. Si \bar{X} es la medida de una muestra de distribución $B(1, p)$, probar que el intervalo de confianza más pequeño de coeficiente γ , está dado por los valores de p que satisfacen $\frac{(\bar{X} - p) \sqrt{n}}{\sqrt{p(1-p)}} = \pm y_\gamma$, donde $P(|N(0, 1)| < y_\gamma) = \gamma$, cuando $n \rightarrow \infty$.
17. Sea \bar{X} la medida de muestra de tamaño n de distribución $P(\lambda)$. Verificar que el intervalo de confianza para λ , más pequeño de coeficiente γ , cuando $n \rightarrow \infty$, está dado por los valores de γ tales que $\frac{(\bar{X} - \lambda) \sqrt{n}}{\sqrt{\lambda}} = \pm y_\gamma$.
18. Si (X_{1i}, \dots, X_{ki}) , $i = 1, \dots, n$, es una muestra de tamaño n de distribución $M(1, p_1, \dots, p_k)$ verificar que el estimador de máxima verosimilitud para (p_1, \dots, p_k) es $(\bar{X}_1, \dots, \bar{X}_k)$ y que la matriz de covarianza de los estimadores es $(\frac{1}{n}(\delta_{ij}p_i - p_i p_j))$.
19. Sea (X_{1i}, \dots, X_{ki}) , $i = 1, \dots, n$, una muestra de tamaño n de distribución normal $N(\mu, \Sigma)$. Probar que el estimador de máxima verosimilitud para μ , es el vector $(\bar{X}_1, \dots, \bar{X}_k)$ y que el estimador de máxima verosimilitud para Σ es $(\frac{n-1}{n} S_{ij})$, con $S_{ij} = \frac{1}{n-1} \sum_{k=1}^n (X_{ik} - \bar{X}_i)(X_{jk} - \bar{X}_j)$.
20. Supongamos que Y_1, \dots, Y_n son variables aleatorias independientes de distribuciones $N(\beta_1 x_{11} + \dots + \beta_k x_{k1}, \sigma^2), \dots, N(\beta_1 x_{1n} + \dots + \beta_k x_{kn}, \sigma^2)$ respectivamente, donde $x_{11}, \dots, x_{k1}, \dots, x_{1n}, \dots, x_{kn}$ son constantes.

$$\text{Sea } a_{pq} = \sum_{k=1}^n x_{pk} x_{qk}, \quad p, q = 1, \dots, k, \quad a_{p0} = \sum_{k=1}^n x_{pk} y_k, \quad p = 1, \dots, k.$$

Probar que el estimador de máxima verosimilitud para $(\beta_1, \dots, \beta_k)$ es:

$$\left(\sum_{p=1}^k a_{p0} a^{p1}, \dots, \sum_{p=1}^k a_{p0} a^{pk} \right),$$

donde $(a^{pq}) = (a_{pq})^{-1}$, con (a_{pq}) no singular.

Probar que el estimador de máxima verosimilitud para σ^2 es:

$$\frac{1}{n} \frac{|a_{p^* q^*}|}{|a_{pq}|}, \quad p^*, q^* = 0, \dots, k; \quad p, q = 1, \dots, k,$$

$$\text{donde } a_{00} = \sum_{i=1}^n y_i^2.$$

21. Sea (X_1, \dots, X_n) una variable aleatoria con función de densidad $f_n(x_1, \dots, x_n, \theta_0)$ y sea $\tilde{\theta}$ un estimador insesgado para θ_0 . Probar que $\forall \delta > 0$:

$$\text{var}_{\theta_0}(\tilde{\theta}) \geq \frac{1}{E_{\theta_0}(g^2)},$$

$$\text{donde } g = \frac{f_n(x_1, \dots, x_n, \theta_0 + \delta) - f_n(x_1, \dots, x_n, \theta_0)}{\delta f_n(x_1, \dots, x_n, \theta_0)} \text{ i.e. } \text{var}_{\theta_0}(\tilde{\theta}) \geq \frac{1}{\inf_{\delta} E_{\theta_0}(g^2)}.$$

22. a) Sea (X_1, \dots, X_n) una variable aleatoria de densidad $f_n(x_1, \dots, x_n, \theta)$, tal que T es una estadística suficiente para θ . Se supone que $f_n(\cdot, \theta)$ y $g(t, \theta)$ tiene derivadas parciales de segundo orden en todo punto de $\Omega^n \times \Theta$, excepto en un conjunto de medida cero. Probar que $f_n(x_1, \dots, x_n, \theta)$ es de la forma:

$$e^{K_1(\theta)S(t) + K_2(\theta) + h(x_1, \dots, x_n)},$$

donde $K_1(\theta)$ y $K_2(\theta)$ no dependen de (x_1, \dots, x_n) , h_n no depende de θ y $S(t)$ depende de t y no de θ .

b) Sea $u(S(T))$ una función de T de medida $q(\theta)$ y sea $v(S(T))$ otra función de medida $q(\theta)$, entonces si $w(S(T)) = u(S(T)) - v(S(T))$ se tiene:

$$\int w(s) e^{K_1(\theta)s + K_2(\theta) + h_n} d\mu = 0, \quad \forall \theta \in \Theta.$$

Probar que si esta expresión se puede derivar bajo el signo de integral con respecto a θ y $w(s)$ se puede representar por una serie de Taylor¹, entonces:

$$E_{\theta}(w(S)^2) = 0, \quad \forall \theta \in \Theta, \text{ i.e. } w(s) = 0, \quad \mu\text{-c.p.d.}, \forall \theta \in \Theta.$$

¹**Brook Taylor (1685-1731)** Nace el 18 de agosto de 1685 en Edmonton, Inglaterra. Muere el 29 de diciembre de 1731 en Londres, Inglaterra. En 1708 Taylor produjo una solución al problema del centro de oscilación, la cual desde que fue difundida hasta 1724, resultó ser la disputa principal con Johann Bernoulli.

En "Los métodos de incrementos directo e inverso" Taylor (1715) agrega a las matemáticas una nueva rama llamada ahora "El cálculo de las diferencias finitas", inventa la integración por partes y descubrió la célebre fórmula conocida como la Serie de Taylor. La importancia de esta fórmula no fue reconocida hasta 1772, cuando Lagrange proclamó los principios básicos del cálculo diferencial.

Taylor también desarrolló los principios fundamentales de la perspectiva en "Perspectivas lineales" (1715), junto con "Los nuevos principios de la perspectiva lineal".

Taylor da cuenta de un experimento para descubrir las leyes de la atracción magnética (1715) y un método no probado para aproximar las raíces de una ecuación, dando un método nuevo para logaritmos computacionales (1717).

Taylor fue elegido socio de la Real Sociedad en 1712 y fue nombrado en ese año para integrar un comité para la adjudicación de las demandas de Newton y de Leibniz de haber inventado el cálculo.

Capítulo 5

Complementos de la teoría de pruebas de hipótesis

Pruebas de hipótesis compuestas con propiedades óptimas, para muestras finitas, son obtenidas por un principio importante debido a Neyman y Pearson: el principio del cociente de verosimilitud.

La prueba del cociente de verosimilitud, para grandes muestras, tiene propiedades asintóticas óptimas en ciertas condiciones, las cuales vamos a considerar en detalle. La prueba del cociente de verosimilitud para grandes muestras, tiene su fundamento en la teoría estadística, así como la teoría de estimación que hemos considerado.

5.1 Distribución asintótica del cociente de verosimilitud para grandes muestras

Consideremos un conjunto $\Theta \subset \mathbb{R}$, que contiene un intervalo abierto Θ_0 tal que $\theta_0 \in \Theta_0 \subset \Theta$, así como la hipótesis simple $H_0 = \{\theta_0\}$.

Asumimos que para n grande, existe un estimador de máxima verosimilitud $\hat{\theta}_n$ para θ_0 tal que la sucesión $\hat{\theta}_n \xrightarrow{\text{Pr}} \theta_0$. La sucesión de cocientes de verosimilitud está dada por:

$$\Lambda = \frac{f(X_1, \dots, X_n, \theta_0)}{f(X_1, \dots, X_n, \hat{\theta}_n)} = \prod_{i=1}^n \frac{f(X_i, \theta_0)}{f(X_i, \hat{\theta}_n)}$$

y la prueba asociada a Λ se define por:

$$\Phi = \begin{cases} 1 & \text{si } \Lambda < \lambda_0 \\ 0 & \text{si } \Lambda \geq \lambda_0. \end{cases}$$

Vamos a probar que bajo ciertas circunstancias tenemos que:

$$-2 \log \Lambda \xrightarrow{\text{ley}} \chi_1^2,$$

si H_0 es cierta. Suponemos que $f(x, \theta)$ es regular respecto a la segunda derivada θ en θ_0 , i.e.

$$H(\theta_0, \theta) = \int \log f(\cdot, \theta) L(\cdot, \theta_0) d\mu$$

se puede derivar dos veces con respecto a θ y tenemos:

$$H'(\theta_0, \theta_0) = 0$$

$$H''(\theta_0, \theta_0) = -B^2(\theta_0, \theta_0).$$

Así $H(\theta_0, \theta)$ tiene un máximo $H(\theta_0, \theta_0)$ en $\theta = \theta_0$.

Si $f_1(x)$, $f_2(x)$ y $g(x)$ son funciones de densidad absolutamente continuas unas respecto a las otras, Kullback llama información media, o información integral para discriminar $f_1(x)$ contra $f_2(x)$, por observación de $g(x)$, al término:

$$I(1, 2) = \int \log \frac{f_1}{f_2} g d\mu.$$

Así:

$$H(\theta_0, \theta_0) - H(\theta_0, \theta_1) = \int \log \frac{f(\cdot, \theta_0)}{f(\cdot, \theta_1)} f(\cdot, \theta_0) d\mu$$

es, en el sentido de Kullback, la información integral para discriminar $f(x, \theta_0)$ contra $f(x, \theta_1)$ por observación de $f(x, \theta_0)$.

Observemos que $B^2(\theta_0, \theta_0)$ definida por Fisher como la información sobre θ_0 por observación de $f(x, \theta_0)$, es la segunda derivada de la información integral de Kullback $[H(\theta_0, \theta_0) - H(\theta_0, \theta)]$ en $\theta = \theta_0$. La cantidad $B^2(\theta_0, \theta_0)$ mide la curvatura de $H(\theta_0, \theta)$ en su valor máximo. En resumen:

Teorema 5.1.1 Si $f(x, \theta)$ es regular con respecto a su segunda θ derivada en Θ_0 , entonces $H(\theta_0, \theta)$ tiene un máximo relativo en $\theta = \theta_0$. La segunda derivada de $[H(\theta_0, \theta_0) - H(\theta_0, \theta)]$ en $\theta = \theta_0$ es $B^2(\theta_0, \theta_0)$.

Teorema 5.1.2 Sea (X_1, \dots, X_n) una muestra de función de densidad $f(x, \theta_0)$, donde $f(x, \theta)$ es regular con respecto a su primera derivada en Θ_0 , entonces las sucesiones:

$$\frac{1}{n} \sum_{i=1}^n \log f(X_i, \theta_0), \quad \frac{1}{n} \sum_{i=1}^n \log f(X_i, \hat{\theta}),$$

convergen en probabilidad a $H(\theta_0, \theta_0)$, cuando $n \rightarrow \infty$.

Demostración Dado que la muestra tiene función de densidad $f(x, \theta_0)$, entonces tenemos que $\frac{1}{n} \sum_{i=1}^n \log f(X_i, \theta_0)$ es la media muestral de una distribución de media $H(\theta_0, \theta_0)$ i.e. converge en probabilidad a $H(\theta_0, \theta_0)$, cantidad que existe por la regularidad de $f(x, \theta)$. Además bajo estas condiciones $\hat{\theta}_n \xrightarrow{\text{Pr}} \theta_0$, cuando $n \rightarrow \infty$, o sea:

$$\frac{1}{n} \sum_{i=1}^n \log f(X_i, \hat{\theta}_n) \xrightarrow{\text{Pr}} H(\theta_0, \theta_0),$$

cuando $n \rightarrow \infty$.

Teorema 5.1.3 Si $\hat{\theta}_n$ es asintóticamente normal como en las condiciones del teorema 4.2.4, entonces:

$$\lim_{n \rightarrow \infty} P_{\theta_0}(-2 \log \Lambda < \chi^2) = \frac{1}{\sqrt{2\pi}} \int_0^{\chi^2} u^{-\frac{1}{2}} e^{-\frac{1}{2}u} du,$$

es decir si H_0 es cierta $-2 \log \Lambda \xrightarrow{\text{ley}} \chi_1^2$.

Demostración Como $\hat{\theta} \xrightarrow{\text{c.s.}} \theta_0$, cuando $n \rightarrow \infty$, $\forall \epsilon > 0$, $\forall \delta > 0$, existe $n_{\epsilon, \delta}$ tal que:

$$P_{\theta_0}(E_n : \forall n \geq n_{\epsilon, \delta}) > 1 - \epsilon.$$

Sobre E_n tenemos que:

$$\sum_{i=1}^n \log f(X_i, \theta_0) = \sum_{i=1}^n \log f(X_i, \hat{\theta}) + \frac{1}{2} \sum_{i=1}^n \left(\frac{\partial^2}{\partial \theta^2} \log f(X_i, \theta) \right)_{\theta=\theta^*} (\theta_0 - \hat{\theta})^2,$$

$\forall n \geq n_{\epsilon, \delta}$, donde θ^* es una variable aleatoria tal que $|\theta_0 - \theta^*| < |\theta_0 - \hat{\theta}|$, es decir:

$$-2 \sum_{i=1}^n \log \frac{f(X_i, \theta_0)}{f(X_i, \hat{\theta})} = -\frac{1}{n} \sum_{i=1}^n \left(\frac{\partial^2}{\partial \theta^2} \log f(X_i, \theta) \right)_{\theta=\theta^*} [\sqrt{n}(\theta_0 - \hat{\theta})]^2,$$

para todo $n \geq n_{\epsilon, \delta}$, con probabilidad mayor que $1 - \epsilon$. Pero:

$$-\frac{1}{n} \sum_{i=1}^n \left(\frac{\partial^2}{\partial \theta^2} \log f(X_i, \theta) \right)_{\theta=\theta^*} \xrightarrow{\text{Pr}} B^2(\theta_0, \theta_0)$$

y además $[\sqrt{n}B(\theta_0, \theta_0)(\theta_0 - \hat{\theta})]^2$ converge en ley a un χ_1^2 , es decir:

$$-2 \log \Lambda \xrightarrow{\text{ley}} \chi_1^2,$$

cuando H_0 es verdadera.

5.1.1 Consistencia de la prueba del cociente de verosimilitud

Teorema 5.1.4 Si $f(x, \theta)$ es regular con respecto a su segunda θ derivada, la prueba del cociente de verosimilitud es una prueba consistente para H_0 .

Demostración Por el teorema 5.1.3 tenemos que $\lim_{n \rightarrow \infty} P_{\theta_0}(-2 \log \Lambda > \chi_\alpha^2) = \alpha$, donde $P(\chi_1^2 > \chi_\alpha^2) = \alpha$.

Sea $\theta_1 \neq \theta_0$ un elemento de Θ_0 y consideramos $\Lambda = \Lambda' \nu$, donde:

$$\Lambda' = \frac{f(X_1, \dots, X_n, \theta_1)}{f(X_1, \dots, X_n, \hat{\theta})}, \quad \nu = \frac{f(X_1, \dots, X_n, \theta_0)}{f(X_1, \dots, X_n, \theta_1)}$$

Consideremos:

$$u_n = -2 \log \Lambda', \quad nv_n = -2 \log \nu.$$

Si θ_1 es el verdadero valor de θ , por el teorema 4.2.3, $u_n \xrightarrow{\text{ley}} \chi_1^2$, cuando $n \rightarrow \infty$.

Además:

$$\nu_n = 2 \left(\frac{1}{n} \sum_{i=1}^n \log f(X_i, \theta_1) - \frac{1}{n} \sum_{i=1}^n \log f(X_i, \theta_0) \right)$$

y si θ_1 es el verdadero valor de θ , $\nu_n \xrightarrow{\text{Pr}} 2[H(\theta_1, \theta_1) - H(\theta_1, \theta_0)] > 0$, por lo que:

$$\lim_{n \rightarrow \infty} P_{\theta_1}(u_n + nv_n > \chi_\alpha^2) = 1.$$

Así definimos:

$$\Phi = \begin{cases} 1 & \text{si } u_n + nv_n > \chi_\alpha^2 \\ 0 & \text{si } u_n + nv_n \leq \chi_\alpha^2, \end{cases}$$

entonces la potencia:

$$\beta_\Phi(\theta) \rightarrow \zeta(\theta) = \begin{cases} \alpha & \theta = \theta_0 \\ 1 & \theta \neq \theta_0, \end{cases}$$

cuando $n \rightarrow \infty$, es decir que Φ es una prueba consistente.

5.1.2 Potencia asintótica de la prueba del cociente de verosimilitud

Si consideramos las hipótesis del teorema 4.2.4, para grandes muestras, la prueba del cociente de verosimilitud Φ con un error de tipo I igual a α , determinado por un χ_1^2 , es una prueba consistente cuya

potencia converge a $\zeta(\theta)$. Vamos a demostrar que bajo ciertas condiciones la potencia de cualquier prueba consistente de nivel α convergiendo a $\zeta(\theta)$, no lo hace más que rápidamente de Φ .

Para esto volvamos a la función regular de estimación $g_n(x_1, \dots, x_n, \theta)$, pues la sucesión:

$$\Lambda^* = ng_n^2(X_1, \dots, X_n, \theta_0) \xrightarrow{\text{ley}} \chi_1^2.$$

Sea Φ^* la prueba definida por:

$$\Phi^* = \begin{cases} 1 & \text{si } \Lambda^* > \chi_\alpha^2 \\ 0 & \text{si } \Lambda^* \leq \chi_\alpha^2, \end{cases}$$

por las propiedades de regularidad de la función de estimación g , si $\theta_1 \neq \theta_0$ es el verdadero valor de θ con $\theta_1 \in \Theta_0$, las sucesiones:

$$ng_n^2(X_1, \dots, X_n, \theta_0), \quad w_n^* = (g'_n(X_1, \dots, X_n, \tilde{\theta}^*)[\sqrt{n}(\theta_1 - \tilde{\theta}^*) + \sqrt{n}(\theta_0 - \theta_1)])^2,$$

donde $|\tilde{\theta}^* - \theta_0| < |\tilde{\theta} - \theta_0|$, convergen en ley si una de ellas converge.

Como $w_n^* = u_n^* + nv_n^*$, donde $u_n^* = (g'(X_1, \dots, X_n, \tilde{\theta}^*)\sqrt{n}(\theta_1 - \tilde{\theta}^*))^2$, $v_n^* = \frac{1}{n}(w_n^* - u_n^*)$ y si el verdadero valor de θ es θ_1 , la sucesión $u_n^* \xrightarrow{\text{ley}} \chi_1^2$, $v_n^* \xrightarrow{\text{Pr}} v_0^*$, donde $v_0^* = B^{*2}(\theta_1, \theta_1)(\theta_0 - \theta_1)^2$, pues $\frac{1}{n}w_n^* \xrightarrow{\text{Pr}} B^{*2}(\theta_1, \theta_1)(\theta_0 - \theta_1)^2$ y $\frac{1}{n}u_n^* \xrightarrow{\text{Pr}} 0$.

Por otro lado $v_0^* > 0$, por lo tanto:

$$\lim_{n \rightarrow \infty} P_{\theta_1}(u_n^* + nv_n^* > \chi_\alpha^2) = 1.$$

Así definimos la prueba Φ^* por:

$$\Phi^* = \begin{cases} 1 & \text{si } u_n^* + nv_n^* > \chi_\alpha^2 \\ 0 & \text{si } u_n^* + nv_n^* \leq \chi_\alpha^2 \end{cases}$$

y entonces $\beta_{\Phi^*}(\theta) \rightarrow \zeta(\theta)$, cuando $n \rightarrow \infty$, i.e. θ^* es consistente.

Nuestro problema es ahora comparar las potencias de la prueba Φ y la prueba Φ^* . Para resolver este problema definiremos algunos conceptos.

Definición 5.1.1 Si Φ es una prueba para H_0 basada en la muestra (X_1, \dots, X_n) de potencia límite

$\eta(\theta)$, cuando $n \rightarrow \infty$ definida por:

$$\begin{aligned}\eta(\theta_0) &= \alpha \\ \alpha &\leq \eta(\theta) \leq 1, \quad \theta \neq \theta_0,\end{aligned}$$

se dice que Φ es asintóticamente insesgado como prueba de H_0 , de nivel α .

Definición 5.1.2 Sean Φ_1 y Φ_2 dos pruebas asintóticamente insesgadas para la hipótesis H_0 , de nivel α , de potencias asintóticas $\eta_1(\theta)$, $\eta_2(\theta)$ para $\theta \in \Theta_0$.

Si $\eta_1(\theta) \geq \eta_2(\theta)$, con $\eta_1(\theta_0) = \eta_2(\theta_0) = \alpha$, se dice que Φ_1 es asintóticamente más potente que Φ_2 para H_0 .

Si $\eta_1(\theta) = \eta_2(\theta)$, para $\theta \neq \theta_0$, se dice que Φ_1 y Φ_2 son pruebas insesgadas asintóticamente equivalentes.

Sea Φ_c y Φ_c^* pruebas determinadas por desigualdades:

$$u_n + cv_n > \chi_\alpha^2, \quad u_n^* + cv_n^* > \chi_\alpha^2,$$

donde $c > 0$. Como u_n y u_n^* son no negativos vamos a probar que $\forall c > 0$, Φ_c es asintóticamente más potente que Φ_c^* en cuyo caso diremos, adoptando la terminología de Wald, que Φ es más conveniente que Φ^* para probar H_0 .

Si Φ_c y Φ_c^* son asintóticamente equivalentes para todo $c > 0$, diremos que Φ y Φ^* son asintóticamente igualmente convenientes para probar H_0 .

De acuerdo a estas definiciones podemos establecer el siguiente teorema.

Teorema 5.1.5 Sea Φ la prueba el cociente de verosimilitud y Φ^* la prueba definida por la región crítica $nh_n^2(X_1, \dots, X_n, \theta_0) > \chi_\alpha^2$, con $h_n(\cdot, \theta) = \frac{S(\cdot, \theta)}{nB(\theta_0, \theta)}$, entonces si la función de densidad $f(x, \theta)$ es regular respecto a la segunda θ derivada en Θ_0 , Φ y Φ^* son pruebas asintóticamente igualmente convenientes para probar H_0 .

Demostración Sabemos que (u_n, v_n) converge en ley a una distribución degenerada bidimensional, la cual es un χ_1^2 a lo largo de $v = v_0$, $u > 0$, donde $v_0 = 2[H(\theta_1, \theta_1) - H(\theta_1, \theta_0)]$.

Igualmente (u_n^*, v_n^*) converge en ley a una distribución degenerada bidimensional, la cual es un χ_1^2 a lo largo de $v = v_0^*$, $u > 0$, donde $v_0^* = B^{*2}(\theta_1, \theta_1)(\theta_0 - \theta_1)^2$.

Pero $H(\theta_1, \theta_0) = H(\theta_1, \theta_1) + H'(\theta_1, \theta_1)(\theta_0, \theta_1) + \frac{1}{2}H''(\theta_1, \theta_1^*)(\theta_0 - \theta_1)^2$, donde $|\theta_0 - \theta_1^*| < |\theta_0 - \theta_1|$ y como $H'(\theta_1, \theta_1) = 0$, tenemos:

$$2[H(\theta_1, \theta_1) - H(\theta_1, \theta_0)] = -H''(\theta_1, \theta_1^*)(\theta_0 - \theta_1)^2.$$

Como $f(x, \theta)$ es regular respecto a la segunda θ derivada, tenemos que:

$$B^{*2}(\theta_1, \theta_1) \leq B^2(\theta_1, \theta_1)$$

y como $H''(\theta_1, \theta)$ es continua en Θ_0 , entonces $\forall \epsilon > 0$, existe $\delta > 0$ tal que si $|\theta_1 - \theta| < \delta$, implica que $|-H''(\theta_1, \theta_1) + H''(\theta_1, \theta)| < \epsilon$, o sea que:

$$B^2(\theta_1, \theta_1) < -H''(\theta_1, \theta) + \epsilon, \quad \text{si } |\theta_1 - \theta| < \delta.$$

Si $|\theta_0 - \theta_1| < \delta$, entonces $|\theta_0 - \theta_1^*| < \delta$ i.e. $B^{*2}(\theta_1, \theta_1) < -H''(\theta_1, \theta_1^*) + \epsilon$, $\forall \epsilon > 0$, así tenemos que:

$$0 < v_0^* \leq v_0.$$

Ahora:

$$\beta_{\Phi_c}(\theta_1) = P_{\theta_1}(u_n + cv_n > \chi_\alpha^2), \quad \beta_{\Phi_c^*}(\theta_1) = P_{\theta_1}(u_n^* + cv_n^* > \chi_\alpha^2),$$

tiene por límites $\eta_c(\theta_1)$ y $\eta_c^*(\theta_1)$ respectivamente, cuando $n \rightarrow \infty$.

Debido a la convergencia de (u_n, v_n) y (u_n^*, v_n^*) tenemos:

$$\eta_c(\theta_1) = \eta_c^*(\theta_1) = 1, \quad \text{para } \theta_1 \text{ tal que } v_0^* > \frac{\chi_\alpha^2}{c},$$

$$1 \geq \eta_c(\theta_1) > \eta_c^*(\theta_1) > \alpha, \quad \text{para } \theta_1 \text{ tal que } 0 < v_0^* \leq \frac{\chi_\alpha^2}{c}$$

$$\eta_c(\theta_0) = \eta_c^*(\theta_0) = \alpha, \quad \text{para } \theta_1 = \theta_0 \text{ i.e. } v_0^* = 0.$$

Así Φ_c es asintóticamente más potente que Φ_c^* para probar H_0 , $\forall c > 0$ i.e. Φ es asintóticamente más conveniente que Φ^* , para probar H_0 .

Si $v_0^* = v_0$, entonces $1 > \eta_c(\theta_1) = \eta_c^*(\theta_1) > \alpha$, $\forall \theta_1$ tal que $0 < v_0^* \leq \frac{\chi_\alpha^2}{c}$, o sea Φ_c y Φ_c^* son pruebas insesgadas asintótica igualmente convenientes. En resumen tenemos el siguiente teorema.

Teorema 5.1.6 Sea (X_1, \dots, X_n) una muestra con función de densidad $f(x, \theta)$, donde $f(x, \theta)$ es regular con respecto a su segunda θ derivada en Θ_0 . Sea Φ la prueba del cociente de verosimilitud de nivel α y sea $g(x_1, \dots, x_n, \theta)$ una función regular de estimación para θ , induciendo la prueba Φ^* de nivel α , entonces Φ es asintótica igualmente conveniente como Φ para probar H_0 . Además si $g_n = h_n$, Φ_c^* es asintóticamente igualmente conveniente como Φ para probar H_0 .

5.2 Prueba del cociente de verosimilitud de una hipótesis simple

Consideremos $\Theta_r \subset \mathbb{R}^r$ y sea Θ_{r0} un intervalo abierto tal que $\theta_0 \in \Theta_{r0} \subset \Theta_r$.

El cociente de verosimilitud Λ , para $H_0 = \{\theta_0\}$ es:

$$\Lambda = \frac{L(x_1, \dots, x_n, \theta_0)}{L(x_1, \dots, x_n, \theta)} = \prod_{k=1}^n \frac{f(x_k, \theta_0)}{f(x_k, \theta)}$$

y la prueba se define por:

$$\Phi_\Lambda = \begin{cases} 1 & \text{si } \Lambda < \lambda_0 \\ 0 & \text{si } \Lambda \geq \lambda_0. \end{cases}$$

Vamos a probar que $-2 \log \Lambda \xrightarrow{\text{ley}} \chi_r^2$, si H_0 es cierta.

Teorema 5.2.1 Si $f(x, \theta)$ es regular con respecto a sus primeras derivadas parciales en Θ_{r0} , entonces $H(\theta_0, \theta)$ tiene un máximo en $\theta = \theta_0$.

La matriz de segundas derivadas parciales de $[H(\theta_0, \theta_0) - H(\theta_0, \theta)]$ en $\theta = \theta_0$ es (B_{pq}) .

Teorema 5.2.2 Sea (X_1, \dots, X_n) una muestra con función de densidad $f(x, \theta_0)$, donde $f(x, \theta)$ es regular con respecto a las primeras θ derivadas en Θ_{r0} , entonces si $\hat{\theta}_n$ es el estimador de máxima verosimilitud, las sucesiones:

$$\frac{1}{n} \sum_{i=1}^n \log f(x_i, \theta_0), \quad \frac{1}{n} \sum_{i=1}^n \log f(x_i, \hat{\theta}_n),$$

convergen en probabilidad a $H(\theta_0, \theta_0)$, cuando $n \rightarrow \infty$.

Teorema 5.2.3 Si $\hat{\theta}$ es asintóticamente normal como en las condiciones del teorema 4.2.4, entonces:

$$\lim_{n \rightarrow \infty} P_{\theta_0}(-2 \log \Lambda < \chi^2) = \int_0^{\chi^2} \frac{e^{-\frac{1}{2}x} x^{\frac{r}{2}-1}}{2^{\frac{r}{2}} \Gamma(\frac{r}{2})} dx,$$

es decir si H_0 es cierta, $-2 \log \Lambda \xrightarrow{\text{ley}} \chi_r^2$.

Demostración Como $\hat{\theta} \xrightarrow{\text{c.s.}} \theta_0$, $\forall \epsilon > 0$, $\forall \delta > 0$, existen $n_{\epsilon, \delta}$ tal que:

$$P_{\theta_0}(E_n : \forall n \geq n_{\epsilon, \delta}) > 1 - \epsilon.$$

Sobre E_n tenemos que:

$$\sum_{i=1}^n \log f(X_i, \theta_0) = \sum_{i=1}^n \log f(X_i, \hat{\theta}) + \frac{1}{2}(\hat{\theta} - \theta_0)' B_{\theta^*} (\hat{\theta} - \theta_0), \quad \forall n \geq n_{\epsilon, \delta},$$

con θ^* variable aleatoria tal que $\|\theta_0 - \theta^*\| < \|\theta_0 - \hat{\theta}\|$ y $B_{\theta^*} = (B_{pq}(\theta_0, \theta^*))$, o sea:

$$-2 \sum_{i=1}^n \log \frac{f(X_i, \theta_0)}{f(X_i, \hat{\theta})} = -(\sqrt{n}(\hat{\theta} - \theta_0))' \left(\frac{1}{n} B_{\theta^*}\right) (\sqrt{n}(\hat{\theta} - \theta_0)), \quad \forall n \geq n_{\epsilon, \delta},$$

con probabilidad mayor de $1 - \epsilon$. Pero:

$$-\frac{1}{n} B_{\theta^*} \xrightarrow{\text{Pr}} B(\theta_0, \theta_0) = B = I_{\theta_0},$$

por lo que $(\sqrt{n}(\hat{\theta} - \theta_0))' B (\sqrt{n}(\hat{\theta} - \theta_0)) \xrightarrow{\text{ley}} \chi_r^2$, dado que $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{\text{ley}} N(0, B^{-1})$.

Teorema 5.2.4 Consistencia Si $f(x, \theta)$ es regular con respecto a sus segundas θ derivadas parciales, la prueba del cociente de verosimilitud es consistente para probar la hipótesis H_0 .

Demostración Por el teorema 5.2.3 se tiene $\lim_{n \rightarrow \infty} P_{\theta_0}(-2 \log \Lambda > \chi_\alpha^2) = \alpha$, donde el nivel de la prueba está dado por $P(\chi_r^2 > \chi_\alpha^2) = \alpha$ (nivel asintótico).

Sea $\theta_1 \neq \theta_0$, con $\theta_1 \in \Theta_{r0}$ y consideremos $\Lambda = \Lambda' \nu$, donde:

$$\Lambda' = \frac{\mathbb{L}(X_1, \dots, X_n, \theta_1)}{\mathbb{L}(X_1, \dots, X_n, \hat{\theta})}, \quad \nu = \frac{\mathbb{L}(X_1, \dots, X_n, \theta_0)}{\mathbb{L}(X_1, \dots, X_n, \theta_1)}.$$

Sea $u_n = -2 \log \Lambda'$, $nv_n = -2 \log \nu$. Si θ_1 es el verdadero valor de θ , por el teorema 5.2.3, $u_n \xrightarrow{\text{ley}} \chi_r^2$, cuando $n \rightarrow \infty$. Además:

$$\nu_n = 2 \left[\frac{1}{n} \sum_{i=1}^n \log f(X_i, \theta_1) - \sum_{i=1}^n \log f(X_i, \theta_0) \right] \xrightarrow{\text{Pr}} 2[H(\theta_1, \theta_1) - H(\theta_1, \theta_0)] > 0,$$

si θ_1 es el verdadero valor de θ . Así tenemos que:

$$\lim_{n \rightarrow \infty} P_{\theta_1}(u_n + nv_n > \chi_\alpha^2) = 1,$$

por lo que definimos:

$$\Phi = \begin{cases} 1 & \text{si } u_n + nv_n > \chi_\alpha^2 \\ 0 & \text{si } u_n + nv_n < \chi_\alpha^2 \end{cases}$$

y la potencia $\beta_\Phi(\boldsymbol{\theta}) \rightarrow \zeta(\boldsymbol{\theta})$, cuando $n \rightarrow \infty$.

Una extensión de los teoremas 5.1.5 y 5.1.6 es posible y se deja de ejercicio.

5.3 Prueba del cociente de verosimilitud de una hipótesis compuesta

En las pruebas de hipótesis de r parámetros, nos interesamos en un subespacio de $\Theta_r \subset \mathbb{R}^r$ de elementos de la forma $(\theta_1, \dots, \theta_s, \theta_{s+1,0}, \dots, \theta_{r,0})$, donde $\theta_{s+1,0}, \dots, \theta_{r,0}$ son fijos. Denotaremos la hipótesis compuesta por $\Theta_s \subset \Theta_r$. El cociente de verosimilitud es:

$$\Lambda_{r-s} = \frac{\sup_{\boldsymbol{\theta} \in \Theta_s} \mathbb{L}(X_1, \dots, X_n, \boldsymbol{\theta})}{\sup_{\boldsymbol{\theta} \in \Theta_r} \mathbb{L}(X_1, \dots, X_n, \boldsymbol{\theta})}$$

Para grandes muestras, nos interesamos en un abierto $\Theta_{r,0} \subset \Theta_r$ conteniendo a $\boldsymbol{\theta}_0$ el verdadero valor de $\boldsymbol{\theta}$.

Teorema 5.3.1 Sea X_1, \dots, X_n una muestra con función de densidad $f(x, \boldsymbol{\theta})$, con $\boldsymbol{\theta} \in \Theta_r \subset \mathbb{R}^r$, la cual es regular respecto a las segundas derivadas parciales para $\boldsymbol{\theta} \in \Theta_{r,0}$, entonces:

$$\lim_{n \rightarrow \infty} P(-2 \log \Lambda_{r-s} < \chi^2 / \boldsymbol{\theta} \in \Theta_s) = \int_0^{\chi^2} \frac{e^{-\frac{1}{2}x} x^{\frac{1}{2}(r-s)-1}}{2^{\frac{1}{2}(r-s)} \Gamma(\frac{1}{2}(r-s))} dx,$$

es decir si $\boldsymbol{\theta} \in \Theta_s$ es cierta, $-2 \log \Lambda_{r-s} \xrightarrow{\text{ley}} \chi_{r-s}^2$.

Demostración Sea $(\hat{\theta}_1, \dots, \hat{\theta}_r)$ el estimador de máxima verosimilitud como definido en el teorema 4.3.3 y sea $\hat{\theta}'_1, \dots, \hat{\theta}'_s$ la solución de $S_{p'n}(x_1, \dots, x_n, \boldsymbol{\theta}) = 0$, $p' = 1, \dots, s$, con respecto a $\theta_1, \dots, \theta_s$, cuando las otras componentes $\theta_{s+1}, \dots, \theta_r$ son fijas e iguales a $\theta_{s+1,0}, \dots, \theta_{r,0}$ respectivamente. Bajo las hipótesis del teorema 4.3.4 tenemos que $(\hat{\theta}'_1, \dots, \hat{\theta}'_s)$ se distribuye asintóticamente como una normal $N(\boldsymbol{\theta}_{p'}, (nB_{p'q'})^{-1})$, con $p', q' = 1, \dots, s$.

Si $\theta_0 \in \Theta_s$, denotemos $\zeta, \eta, \eta^1, A(\theta), A_1(\theta)$ con componentes:

$$\begin{aligned}\frac{1}{\sqrt{n}}S_{pn}(X_1, \dots, X_n, \theta_0) &= \zeta_{pn} \\ \sqrt{n}(\theta_{p0} - \hat{\theta}_p) &= \eta_{pn} \\ \sqrt{n}(\theta_{p'0} - \hat{\theta}'_{p'}) &= \eta^1_{p'n} \\ \frac{1}{n}S_{pqn}(X_1, \dots, X_n, \theta) &= A_{pq}(\theta) \\ \frac{1}{n}S_{p'q'n}(X_1, \dots, X_n, \theta_1, \dots, \theta_s, \theta_{s+1,0}, \dots, \theta_{r0}) &= A_{1pq}(\theta')\end{aligned}$$

donde $p, q = 1, \dots, r; p', q' = 1, \dots, s$.

Dada la regularidad de $f(x, \theta)$, $\forall \epsilon > 0$, existe n_ϵ tal que la probabilidad es mayor que $1 - \epsilon$ de que las siguientes ecuaciones se tengan, para todo $n \geq n_\epsilon$:

$$\begin{aligned}\zeta &= A(\theta^*)\eta \\ \zeta_1 &= A_1(\theta^{*'})\eta^1 \\ \sum_{i=1}^n \log f(X_i, \theta_0) &= \sum_{i=1}^n \log f(X_i, \hat{\theta}_1, \dots, \hat{\theta}_r) + \frac{1}{2}\eta' A(\theta_1^*)\eta \\ \sum_{i=1}^n \log f(X_i, \theta_0) &= \sum_{i=1}^n \log f(X_i, \hat{\theta}'_1, \dots, \hat{\theta}'_s, \theta_{s+1,0}, \dots, \theta_{r0}) + \frac{1}{2}\eta^{1'} A_1(\theta_1^{*'})\eta^1,\end{aligned}$$

con $\|\theta_1^* - \theta_0\| < \|\hat{\theta} - \theta_0\|$, $\|\theta_1^{*'} - \theta_0\| < \|\hat{\theta}' - \theta_0\|$, $\|\theta^* - \theta_0\| < \|\hat{\theta} - \theta_0\|$, $\|\theta^{*'} - \theta_0\| < \|\hat{\theta}' - \theta_0\|$,
 $\theta^*, \theta_1^* \in \Theta_{r0}$, $\theta^{*'}, \theta_1^{*'} \in \Theta_s \cap \Theta_r$, $\zeta_1 = (\zeta_{1n}, \dots, \zeta_{sn})$.

Los vectores pueden expresarse como:

$$\begin{aligned}\eta &= [A(\theta^*)]^{-1}\zeta \\ \eta^1 &= [A_1(\theta^{*'})]^{-1}\zeta_1.\end{aligned}$$

Además $-2 \log \Lambda_{r-s} = -2[\sum_{i=1}^n \log f(X_i, \hat{\theta}'_1, \dots, \hat{\theta}'_s, \theta_{s+1,0}, \dots, \theta_{r0}) - \sum_{i=1}^n \log f(X_i, \hat{\theta}_1, \dots, \hat{\theta}_r)]$, por lo que:

$$P_\theta(-2 \log \Lambda_{r-s} = \eta^{1'} A_1(\theta_1^{*'})\eta^1 - \eta' A(\theta_1^*)\eta/n \geq n_\epsilon) > 1 - \epsilon.$$

Si $\theta_0 \in \Theta_s$ es el verdadero valor de θ , $\zeta \xrightarrow{\text{ley}} N(0, (B_{pq}))$ i.e. las matrices $(A_{pq}(\theta^*))$ y $(A_{pq}(\theta_1^*))$ convergen en probabilidad a $(-B_{pq})$, cuando $n \rightarrow \infty$, dado que si $n \rightarrow \infty$, $\hat{\theta} \xrightarrow{\text{Pr}} \theta_0$, lo que implica que $\theta_1^*, \theta^* \xrightarrow{\text{Pr}} \theta_0$. Similarmente la variable $\zeta_1 \xrightarrow{\text{ley}} N(0, (B_{p'q'}))$ y las matrices $(A_{1p'q'}(\theta^{*'}))$ y $(A_{1p'q'}(\theta_1^{*'}))$ convergen en probabilidad a $(-B_{p'q'})$.

Así tenemos que:

$$\eta_1' A_1(\theta_1^*) \eta^1 - \eta' A(\theta_1^*) \eta =$$

$$\zeta_1' A_1(\theta_1^*)^{-1} A_1(\theta^*) A_1(\theta_1^*)^{-1} \zeta_1 - \zeta' A(\theta_1^*)^{-1} A(\theta^*) A(\theta_1^*)^{-1} \zeta \xrightarrow{\text{Ley}} Z,$$

con $Z = Y' B^{-1} Y - Y_1' B_{[s]}^{-1} Y_1$, donde $B_{[s]}^{-1} = (B_{p'q'})^{-1}$, el vector $Y \sim N(0, (B_{pq}))$ y $Y' = (Y_1', y_{s+1}, \dots, y_r)$.

Además $Z \sim \chi_{r-s}^2$.

En efecto:

$$B^* = B^{-1} - \begin{bmatrix} B_{[s]}^{-1} & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} B_{11.2}^{-1} - B_{11} & -B_{11.2}^{-1} B_{12} B_{22}^{-1} \\ -B_{22}^{-1} B_{21} B_{11.2}^{-1} & B_{22}^{-1} + B_{22}^{-1} B_{21} B_{11.2}^{-1} B_{12} B_{22}^{-1} \end{bmatrix}$$

$$\text{donde } B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}, \quad B_{[s]} = B_{11}, \quad B_{11}, \quad B_{11.2} = B_{11} - B_{12} B_{22}^{-1} B_{21}.$$

$$\text{Además, } B^* B = \begin{bmatrix} 0 & B_{11}^{-1} B_{12} \\ 0 & I_{r-s} \end{bmatrix} \text{ es idempotente y de rango } r - s.$$

Teorema 5.3.2 *Bajo las hipótesis del teorema 5.3.1 la prueba del cociente de verosimilitud es consistente.*

Demostración Sea $\theta_1 \neq \theta_0$, $\theta_1 \notin \Theta_s$, $\theta_1 = (\theta_{11}, \dots, \theta_{1s}, \theta_{1s+1}, \dots, \theta_{1r})$,

$$\Lambda_{r-s} = \frac{L(\cdot, \hat{\theta}'_1, \dots, \hat{\theta}'_s, \theta_{s+10}, \dots, \theta_{r0})}{L(\cdot, \hat{\theta}_1, \dots, \hat{\theta}_s, \theta_{s+1}, \dots, \hat{\theta}_{r0})} =$$

$$\frac{L(\cdot, \hat{\theta}''_1, \dots, \hat{\theta}''_s, \theta_{1s+1}, \dots, \theta_{1r})}{L(\cdot, \hat{\theta}_1, \dots, \hat{\theta}_s, \hat{\theta}_{s+1}, \dots, \hat{\theta}_r)} \cdot \frac{L(\cdot, \hat{\theta}'_1, \dots, \hat{\theta}'_s, \theta_{s+10}, \dots, \theta_{r0})}{L(\cdot, \hat{\theta}''_1, \dots, \hat{\theta}''_s, \theta_{1s+1}, \dots, \theta_{1r})}$$

donde $\hat{\theta}'$ es el estimador de máxima verosimilitud para $\theta \in \Theta_s$ y $\hat{\theta}''$ es el estimador de máxima verosimilitud para $\theta \notin \Theta_s$ i.e. $\theta \in \Theta'_s$, con $\Theta'_s = \{(t_1, \dots, t_s, \theta_{1s+1}, \dots, \theta_{1r}) \in \Theta_r\}$. Así tenemos que:

$$-2 \log \Lambda_{r-s} = -2 \log \Lambda_{r-s}^* + 2[\log L(\cdot, \hat{\theta}''_1, \dots, \hat{\theta}''_s, \theta_{1s+1}, \dots, \theta_{1r})$$

$$- \log L(\cdot, \hat{\theta}'_1, \dots, \hat{\theta}'_s, \theta_{s+10}, \dots, \theta_{r0})]$$

$$= u_n + nu_v.$$

Si θ_1 es el verdadero valor de θ :

$$u_n = -2 \log \Lambda_{r-s}^* \xrightarrow{\text{ley}} \chi_{r-s}^2,$$

$$v_n \xrightarrow{\text{Pr}} 2[H(\theta_1, \theta_1) - H(\theta_1, \theta'_1)] > 0,$$

pues $\hat{\theta}'' \xrightarrow{\text{Pr}} \theta_1$ y $\hat{\theta}' \xrightarrow{\text{Pr}} \theta'_1$, para algún $\theta'_1 \in \Theta_s$ i.e. si $\theta_1 \notin \Theta_s$:

$$\lim_{n \rightarrow \infty} P_{\theta_1}(u_n + nv_n > \chi_\alpha^2) = 1.$$

Utilizando funciones regulares de estimación se puede construir una clase de pruebas que son consistentes. Ninguna de éstas pruebas es más conveniente que la prueba del cociente de verosimilitud, cuando la noción de conveniencia asintótica se extiende a hipótesis compuestas.

5.4 Ejercicios

1. Una muestra (X_1, \dots, X_n) con función de densidad $\theta e^{-\theta x}$, $x > 0$, $\theta > 0$, se toma para probar la hipótesis $H_0: \theta = \theta_0$ contra $H_1: \theta = \theta_1$, $\theta_1 > \theta_0$. Determinar una prueba de nivel α y estudiar la potencia de la prueba.
2. Sea n_1 , \bar{X}_1 , S_1^2 el tamaño, la media y la varianza muestral de una población $N(\mu_1, \sigma^2)$ y n_2 , \bar{X}_2 , S_2^2 cantidades similares para una muestra de distribución $N(\mu_2, \sigma^2)$ independiente de la primera. Consideremos la hipótesis $H_0: \mu_1 = \mu_2$.

Probar que el cociente de verosimilitud para H_0 es equivalente a un t de Student:

$$t = \frac{(\bar{X}_1 - \bar{X}_2)}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},$$

donde $s^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$. Verificar que la prueba es insesgada.

3. Si el problema precedente se generaliza al caso k muestras, donde n_i , \bar{X}_i , S_i^2 , $i = 1, \dots, k$ son el tamaño, media y varianza muestrales de k muestras independientes de distribución $N(\mu_i, \sigma^2)$, $i = 1, \dots, k$ respectivamente y H_0 es la hipótesis $\mu_1 = \mu_2 = \dots = \mu_k$.

Probar que el cociente de verosimilitud es equivalente a un $F(k, n - k)$ de Fisher-Snedecor, donde $n = \sum_{i=1}^k n_i$. Determinar la estadística y probar que la prueba es insesgada.

4. Si S_1^2 y S_2^2 son varianzas muestrales en muestras de tamaño n_i de una población $N(\mu_i, \sigma_i^2)$, $i = 1, 2$ y si H_0 es la hipótesis $\sigma_1^2 = \sigma_2^2$, probar que el cociente de verosimilitud es equivalente al cociente de Fisher-Snedecor:

$$F = \frac{S_1^2}{S_2^2},$$

el cual tiene distribución $F(n_1 - 1, n_2 - 1)$, si H_0 es verdadera.

5. Supongamos que X_1, X_2 son variables aleatorias independientes de distribución binomial $B(n_i, p_i)$, $i = 1, 2$ respectivamente. Verificar que el cociente de verosimilitud para probar $H_0: p_1 = p_2$, está dado por:

$$\Lambda = \left(\frac{x}{n}\right)^x \left(1 - \frac{x}{n}\right)^{n-x} \left(\frac{x_1}{n_1}\right)^{-x_1} \left(\frac{x_2}{n_2}\right)^{-x_2} \left(1 - \frac{x_1}{n_1}\right)^{-n_1+x_1} \left(1 - \frac{x_2}{n_2}\right)^{-n_2+x_2},$$

donde $X = X_1 + X_2$ y $n = n_1 + n_2$. Probar que la distribución límite de $-2 \log \Lambda$, cuando $n_i \rightarrow \infty$, $i = 1, 2$, es una χ_1^2 si H_0 es cierta.

6. Generalizar el resultado anterior cuando $X_i \sim B(n_i, p_i)$, $i = 1, \dots, k$ y son independientes.
7. a) Supongamos que n_{ij} , $i = 1, \dots, r$, $j = 1, \dots, s$, son realizaciones de una variable aleatoria de distribución multinomial de parámetros $p_{ij} > 0$, donde $n = \sum_{i=1}^r \sum_{j=1}^s n_{ij}$, $n_{.j} = \sum_{i=1}^r n_{ij}$, $n_{i.} = \sum_{j=1}^s n_{ij}$ y $\sum_{i=1}^r \sum_{j=1}^s p_{ij} = 1$. Sea H_0 el espacio en el cual $p_{ij} = p_i q_j$, donde $\sum_{i=1}^r p_i = \sum_{j=1}^s q_j = 1$. Verificar que el cociente de verosimilitud para probar H_0 está dado por:

$$\Lambda = \frac{\prod_{i=1}^r \binom{n_{i.}}{n}^{n_{i.}} \prod_{j=1}^s \binom{n_{.j}}{n}^{n_{.j}}}{\prod_{j=1}^s \prod_{i=1}^r \binom{n_{ij}}{n}^{n_{ij}}}$$

y que la distribución límite de $-2 \log \Lambda$ es $\chi_{(r-1)(s-1)}^2$.

- b) Sea $g = \sum_{j=1}^s \sum_{i=1}^r \frac{n_{ij} - \frac{n_{i.} n_{.j}}{n}}{\frac{n_{i.} n_{.j}}{n}}$, probar que si H_0 es cierta, $-2 \log \Lambda$ y g convergen conjuntamente en ley a una distribución $\chi_{(s-1)(r-1)}^2$.

8. Sea n_{ijk} , $i = 1, \dots, r$, $j = 1, \dots, s$, $k = 1, \dots, t$, una variable aleatoria $(rst - 1)$ dimensional de distribución multinomial de parámetros $p_{ijk} > 0$, donde $n = \sum_{i=1}^r \sum_{j=1}^s \sum_{k=1}^t n_{ijk}$, $n_{..k} = \sum_{i=1}^r \sum_{j=1}^s n_{ijk}$, $n_{i.j.} = \sum_{k=1}^t n_{ijk}$ y $\sum_{i=1}^r \sum_{j=1}^s \sum_{k=1}^t p_{ijk} = 1$.

Sea H_0 la hipótesis $p_{ijk} = p_{ij} q_k$, con $\sum_{i=1}^r \sum_{j=1}^s p_{ij} = \sum_{k=1}^t q_k = 1$, verificar que el cociente de verosimilitud para probar H_0 , está dado por:

$$\Lambda = \frac{\prod_{j=1}^s \prod_{i=1}^r \binom{n_{ij.}}{n}^{n_{ij.}} \prod_{k=1}^t \binom{n_{..k}}{n}^{n_{..k}}}{\prod_{i=1}^r \prod_{j=1}^s \prod_{k=1}^t \binom{n_{ijk}}{n}^{n_{ijk}}}$$

y que el límite de $-2 \log \Lambda$, cuando $n \rightarrow \infty$, es un $\chi_{(rs-1)(t-1)}^2$, si H_0 es cierta.

9. Supongamos que $(X_{11}, \dots, X_{1n_1})$ y $(X_{21}, \dots, X_{2n_2})$ son muestras independientes de distribución de Poisson $P(\mu_1)$ y $P(\mu_2)$ respectivamente. Sea $H_0: \mu_1 = \mu_2$, probar que el cociente de verosimilitud está dada por:

$$\Lambda = \frac{\bar{X}^{n\bar{X}}}{\bar{X}_1^{n_1\bar{X}_1} \bar{X}_2^{n_2\bar{X}_2}},$$

donde \bar{X}_1 y \bar{X}_2 son las medias de las muestras y $\bar{X} = \frac{n_1\bar{X}_1 + n_2\bar{X}_2}{n_1 + n_2}$, $n = n_1 + n_2$.

Probar que $-2 \log \Lambda$, cuando $n_1, n_2 \rightarrow \infty$ es un χ_1^2 , si H_0 es cierta. Generalizar éste resultado para k muestras.

10. a) La variable aleatoria $(k-1)$ -dimensional con realización (n_1, \dots, n_k) , tiene distribución multinomial de parámetros p_1, \dots, p_k , donde $\sum_{i=1}^k p_i = 1$ y $\sum_{i=1}^k n_i = n$, demostrar que el cociente de verosimilitud Λ , para probar la hipótesis $H_0: p_1 = p_{10}, \dots, p_k = p_{k0}$ está dada por:

$$\Lambda = \left[\left(\frac{p_{10}}{\hat{p}_1} \right)^{\hat{p}_1} \dots \left(\frac{p_{k0}}{\hat{p}_k} \right)^{\hat{p}_k} \right]^n,$$

donde $\hat{p}_i = \frac{n_i}{n}$ y probar que $-2 \log \Lambda$, cuando $n \rightarrow \infty$, es la distribución de un χ_{k-1}^2 , si H_0 es verdadera.

- b) Probar que cuando $n \rightarrow \infty$, $-2 \log \Lambda$ y $\sum_{i=1}^k \frac{(n_i - np_{i0})^2}{np_{i0}}$ convergen juntos a una distribución χ_{k-1}^2 , cuando $n \rightarrow \infty$, si H_0 es verdadera.

11. Si S_1^2, \dots, S_k^2 son las varianzas de muestras independientes de tamaños n_1, \dots, n_k , de distribuciones $N(\mu_1, \sigma_1^2), \dots, N(\mu_k, \sigma_k^2)$ y si $H_0: \sigma_1^2 = \dots = \sigma_k^2$, verificar que el cociente de verosimilitud Λ , para probar H_0 , está dado por:

$$\Lambda = \left[\frac{(n_1 - 1)S_1^2}{n_1 S^2} \right]^{\frac{1}{2}n_1} \dots \left[\frac{(n_k - 1)S_k^2}{n_k S^2} \right]^{\frac{1}{2}n_k},$$

donde $S^2 = \frac{(n_1 - 1)S_1^2 + \dots + (n_k - 1)S_k^2}{n_1 + \dots + n_k}$.

12. Si S^2 es la varianza de una muestra de tamaño n de distribución $N(\mu, \sigma^2)$, demostrar que usando el cociente de verosimilitud para probar $H_0: \sigma^2 = \sigma_0^2$, contra la hipótesis $H_1: \sigma^2 < \sigma_0^2$, es equivalente a usar $\frac{(n-1)S^2}{\sigma_0^2} \sim \chi_{n-1}^2$, si H_0 es cierta.

13. Supongamos que Y_{11}, \dots, Y_{1n_1} son variables aleatorias independientes de distribución $N(\beta_{10} + \beta_{11}x_{1v}, \sigma^2)$, $v = 1, \dots, n_1$ y que Y_{21}, \dots, Y_{2n_2} , son variables aleatorias independientes de distribución $N(\beta_{20} + \beta_{21}x_{1v}, \sigma^2)$, $v = 1, \dots, n_2$. Sea $H_0: \beta_{11} = \beta_{21}$ (hipótesis de líneas de regresión paralelas), probar que el cociente de verosimilitud es equivalente a:

$$F = \frac{S_w - S_\Omega}{\frac{S_w}{(n_1 + n_2 - 4)}},$$

donde $F \sim F(1, n_1 + n_2 - 4)$, si H_0 es cierta y donde:

$$\begin{aligned} S_w &= \sum_{p=1}^2 \sum_{v=1}^{n_p} [(Y_{pv} - \bar{Y}_p) - \hat{\beta}_p(x_{pv} - \bar{x}_p)]^2 \\ \bar{Y}_p &= \frac{1}{n_p} \sum_{v=1}^{n_p} Y_{pv}, \quad \bar{x}_p = \frac{1}{n_p} \sum_{v=1}^{n_p} x_{pv}, \quad \hat{\beta}_p = \frac{a_p}{b_p}, \\ a_p &= \sum_{v=1}^{n_p} (Y_{pv} - \bar{Y}_p)(x_{pv} - \bar{x}_p), \quad b_p = \sum_{v=1}^{n_p} (x_{pv} - \bar{x}_p)^2, \quad p = 1, 2, \\ S_\Omega &= \sum_{p=1}^2 \sum_{v=1}^{n_p} [(Y_{pv} - \bar{Y}_p) - \hat{\beta}(x_{pv} - \bar{x}_p)]^2, \quad \hat{\beta} = \frac{a_1 + a_2}{b_1 + b_2}. \end{aligned}$$

14. Si X_{ij} , $i = 1, \dots, r$, $j = 1, \dots, s$, son variables aleatorias independientes de distribución $N(\mu + \mu_{\zeta} + \mu_{\eta}, \sigma^2)$, donde $\sum_{\zeta=1}^r \mu_{\zeta} = 0$, $\sum_{\eta=1}^s \mu_{\eta} = 0$. Sea Θ el espacio de todos los vectores de la forma $(\mu, \mu_1, \dots, \mu_r, \mu_1, \dots, \mu_s, \sigma^2)$ sujeta a las restricciones $\sum_{\zeta=1}^r \mu_{\zeta} = \sum_{\eta=1}^s \mu_{\eta} = 0$ y Θ_0 el subconjunto de Θ , con $\mu_1 = \dots = \mu_r = 0$. Demostrar que el cociente de verosimilitud es equivalente a:

$$F = \frac{\frac{S_{\cdot 0}}{r-1}}{\frac{S_{\cdot\cdot}}{(r-1)(s-1)}} \sim F(r-1, (r-1)(s-1)),$$

donde $S_{\cdot 0} = s \sum_{\zeta=1}^r (\bar{X}_{\zeta} - \bar{X})^2$, $S_{\cdot\cdot} = \sum_{\zeta=1}^r \sum_{\eta=1}^s (X_{\zeta\eta} - \bar{X}_{\zeta} - \bar{X}_{\eta} + \bar{X})^2$, $\bar{X} = \frac{1}{rs} \sum_{\zeta=1}^r \sum_{\eta=1}^s X_{\zeta\eta}$, $\bar{X}_{\zeta} = \frac{1}{s} \sum_{\eta=1}^s X_{\zeta\eta}$, $\bar{X}_{\eta} = \frac{1}{r} \sum_{\zeta=1}^r X_{\zeta\eta}$, si Θ_0 es verdadera.

Apéndice A

Resumen de Teoría de la media e integración

A.1 Propiedades y definiciones

Definición A.1.1 Sea $\mathfrak{A} \subset \mathfrak{P}(\Omega)$, se llama anillo booleano¹ (clan) de partes de Ω , si:

$$A, B \in \mathfrak{A} \implies A \cup B \in \mathfrak{A}, \quad A \setminus B \in \mathfrak{A}.$$

Proposición A.1.1 Sea \mathfrak{A} un anillo booleano, entonces $A, B \in \mathfrak{A} \implies A \cap B \in \mathfrak{A}$.

Si \mathfrak{A} es unitario, $A \in \mathfrak{A} \implies \bar{A} \in \mathfrak{A}$.

Definición A.1.2 Se llama σ -anillo booleano (σ -clan) un anillo booleano \mathfrak{A} tal que:

$$\forall (A_n)_{n \in \mathbb{N}}, A_n \in \mathfrak{A}, n \in \mathbb{N} \implies \bigcup_{n \in \mathbb{N}} A_n \in \mathfrak{A}.$$

Se llama tribu (σ -álgebra, cuerpo de Borel, σ -clan unitario) a todo σ -anillo booleano \mathfrak{A} que satisface $\Omega \in \mathfrak{A}$.

Proposición A.1.2 Todo σ -anillo booleano cumple que si $(\forall n \in \mathbb{N})(A_n \in \mathfrak{A})$, $\bigcap_{n \in \mathbb{N}} A_n \in \mathfrak{A}$.

¹**George Boole (1815-1864)** Nace el 2 de Noviembre de 1815 en Lincoln, Lincolnshire, Inglaterra. Muere el 8 de Diciembre de 1864 en Ballintemple, County Cork, Irlanda. Autodidacta, es el precursor del concepto de conjunto y del cálculo sobre conjuntos (clases de objetos), que vinculan la lógica matemática con el cálculo algebraico (The Mathematical Analysis of Logic, 1847). Boole solucionó así, uno de los problemas fundamentales de formalización del lenguaje y del razonamiento, que se planteaban los matemáticos desde Leibniz. La pertinencia de sus trabajos le permitieron la obtención (1849) de un Cátedra de Matemáticas en el Queen's College de Cork (Irlanda del Sur). Se le puede considerar como el creador de la lógica moderna. Boole recluyó la lógica a una álgebra simple. También trabajó en ecuaciones diferenciales, el cálculo de diferencias finitas y métodos generales en probabilidad.

El álgebra booleana que desarrolla en 1854, en su tratado "An investigation of the laws of thought" se utiliza hoy día en el afinamiento de máquinas automáticas (álgebra de circuitos).

Definición A.1.3 Se llama *semi-anillo booleano* al conjunto $\mathfrak{A} \subset \mathfrak{P}(\Omega)$ si:

$$A, B \in \mathfrak{A} \implies A \cap B \in \mathfrak{A},$$

$$A, B \in \mathfrak{A}, A \subset B \implies \exists A_1, \dots, A_m, A_i \in \mathfrak{A} \text{ tales que } A_i \cap A_j = \emptyset, \text{ si } i \neq j \text{ y } B \setminus A = \bigcup_{i=1}^m A_i.$$

Definición A.1.4 Se llama *espacio medible* al par (Ω, \mathfrak{B}) , donde \mathfrak{B} es un σ -anillo booleano de partes del conjunto Ω .

Cuando \mathfrak{B} es una tribu, el par (Ω, \mathfrak{B}) , se llama *espacio probabilizable*.

Si $\mathfrak{B}_{\mathbb{R}^n}$ es la tribu de Borel de \mathbb{R}^n , $(\mathbb{R}^n, \mathfrak{B}_{\mathbb{R}^n})$ es un espacio probabilizable.

Definición A.1.5 Sean $(\Omega_1, \mathfrak{B}_1)$ y $(\Omega_2, \mathfrak{B}_2)$ espacios medibles, una función f de Ω_1 en Ω_2 se dice *medible* o \mathfrak{B}_1 -medible $\iff \forall B \in \mathfrak{B}_2 \implies f^{-1}(B) \in \mathfrak{B}_1$.

Definición A.1.6 Sea (Ω, \mathfrak{B}) espacio probabilizable y f una función $f: \Omega \rightarrow \mathbb{R}$ (o en $\bar{\mathbb{R}}$), f es *medible* \iff la aplicación f de (Ω, \mathfrak{B}) en $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}})$ (o $(\bar{\mathbb{R}}, \mathfrak{B}_{\bar{\mathbb{R}}})$) es medible.

Proposición A.1.3 Sean Ω' un conjunto, (Ω, \mathfrak{B}) un espacio medible y f una función $f: \Omega' \rightarrow \Omega$:

- $f^{-1}(\mathfrak{B})$ es un σ -anillo booleano.
- Si \mathcal{E} engendra \mathfrak{B} , $f^{-1}(\mathcal{E})$ engendra $f^{-1}(\mathfrak{B})$, o sea $\langle f^{-1}(\mathcal{E}) \rangle = f^{-1}(\langle \mathcal{E} \rangle)$.

Teorema A.1.1 Sean $(\Omega_1, \mathfrak{B}_1)$ y $(\Omega_2, \mathfrak{B}_2)$ dos espacios medibles y sea $f: \Omega_1 \rightarrow \Omega_2$, si \mathcal{E} engendra a \mathfrak{B}_2 , entonces f es medible $\iff f^{-1}(\mathcal{E}) \subset \mathfrak{B}_1$.

Corolario A.1.1 Sea (Ω, \mathfrak{B}) espacio probabilizable y sea $f: \Omega \rightarrow \mathbb{R}$, entonces f es medible $\iff f^{-1}(] - \infty, t]) \in \mathfrak{B}$, (o bien $f^{-1}(] - \infty, t]) \in \mathfrak{B}$).

Corolario A.1.2 Sean Ω_1, Ω_2 espacios topológicos y $\mathfrak{B}_1, \mathfrak{B}_2$ las respectivas tribus de Borel. Si $f: \Omega_1 \rightarrow \Omega_2$ es continua, entonces f es medible.

Definición A.1.7 Una función $f: E \rightarrow F$ se dice *escalonada* si toma un número finito de valores $\{y_1, \dots, y_n\} \subset F$.

Sea $\mathfrak{A} \subset \mathfrak{P}(E)$ y si $\forall i = 1, \dots, n, f^{-1}(\{y_i\}) \in \mathfrak{A}$, f se llama *escalonada sobre* \mathfrak{A} .

– Una función real escalonada sobre \mathfrak{A} es de la forma $f = \sum_{i=1}^n y_i I_{A_i}$, con $A_i = f^{-1}(\{y_i\})$, $A_i \cap A_j = \emptyset$, $i \neq j$.

– La función f escalonada es \mathfrak{A} -medible $\iff f$ es escalonada sobre \mathfrak{A} .

– Sea \mathfrak{A} anillo booleano, toda función real de la forma $f = \sum_{i=1}^n y_i I_{A_i}$, con $(A_i)_{i=1, \dots, n}$ familia finita de elementos de \mathfrak{A} , es escalonada sobre \mathfrak{A} .

A.2 Propiedades de funciones medibles numéricas

Consideramos que f es una función de Ω en \mathbb{R} , denotamos por convención:

$$f^{-1}(B) = \{x \in \Omega / f(x) \in B\} = [f \in B]$$

$$[f > \alpha] = \{x \in \Omega / f(x) > \alpha\}.$$

Teorema A.2.1 Sea (Ω, \mathfrak{B}) un espacio probabilisable y sean f y g funciones medibles sobre \mathbb{R} (o $\bar{\mathbb{R}}$), entonces la función $x \mapsto (f(x), g(x))$ de Ω en $\mathbb{R} \times \mathbb{R}$ (o $\bar{\mathbb{R}} \times \bar{\mathbb{R}}$) es medible, con la tribu de Borel sobre $\mathbb{R} \times \mathbb{R}$ (o $\bar{\mathbb{R}} \times \bar{\mathbb{R}}$).

Teorema A.2.2 Sea (Ω, \mathfrak{B}) espacio probabilisable y f, g dos funciones medibles sobre \mathbb{R} (resp. $\bar{\mathbb{R}}$), entonces:

- 1) $f + g$ (resp. si $f + g$ está definida), $\sup(f, g)$, $\inf(f, g)$, $f \cdot g$ son medibles; αf es medible $\forall \alpha \in \mathbb{R}$.
(Por convención $\infty \cdot 0 = 0 \cdot \infty = 0$, $+\infty - \infty$ no tiene sentido en $\bar{\mathbb{R}}$).
- 2) f sobre $\bar{\mathbb{R}}$ es medible $\iff f^+ = \sup(f, 0)$ y $f^- = \sup(-f, 0)$ son medibles.

Teorema A.2.3

- 1) Sea $(f_n)_n$ una sucesión de funciones medibles, con valores en $\bar{\mathbb{R}}$, las funciones $\sup_n f_n$, $\inf_n f_n$, $\limsup_n f_n$, $\liminf_n f_n$ son medibles.
- 2) Si $(f_n)_n$ es una sucesión de funciones medibles en $\bar{\mathbb{R}}$ tales que $\lim_n f_n = f$, entonces f es medible.

Teorema A.2.4 Sea (Ω, \mathfrak{B}) un espacio probabilisable, entonces toda función medible f en $\bar{\mathbb{R}}^+$ es el límite simple de una sucesión creciente de funciones escalonadas, con valores en \mathbb{R}^+ sobre \mathfrak{B} .

Corolario A.2.1 Toda función numérica medible f , con valores en $\bar{\mathbb{R}}$, es límite simple de una sucesión de funciones escalonadas finitas.

A.3 Medidas positivas

Extenderemos las operaciones suma y producto en $\bar{\mathbb{R}}$ definiendo:

$$\pm\infty \cdot x = \pm\infty, \quad x > 0$$

$$\pm\infty \cdot 0 = 0$$

$$x \in \mathbb{R}, \quad x + \infty = +\infty, \quad x - \infty = -\infty \quad \text{y} \quad -\infty + \infty \text{ no tiene sentido.}$$

Definición A.3.1 Se llama espacio de medida al triplete $(\Omega, \mathfrak{B}, \mu)$ donde (Ω, \mathfrak{B}) es un espacio medible y μ es una medida sobre \mathfrak{B} .

Se llama espacio de probabilidad al triplete $(\Omega, \mathfrak{B}, P)$ donde (Ω, \mathfrak{B}) es un espacio probabilizable y P es una medida sobre \mathfrak{B} tal que $P(\Omega) = 1$.

Teorema A.3.1 Sea S un semi-anillo de Ω y $\mu: S \rightarrow \mathbb{R}^+$ con las propiedades:

i) Toda sucesión finita $(S_i)_{i=1, \dots, n}$, $S_i \in S$ dos a dos disjuntos es tal que:

$$\mu\left(\bigcup_{i=1}^n S_i\right) = \sum_{i=1}^n \mu(S_i) \quad \text{si} \quad \bigcup_{i=1}^n S_i \in S.$$

ii) Toda sucesión $(S_n)_{n \in \mathbb{N}}$, $S_n \in S$ dos a dos disjuntos tales que $\bigcup_{n \in \mathbb{N}} S_n \in S$ cumple:

$$\mu\left(\bigcup_{n \in \mathbb{N}} S_n\right) = \sum_{n \in \mathbb{N}} \mu(S_n).$$

iii) Existe $(S_n)_{n \in \mathbb{N}}$, $S_n \in S$ tal que $\Omega \subset \bigcup_{n \in \mathbb{N}} S_n$.

Sea $X \in \mathfrak{P}(\Omega)$, $\mathfrak{B}(X) = \{(S_i)_i \text{ finitas o infinitas} / S_i \in S, X \subset \bigcup_i S_i\} \neq \emptyset$ por iii).

Sea $\mu^*(X) = \inf\{\sum_i \mu(S_i) / (S_i)_i \in \mathfrak{B}(X)\}$, entonces:

1) $\mu^*: \mathfrak{P}(\Omega) \rightarrow \bar{\mathbb{R}}^+$ es tal que

$$X \mapsto \mu^*(X)$$

a) $\mu^*(\emptyset) = 0$

b) $A \subset B \implies \mu^*(A) \leq \mu^*(B)$

$$c) \forall (A_n)_{n \in \mathbb{N}}, A_n \in \mathfrak{F}(\Omega), \mu^* \left(\bigcup_{n=1}^{\infty} A_n \right) \leq \sum_{n=1}^{\infty} \mu^*(A_n).$$

2) El conjunto $\mathfrak{B} = \{A \subset \Omega / \mu^*(X \cap A) + \mu^*(X \setminus A) = \mu^*(X), \forall X \in \mathfrak{F}(\Omega)\}$ es una tribu y $S \subset \mathfrak{B}$.

3) La restricción de μ^* a \mathfrak{B} es una medida y coincide con μ sobre S (μ^* extiende a μ).

4) Si $X \in \mathfrak{F}(\Omega)$ es tal que $\mu^*(X) = 0 \implies X \in \mathfrak{B}$.

- La función $\mu^*: \mathfrak{F}(\Omega) \longrightarrow \bar{\mathbb{R}}^+$ se llama medida exterior.
- La tribu \mathfrak{B} está formada por los conjuntos μ^* medibles.
- Las conclusiones 2) y 3) se conocen como el Teorema de extensión de Carathéodory.
- Si se elimina la condición iii) se define:

$$\mu^*(X) = \begin{cases} \inf \left\{ \sum_i \mu(S_i) / (S_i) \in \mathfrak{B}(X) \right\} & \text{si } \mathfrak{B}(X) \neq \emptyset \\ +\infty & \text{si } \mathfrak{B}(X) = \emptyset \end{cases}$$

y el resultado sigue siendo válido.

Teorema A.3.2 Sea S semi-anillo, $\mu: S \longrightarrow \mathbb{R}^+$ con las propiedades 1), 2) y 3) del teorema A.3.1.

La función μ se prolonga de forma única sobre la tribu engendrada por S . Esta medida es σ -finita.

A.4 Integración

Sea (Ω, \mathfrak{B}) espacio probabilisable y sea $\mathcal{M}(\mathfrak{B}, \bar{\mathbb{R}}) = \{f \text{ medible} / f: (\Omega, \mathfrak{B}) \longrightarrow \bar{\mathbb{R}}\}$ (resp. $\bar{\mathbb{R}}^+$).

Teorema A.4.1 Sea (Ω, \mathfrak{B}) espacio probabilisable, μ medida positiva sobre \mathfrak{B} , entonces existe una única función $\Phi: \mathcal{M}(\mathfrak{B}, \bar{\mathbb{R}}^+) \longrightarrow \bar{\mathbb{R}}^+$ tal que:

1) Para todo $\alpha \in \mathbb{R}^+$, $f \in \mathcal{M}(\mathfrak{B}, \bar{\mathbb{R}}^+)$, $g \in \mathcal{M}(\mathfrak{B}, \bar{\mathbb{R}}^+)$:

$$\Phi(\alpha f) = \alpha \Phi(f)$$

$$\Phi(f + g) = \Phi(f) + \Phi(g)$$

$$f \leq g \implies \Phi(f) \leq \Phi(g).$$

- 2) Para toda sucesión $(f_n)_{n \in \mathbb{N}} \uparrow f_n \in \mathcal{M}(\mathfrak{B}, \bar{\mathbb{R}}^+)$ se tiene $\sup_n \Phi(f_n) = \Phi(\sup_n f_n)$ (propiedad de Beppo Levi²).
- 3) Para todo $B \in \mathfrak{B}$, $\Phi(I_B) = \mu(B)$.

Definición A.4.1 Sea Φ el funcional asociado de forma única a μ (Teorema A.4.1) y sea $f \in \mathcal{M}(\mathfrak{B}, \bar{\mathbb{R}}^+)$:

$$\Phi(f) = \int^* f(x) d\mu(x) = \int^* f(x) \mu(dx) = \int^* f d\mu.$$

Φ es llamado integral superior asociado a μ . Si $B \in \mathfrak{B}$, $\int_B^* f d\mu =: \int^* I_B f d\mu$.

²**Beppo Levi (1875-1961)** Nace el 14 de mayo de 1875 en Turín, Italia y muere el 28 de agosto de 1961 en Rosario, Argentina. Se graduó en matemática en la Universidad de Turín en 1896. Después de este período de 1896 a 1899 fue asistente de Luigi Berzolari y se dedicó a la enseñanza media hasta que en 1906, ganó el puesto de Geometría descriptiva y proyectiva de la Universidad de Cagliari. También enseñó Geometría Analítica y en 1910 fue llamado a Parma a cubrir el puesto de Análisis algebraico. En 1928 se va a Bolonia, donde ocupa el puesto de Teoría de las funciones y de Análisis algebraico, donde entonces lleva a cabo una actividad favoreciendo la Unión Matemático Italiana, de la que fue administrador de 1931 a 1938, tomando a su cargo la escritura del Boletín. En 1938 para escapar del fascismo de la dictadura de Mussolini, se exilia en Argentina, donde fue designado Director del recién formado Instituto Matemático de Rosario y donde fundó algunas revistas matemáticas (Publicaciones y el Mathematicae Notae). Su actividad científico se divide en cuatro períodos de su vida profesional. Antes del traslado a Parma, el espacio desde la Geometría algebraica (issolution of the singularità of superficial algebriche) a la Lógica (el axioma de elección), la teoría de integración y las ecuaciones en derivadas parciales (el teorema famoso de Beppo Levi en integración de la monotonía de las sucesiones de funciones) y los trabajos Sobre el principio de Dirichlet, en el que viene introduce un nuevo espacio que da origen al llamado espacio de Beppo Levi. Sus trabajos se refieren a la teoría de integración de Lebesgue y a la geometría algebraica (estudio de los puntos singulares de una superficie). Durante los 18 años que pasó en Parma, de las previas tradiciones de investigación juntó la Teoría de los números, la Ingeniería Eléctrica, la Teoría de la medida física y Física teórica. En Bolonia a pesar de la dificultad de las tareas didácticas, continuó con intensidad sus artículos de Lógica, Ecuaciones diferenciales, Análisis complejo, publica en la frontera entre el Análisis y la Física y ensancha la actividad divulgativa. En los años de exilio, en su vida científica en Argentina finalmente publicó algunos de sus trabajos: unos explícitamente para introducir a los colegas argentinos en la investigación (expuesto con observaciones y desarrollos recientes), otras son fruto del diálogo que estimuló y conservó con los lectores de las Revistas que había fundado.

Teorema A.4.2 Teorema de Fatou³ Sea $(f_n)_n \subset \mathcal{M}(\mathfrak{B}, \bar{\mathbb{R}}^+)$ entonces:

$$\liminf_n \int^* f_n d\mu \geq \int^* (\liminf_n f_n) d\mu.$$

Teorema A.4.3 Sea $f \in \mathcal{M}(\mathfrak{B}, \bar{\mathbb{R}}^+)$, entonces $\int^* f d\mu = 0 \iff f = 0 \mu\text{-c.p.d.}$

Corolario A.4.1 Para toda $f \in \mathcal{M}(\mathfrak{B}, \bar{\mathbb{R}}^+)$, si $\mu(A) = 0 \implies \int_A^* f d\mu = 0$.

³**Pierre Joseph Louis Fatou (1878-1929)** Nace el 28 de febrero de 1878 en Lorient, Francia y muere el 10 de agosto de 1929 en Pornichet, Francia. Pierre Fatou entró a l'École Normale Supérieure en París en 1898 para estudiar matemática. Se graduó en 1901 y decide obtener un puesto matemático en el observatorio de París. Una vez designado en el puesto de astronomía, Fatou continúa trabajando en su tesis, que sometió en 1906, sobre teoría de integración y teoría de funciones complejas. Fatou probó que si una función es Lebesgue integrable, entonces los límites radiales para las respectivas integrales de Poisson existen casi por doquier. Este resultado condujo a las generalizaciones de Privalov, de Plessner y de Marcel Riesz. Aunque no dio una solución completa, Fatou también hizo una contribución importante a encontrar una solución a la cuestión relacionada con la transformación conforme de regiones de Jordan, sobre el disco abierto que se pueden extender continuamente en la frontera. En 1907 Fatou recibió su doctorado para este importante trabajo. Trabajó en el campo de la dinámica analítica compleja.

En 1915, l'Académie des Sciences en París da este tópico para el Gran Prix de 1918. El premio sería concedido a un estudio de la iteración desde un punto de vista global. Parece que los matemáticos como Appell, Emile Picard y Koenigs habían propuesto la idea a Académie des Sciences, porque esperaban progresos en el concepto de familias normales de Montel. Fatou escribió una extensa memoria que utilizaba las ideas de familias normales de Montel, para desarrollar la teoría fundamental de la iteración en 1917. Aunque no sabemos si se proponía entrar en el Grand Prix, sí parece casi seguro que emprendió el trabajo con esto en mente.

Dado que el asunto había sido propuesto para el premio, no es sorprendente que otros matemático también trabajaran en el asunto y Julia también produjo de hecho, una extensa memoria donde desarrolló la teoría de una manera similar a Fatou. Los dos, sin embargo, eligieron diversas maneras de ir adelante. Durante la segunda mitad de 1917 Julia depositó sus resultados en l'Académie des Sciences en sobres sellados. Fatou, por otra parte, publicó un aviso de sus resultados en una nota en el diciembre 1917 por partes en Comptes Rendus. Llegó a ser más adelante evidente que habían descubierto resultados muy similares.

Julia escribió una carta a Comptes Rendus referente a la prioridad del trabajo y fue publicado el 31 de diciembre de 1917. Julia había pedido a l'Académie des Sciences para que examinara sus sobres sellados y Georges Humbert habían sido designado de realizar la tarea. El mismo el 31 de diciembre 1917 de parte de Comptes Rendus, Georges Humbert recibe una carta reportando los trabajos de Julia. Como resultado de estas cartas Fatou no entró en el Gran Prix y fue concedido a Julia. Fatou sin embargo no perdió del todo, pues aunque no había optado para el premio, l'Académie des Sciences le dio premio su destacado artículo sobre el tema. Su trabajo se conoce ahora como el teorema generalizado de Fatou-Julia.

Fatou recibió el título del "astrónomo" en 1928 y como astrónomo, también hizo contribuciones a ese campo. Usando los teoremas de existencia para las soluciones a las ecuaciones diferenciales, Fatou pudo probar resultados rigurosos relacionados con en órbitas planetarias que Gauss había sugerido, realizando solamente una discusión intuitiva. También estudió el movimiento de un planeta en un medio resistente, con la intención de explicar cómo las estrellas gemelas se formarían con la captura de una, que se mueve en la atmósfera de la otra. Debemos también mencionar su trabajo sobre las series de Taylor, donde examinó la convergencia y la extensión analítica de la serie. Quizás el resultado más famoso de Fatou es que una función armónica $u > 0$ en una bola, tiene un límite no tangencial casi por doquier en la frontera.

A.5 Medidas definidas por densidades

Proposición A.5.1 Sea (Ω, \mathfrak{B}) un espacio probabilizable, μ medida positiva sobre \mathfrak{B} . Para cada función $f \geq 0$, \mathfrak{B} -medible, la aplicación $\nu(B) = \int_B^* f d\mu$ es una medida positiva sobre \mathfrak{B} tal que, $\mu(B) = 0 \implies \nu(B) = 0$.

Teorema A.5.1 Radon–Nikodým Sea μ una medida positiva σ -finita sobre un espacio probabilizable (Ω, \mathfrak{B}) . Sea ν una medida positiva absolutamente continua con respecto a μ ($\nu \ll \mu$), entonces:

– Existe una clase única \tilde{f} de funciones numéricas \mathfrak{B} -medibles μ -equivalentes tales que $\forall f \in \tilde{f}$ se tiene que $\nu = f \cdot \mu$.

– $\{f \in \tilde{f}\}$ son μ -finitas c.p.d. $\iff \mu$ es σ -finita.

– $\{f : f \in \tilde{f}\}$ son μ -integrables $\iff \mu$ es acotada.

Teorema A.5.2 Lebesgue⁴–Nikodým Sea μ medida positiva σ -finita sobre (Ω, \mathfrak{B}) espacio probabilizable y sea ν una medida positiva sobre (Ω, \mathfrak{B}) , existe una única clase \tilde{f} de funciones numéricas \mathfrak{B} -medibles positivas μ -equivalentes y una medida ν' positiva única μ -singular tal que $\nu = f\mu + \nu'$, $\forall f \in \tilde{f}$.

A.6 Medida imagen

Proposición A.6.1 Sean (Ω, \mathfrak{A}) y (E, \mathfrak{B}) espacios probabilizables y $T : (\Omega, \mathfrak{A}) \longrightarrow (E, \mathfrak{B})$ medible.

Si μ es una medida sobre \mathfrak{A} , la aplicación $\nu : \mathfrak{B} \longrightarrow \bar{\mathbb{R}}^+$ es una medida.

$$B \longmapsto \mu(T^{-1}(B))$$

⁴**Henri Léon Lebesgue (1875-1941)** Nace en Francia en 1875. Fue alumno en la Normal Superior y tuvo a Emile Borel como profesor (a quien se deben los primeros trabajos en teoría de la medida). Después de algunos años en el colegio de Nancy, Lebesgue enseña en Rennes. En este período que se dará a conocer por su elegante teoría de la medida. Fue profesor en la Sorbonne, luego en el Collège de France se le elige a la Academia de Ciencias en 1922.

Por su teoría de las funciones medibles (1901) que se apoya en las tribus borelianas (nombre dado en honor Emile Borel), Lebesgue redireccionó profundamente y generalizó el cálculo integral. Su teoría de la integración (1902-1904) respondió a las necesidades de los físicos, permitiendo la investigación y la existencia de primitivas de funciones “irregulares” y amalgama las distintas teorías hasta aquí desarrolladas, apareciendo como casos particulares: Riemann (función escalera, función continua), Darboux (funciones acotadas) y Stieltjes (funciones de variación acotada).

Sobre un intervalo de integración I , Riemann subdivide I , en tanto que Lebesgue subdivide $f(I)$. Se puede decir que la integral de Lebesgue es la integral de Riemann, el conjunto de números reales \mathbb{R} es al conjunto de racionales \mathbb{Q} : el conjunto de funciones integrables en el sentido de Lebesgue se puede obtener, en alguna medida, como la completación de las funciones integrables en el sentido de Riemann. El problema del paso al límite bajo la integral, extremadamente difícil con la integral de Riemann, se vuelve casi simple con esta nueva visión pero la teoría subyacente es más compleja.

Definición A.6.1 La medida ν definida sobre \mathfrak{B} por $\nu(B) = \mu(T^{-1}(B))$ se llama medida imagen de μ por T y se denota $T(\mu)$.

Teorema A.6.1 Fubini⁵ Sean μ_1, μ_2 medibles σ -finitas sobre $(\Omega_1, \mathfrak{B}_1)$ y $(\Omega_2, \mathfrak{B}_2)$ respectivamente.

Sea f una función $\mu_1 \otimes \mu_2$ -integrable, entonces:

– μ_1 -c.p.d. x_1, f_{x_1} es μ_2 -integrable.

– μ_2 -c.p.d. x_2, f_{x_2} es μ_1 -integrable.

– la función $x_1 \mapsto \int f_{x_1} d\mu_2$ está definida μ_1 -c.p.d. y es μ_1 -integrable.

– la función $x_2 \mapsto \int f_{x_2} d\mu_1$ está definida μ_2 -c.p.d. y es μ_2 -integrable y se tiene:

$$\int f d(\mu_1 \otimes \mu_2) = \int d\mu_1(x_1) \left[\int f_{x_1}(x_2) d\mu_2(x_2) \right] = \int d\mu_2(x_2) \left[\int f_{x_2}(x_1) d\mu_1(x_1) \right].$$

⁵**Guido Fubini (1879-1943)** Nace el 19 de enero 1879 en Venecia, Italia. Muere el 6 de junio de 1943 en New York, U.S.A. En 1896 Fubini entró a la Escuela Normal Superior de Pisa. Fue influido por Dini y Bianchi a estudiar geometría. Presentó su tesis doctoral en paralelismo en espacios elípticos de Clifford, en 1900.

Enseñó en la Universidad de Catania, Sicilia, en el Politécnico y en la Universidad de Turín. Por diferencias con el gobierno de Mussolini, Fubini fue forzado a jubilarse de su puesto en Turín. Recibió una invitación del Instituto de Estudios Avanzados en Princeton en 1939 y emigra a los Estados Unidos. Por sus problemas de la salud, muere cinco años después del corazón. Los intereses de Fubini en matemáticas eran muy diversos. Además del área del análisis, trabajó en el cálculo de variaciones donde él estudió la integral de Weierstrass reduciéndola a la integral de Lebesgue y expresa la integral de superficie en términos de dos integraciones simples. Otro tema de análisis que estudió fue las ecuaciones integrales no lineales. Fubini trabajó también en la teoría de grupos, en grupos lineales y grupos de funciones de automorfismos. Sus intereses incluyeron los grupos continuos donde trató el problema de proveer de una métrica al grupo. En espacios no euclidiano, extendió resultados de Appell y Mittag-Leffler. Sus trabajos más importantes fueron en geometría diferencial proyectiva, donde usó el cálculo diferencial. El teorema de la igualdad de las integrales iteradas lleva el nombre de Fubini, quien lo probó con gran generalidad en 1907, aunque Cauchy y sus contemporáneos ya sabían que se cumplía la igualdad para funciones continuas.

A.7 Modos de convergencia. Definiciones

En este capítulo trataremos cuatro tipos de convergencia de variables aleatorias (funciones medibles): convergencia en ley, convergencia en probabilidad, convergencia en L_2 (media cuadrática) y convergencia casi segura o casi por doquier y sus relaciones entre sí.

Definición A.7.1 Sea $(X_n)_{n \in \mathbb{N}}$ una sucesión de variables aleatorias tales que para todo n , X_n es una variable aleatoria vectorial con valores en \mathbb{R}^k y admite una función de repartición F_n . Sea X una variable con valores en \mathbb{R}^k , con función de repartición F , entonces se dice que $(X_n)_n$ converge en ley a X si y sólo si $F_n(x) \rightarrow F(x)$, en todo punto x donde F es continua, y lo denotaremos por $X_n \xrightarrow{\text{ley}} X$.

Observación

- $X_n \xrightarrow{\text{ley}} X$ no quiere decir que las realizaciones de X_n en una experiencia, están próximas a X .
- Si X es normal reducida $N(0, 1)$, escribimos $X_n \xrightarrow{\text{ley}} N(0, 1)$.

Definición A.7.2 Sea $(X_n)_{n \in \mathbb{N}}$ una sucesión de variables aleatorias reales, se dice que $(X_n)_n$ converge en probabilidad a 0 si y sólo si $\forall \epsilon > 0, \forall \eta > 0$, existe $n_0 \in \mathbb{N}$ tal que para todo $n \geq n_0$: $P(|X_n| < \epsilon) > 1 - \eta$ (i.e. para todo $\epsilon > 0, P(|X_n| < \epsilon) \rightarrow 1$ cuando $n \rightarrow +\infty$) y escribiremos $X_n \xrightarrow{\text{Pr}} 0$ si $n \rightarrow +\infty$.

Corolario A.7.1 $X_n \xrightarrow{\text{Pr}} X \iff X_n - X \xrightarrow{\text{Pr}} 0$.

Definición A.7.3 Sea $(\Omega, \mathfrak{A}, P)$ un espacio de probabilidad, $(X_n)_n$ una sucesión de variables aleatorias reales y X una variable aleatoria real, las definiciones siguientes son equivalentes.

- i) X_n converge casi seguramente a X si y sólo si existe $\Omega_1 \in \mathfrak{A}$ tal que $P(\Omega_1) = 1$ y para todo $\omega \in \Omega_1$, $X_n(\omega) \rightarrow X(\omega)$.
- ii) X_n converge casi seguramente a X si y sólo si $\sup_{k \geq n} |X_k - X| \xrightarrow{\text{Pr}} 0$, cuando $n \rightarrow +\infty$.

Observación Se dice que $(X_n)_n$ converge simplemente a $X \iff X_n(\omega) \rightarrow X(\omega), \forall \omega \in \Omega$. Igualmente se puede definir la convergencia uniforme.

A.8 Relaciones entre los tipos de convergencia

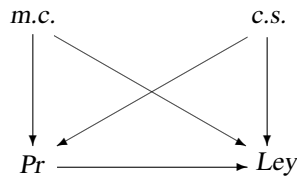
Teorema A.8.1 Si $(X_n)_n$ converge a X en media cuadrática, entonces $(X_n)_n$ converge en probabilidad.

Teorema A.8.2 Si $(X_n)_n$ converge casi seguro a X , entonces $(X_n)_n \xrightarrow{\text{Pr}} X$.

Teorema A.8.3 Si $(X_n)_n$ converge en probabilidad a X , entonces $(X_n)_n$ converge en ley a X .

Teorema A.8.4 Sea $(X_n)_n$ una sucesión de variables aleatorias que converge en ley a $X = a$ c.s., entonces $X_n \xrightarrow{\text{Pr}} a$.

El siguiente diagrama resume las relaciones de los teoremas vistos hasta ahora.



Nota La convergencia simple (i.e. $\forall \omega: X_n(\omega) \rightarrow X(\omega)$) implica la convergencia c.s.

A.9 Ley débil de grandes números

Teorema A.9.1 Teorema de Bernoulli⁶ Sea $(\Omega, \mathfrak{A}, P)$ un espacio de probabilidad y X_n una variable aleatoria real tal que $X_n \sim B(n, p)$, fijo $0 < p < 1$ (i.e. X_n es el número de realizaciones de un evento E en n pruebas independientes con $P(E) = p$). Si consideramos $Y_n = \frac{1}{n}X_n$, entonces $Y_n \xrightarrow{\text{Pr}} p$.

⁶**Jacob (Jacques) Bernoulli (1654-1705)** Nace en Basilea, Suiza el 27 de diciembre de 1654. Se graduó en Teología de Basilea en el año 1676. Recibió enseñanza en matemáticas y astronomía contra los deseos de sus padres. Entre el año 1676 y 1682 Jacob viajó a lo largo de Francia, Inglaterra y los países nórdicos. Se reunió con Boyle y Hooke en Inglaterra. Jacob retornó a Suiza y enseñó mecánica en la Universidad de Basilea desde el año 1683. Fue nombrado como profesor de matemáticas en Basilea en el año 1687 y es bien conocido por sus contribuciones a la teoría de la elasticidad y a la probabilidad matemática. El método para resolver la ecuación de Bernoulli fue descubierto por Leibniz en 1696. Los discípulos de Jacob Bernoulli incluyen a su sobrino Nicolás Bernoulli (1687-1759), quien contribuyó a la teoría de la probabilidad y a las series infinitas y a su hermano más joven Johann Bernoulli (1667-1745), quien tuvo una profunda influencia en el desarrollo del cálculo. Tuvo entre sus discípulos a Gabriel Cramer y a Leonard Euler. Su hijo Daniel Bernoulli (1700-1782) es conocido por su trabajo básico en el movimiento de fluidos y en la teoría cinética de los gases. Jacob fue el primero en usar el término integral en el año 1690. Utilizó tempranamente las coordenadas polares y descubrió el isócrono, la curva que se forma al caer verticalmente un cuerpo con velocidad uniforme. En una disputa matemática con su hermano Johann, inventó el cálculo de las variaciones. También trabajó en la Teoría de la Probabilidad. La distribución de Bernoulli, la ecuación diferencial de Bernoulli y los números de Bernoulli fueron llamados así por Jacob Bernoulli. Con su muerte su puesto en Basilea fue ocupado por su hermano Johann Bernoulli. Falleció el 16 de Agosto de 1705 en Basilea, Suiza.

Teorema A.9.2 Teorema de Bernoulli generalizado Sea $(\Omega, \mathfrak{A}, P)$ un espacio de probabilidad y $(X_n)_n$ sucesión de variables aleatorias reales independientes, $X_n \in L_2$, para todo n , tales que $E(X_n) = \mu$, $\forall n \in \mathbb{N}$ y $\text{var}(X_n) = \sigma^2$, $\forall n \in \mathbb{N}$.

Si definimos $\bar{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$, entonces $\bar{X}_n \xrightarrow{\text{Pr}} \mu$.

A.10 Convergencia en ley. Criterios

Teorema A.10.1 Paul Levy⁷ Sea $(X_n)_n$ sucesión de variables aleatorias con valores en \mathbb{R}^k y X una variable aleatoria con valores en \mathbb{R}^k . Sean $\Phi_n(t)$ y $\Phi(t)$ las funciones características de X_n y X respectivamente, entonces $X_n \xrightarrow{\text{ley}} X \iff \Phi_n(t) \rightarrow \Phi(t), \forall t \in \mathbb{R}^k$.

Teorema A.10.2 Si $(X_n)_n$ es una sucesión de variables aleatorias reales de densidad f_n con respecto a μ (σ -finita) y X una variable aleatoria real de densidad f respecto a μ y si existe g medible tal que $|f_n| \leq g$, para todo n , μ c.p.d. y si $f_n \rightarrow f$ μ -c.p.d., entonces $X_n \xrightarrow{\text{ley}} X$.

Ejemplos de convergencia en ley

1) a) Ley binomial

Si X_n es binomial $B(n, p)$, con p fijo, entonces se tiene que $\frac{1}{n}X_n \xrightarrow{\text{ley}} p$.

⁷Paul Pierre Lévy (1886-1972) Nace el 15 de setiembre de 1886 en Paris, Francia y muere el 15 de diciembre de 1971 en Paris, Francia. Paul proviene de una familia de matemáticos. Su abuelo era profesor de matemático y su padre, Lucien Lévy era un profesor de l'Ecole Polytechnique y escribió artículos en geometría. Paul asistió al Lycée San Louis de Paris, donde obtuvo premios no sólo en matemático, sino que también en griego, química y física. Obtuvo el primer promedio en la Ecole Normale Supérieure y el segundo en l'Ecole Polytechnique. Entró a l'Ecole Polytechnique y siendo estudiante publica sus primeros artículos sobre series semiconvergentes en 1905. Después de graduarse como el mejor, Lévy hizo un año de servicio militar y entra a l'Ecole des Mines en 1907. En tanto asistió a los cursos en la Sorbonne de Darboux y Emile Picard y atendió las lecciones en el Collège de France de Georges Humbert and Hadamard. Hadamard ejerció una gran influencia en los temas que Lévy investiga. Terminados los estudios en l'Ecole des Mines en 1910, inicia su investigación en análisis funcional. Su tesis en el tema fue examinada por Emile Picard, Poincaré y Hadamard en 1911 y recibió su Docteur ès Ciencias en 1912.

Lévy trabaja como profesor en l'Ecole des Mines en Paris en 1913, después profesor de Análisis en l'Ecole Polytechnique en Paris en 1920, donde se queda hasta que se retiró en 1959. Durante la I Guerra Mundial sirvió en la artillería y usó sus habilidades matemáticas en resolver problemas de defensa contra ataques aéreos. Un joven matemático R. Gateaux, fue muerto al inicio de la guerra y Hadamard solicita a Lévy prepararle los trabajos de Gateaux para publicarlos. Realizó la encomienda, pero no se detuvo ahí, sino que tomó las ideas de Gateaux y las amplió y desarrolló en una publicación en 1919. Se debe a Lévy el término Análisis funcional. Realizó importantes trabajos en análisis de Fourier, mecánica de las partículas: movimiento browniano y cálculo de probabilidades, análisis funcional, ecuaciones en derivadas parciales y series. En 1926 se extendió la transformada de Laplace. En 1963 fue elegido miembro honorario de la Sociedad Matemática de Londres. Al año siguiente se eligió a l'Académie des Sciences.

Si consideramos la variable aleatoria $Z_n = \frac{X_n - np}{\sqrt{npq}} = \frac{1}{n} X_n - p \sqrt{n}$, es reducida pues $E(Z_n) = 0$ y $\text{var}(Z_n) = 1$, entonces la función generatriz de momentos es:

$$m_{Z_n}(t) = E(e^{tZ_n}) = e^{-t\sqrt{\frac{np}{q}}} \left(q + pe^{\frac{t}{\sqrt{npq}}} \right)^n \rightarrow e^{-\frac{1}{2}t^2},$$

cuando $n \rightarrow \infty$, es decir $Z_n \xrightarrow{\text{ley}} N(0, 1)$.

b) Sea X_n una variable aleatoria binomial $B(n, p_n)$, donde $p_n = \frac{\lambda}{n}$ con λ fijo. Así $m_{X_n}(t) = \left(1 - \frac{\lambda}{n} + \frac{\lambda}{n} e^t \right)^n \rightarrow e^{-\lambda(1-e^t)}$, $\forall t \in \mathbb{R}$ y $e^{-\lambda(1-e^t)}$ es la función característica de una ley de Poisson⁸ de parámetro λ , es decir $X_n \xrightarrow{\text{ley}} P(\lambda)$.

2) Teorema central del límite (o teorema de límite centrado)

Teorema A.10.3 Sea $(X_n)_n$ una sucesión de variables aleatorias reales independientes que tienen la misma ley de probabilidad, tales que $X_n \in L_2$, $\forall n$ y $E(X_n) = \mu$ y $\text{var}(X_n) = \sigma^2$, entonces si $\bar{X}_n = \frac{1}{n}(X_1 + \dots + X_n)$ se tiene que $(\bar{X}_n - \mu) \sqrt{n} \xrightarrow{\text{ley}} N(0, \sigma^2)$, es decir que $(\bar{X}_n - \mu) \sqrt{n}/\sigma \xrightarrow{\text{ley}} N(0, 1)$.

A.11 Estudio de la convergencia casi segura (c.s.)

Proposición A.11.1 Si $X_n \xrightarrow{\text{c.s.}} X$ y $Y_n \xrightarrow{\text{c.s.}} Y$, entonces $X_n + Y_n \xrightarrow{\text{c.s.}} X + Y$.

Criterios de convergencia casi segura

Teorema A.11.1 Si la serie de término general $P(|X_n| > \epsilon)$ es convergente, para todo $\epsilon > 0$, entonces $X_n \xrightarrow{\text{c.s.}} 0$.

Teorema A.11.2 Si existe $p > 0$ tal que la serie de término general $E(|X_n|^p)$ sea convergente, entonces $X_n \xrightarrow{\text{c.s.}} 0$, si $n \rightarrow +\infty$.

⁸**Siméon Denis Poisson (1781-1840)** Nace en Francia en 1781 y estudia matemática, astronomía y física. Ocupó numerosos e importantes puestos de enseñanza: profesor en la Escuela Politécnica, profesor de Mecánica en la Facultad de Ciencias de París, Director de Enseñanza Matemática de los órganos colegiados de Francia. Fue Decano de la Facultad de Ciencias algunos meses antes de su muerte. Se lo conoce por la ley de probabilidad que lleva su nombre (Teoría del Cálculo de Probabilidades, 1838), pero sus trabajos principales lo realiza en electricidad, magnetismo, mecánica y movimientos vibratorios (teoría del calor, teoría de las ondas) donde, introduciendo numerosos conceptos matemáticos vinculados a las ecuaciones de Laplace (teoría del potencial electrostático, ecuaciones a las derivadas parciales), aparece después de Daniel Bernoulli y Fourier como el constructor de la física matemática moderna, es decir el estudio, por medio del análisis matemático, del comportamiento de los fenómenos, como consecuencia de las leyes dadas por la experimentación.

Siendo miembro de la Academia de las Ciencias, rechazó los trabajos del joven Evariste Galois en 1831.

La ley de Poisson es un caso límite de la ley binomial, en la cual interviene el tiempo.

Teoremas sobre las leyes fuertes de grandes números

Teorema A.11.3 Sea $(X_n)_n$ una sucesión de variables aleatorias reales independientes tales que

$$E(X_n) = \mu_n \text{ y } \text{var}(X_n) = \sigma_n^2. \text{ Si } \lim_{n \rightarrow \infty} \mu_n = \mu \text{ (finito) y } \sum_{n=1}^{\infty} \frac{\sigma_n^2}{n^2} \text{ converge, entonces } \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\text{c.s.}} \mu.$$

Teorema A.11.4 Kolmogorov⁹ Sea $(X_n)_n$ una sucesión de variables aleatorias reales independien-

⁹Andrei Nicolaiévitch Kolmogorov (1903-1987) Nace el 25 de abril de 1903 en Tambow, Rusia y muere en Moscú el 20 de octubre de 1987. Su vida es un ejemplo de consagración al trabajo científico. La extensión y variedad de los intereses de Kolmogorov es realmente sorprendente. Aunque siempre se consideró a sí mismo como un matemático puro, una parte esencial de su obra fue investigaciones en las ramas de las aplicaciones, tanto en las ciencias físicas como en la biología, la oceanología, la geología, la mineralogía, la balística y la lingüística. Desde muy joven se interesó por la historia en general y por la historia de la matemática en particular. Promovió la edición de diversas publicaciones de carácter divulgativo y una cuantiosa parte de su obra está dedicada a este empeño. El padre de Kolmogorov, Nikolai Matveevich Kataev, un calificado agrónomo y estadístico fue exiliado en Yaroslav. Después de la revolución socialista de octubre llegó a ser Director del Departamento de Educación en Narkomzem. Murió en 1919 en el frente sur durante la ofensiva de Denikin. Kolmogorov no llevó el apellido Katáev, pues su madre Mariya Yakovlievna Kolmogorov, nunca se casó y murió en el parto, cuando se dirigía a Crimea. Fue criado con su abuelo materno, Yákovliev Kolmogorov. Sus tías maternas, Vera y Nadieshda cuidaron de él y desarrollaron su curiosidad por la naturaleza y su interés en los libros. Trabajó como obrero ferroviario, conductor de trenes y bibliotecario en el vagón acondicionado para ello. En el verano de 1920, regresó a Moscú donde realiza estudios superiores. El estudio era totalmente gratuito y los estudiantes con dificultades económicas recibían un módico estipendio o una cuota mayor de alimentos. Kolmogorov matriculó matemática en la Universidad Estatal de Moscú Lomonosov e Ingeniería metalúrgica en el Instituto de Tecnología y Química Mendeleiev. Pasó varios meses entre Lomonosov y Mendeleiev, hasta que comprendió que sus aptitudes eran para la matemática. A partir de entonces toda su vida estuvo ligado a las matemáticas y a la Universidad Estatal de Moscú.

Fue alumno de Nikolai Luzin (1883-1950) quien formó la Luzitania y asistió al curso de topología que ofrecían Pavel Uryson (1898-1924) y Pavel Aleksandrov (1896-1982). También asistió al seminario sobre series trigonométricas del profesor Viacheslav Stepanov (1889-1950), donde los 19 años adquirió fama internacional al dar un ejemplo de una función integrable cuya serie de Fourier-Lebesgue era divergente casi por doquier, todo lo contrario de lo que pensaban la mayoría de los matemáticos de entonces. Tres años más tarde Kolmogorov volvió a asombrar a la comunidad científica internacional construyendo una función integrable (según Lebesgue) cuya representación en serie trigonométrica divergía en todo punto. Otros dos miembros de la Luzitania también influyeron significativamente en la formación de Kolmogorov. Sobre series trigonométricas, Dmitri Menshov (1892-1988) y sobre teoría de probabilidades, Aleksandr Khinchine (1894-1959). Ambos, con una fructífera experiencia de trabajo investigativo en la Luzitania, se convirtieron en una suerte de guías científicos, sobre todo en los períodos que Luzin viajaba al extranjero.

Kolmogorov terminó la carrera en 1925 y realizó estudios de postgrado durante el tiempo estipulado de cuatro años, inmediatamente entró a formar parte del Instituto de Investigaciones en Mecánica y Matemática de la Universidad Estatal de Moscú. En esa época en la URSS estaban abolidos los grados científicos y cuando se instauran nuevamente en 1935, le otorgaron a Andrei Nikolayevich Kolmogorov, el grado de Doctor en Ciencias Físicas y Matemáticas sin presentar ninguna tesis especial. Entonces ya tenía 59 publicaciones. A los 30 años, fue nombrado Director del Instituto de Investigaciones en Mecánica y Matemática de la U.E.M. y desde 1938 hasta su muerte mantuvo la Cátedra en el Departamento de Lógica Matemática. En el año 1939 fue Secretario científico de la sección de Física y Matemática de la Academia.

Un hecho trascendental en la vida de Kolmogorov ocurrió cuando en 1935, conjuntamente con su amigo Pavel Aleksandrov, adquirió la mitad de un palacete en Komarovka, a unos 40 kilómetros de Moscú, que convirtieron en casa de estudio.

Los primeros trabajos probabilísticos de Kolmogorov estuvieron enmarcados en la aplicación de los métodos de la teoría de funciones a los teoremas límites, en 1931 sobre la constitución de los procesos estocásticos como una rama de la teoría de

probabilidades y el artículo “Métodos analíticos en la teoría de probabilidades”, dedicado a estudiar los procesos de Markov. Publica en 1933 la célebre monografía “Fundamentos de la Teoría de Probabilidades”. En este libro presenta la construcción axiomática de la teoría de probabilidades que actualmente es ampliamente aceptada y hace uso del concepto de tribu (definidas por Borel) y las recientes teorías de la medida y el cálculo integral en el sentido de Lebesgue. Originalmente publicado en 1933 en alemán como “Grundbegriffe der Wahrscheinlichkeitsrechnung”. La edición más reciente “3a Edition” Rusa fue publicada en 1998 por Phasis, Moscow. En ese año demuestra rigurosamente la aproximación de la función de distribución empírica a la distribución de probabilidades de la variable aleatoria, siempre que la muestra que se tome sea suficientemente grande, dando una forma efectiva de estimar el error máximo que se comete en la aproximación. Este resultado de Kolmogorov inmediatamente entró a formar parte de todos los textos de estadística matemática. Sus trabajos se refieren también al análisis de Fourier y los sistemas dinámicos.

Kolmogorov estableció una estrecha colaboración con otros científicos no matemáticos, en particular con físicos y biólogos. Trabajó en el análisis de los problemas de las turbulencias desde un punto de vista estocástico y el estudio de los procesos ramificados. La importancia del primero fue tal que, en 1946, fue designado director del Laboratorio de Turbulencia en el Instituto de Geofísica Teórica de la Academia de Ciencias. En cuanto al segundo, abrió toda una rama dentro de la teoría de los procesos estocásticos y sus aplicaciones.

En los años 50 trabaja en la teoría de la información. Realizó junto a sus colaboradores, el análisis estadístico de los textos, tanto en prosa como en verso. Se interesó por uno de los problemas que Hilbert había propuesto en 1900 y que aún estaba sin solución: “Una función continua de n variables puede escribirse como una superposición de funciones de dos variables”. Introdujo la noción de entropía de un conjunto de funciones y demostró que cualquier función continua de n variables es representable en forma de superposición de funciones continuas de solo tres variables. Más tarde su estudiante Vladimir Arnold (1937–) completó la respuesta.

Los últimos 20 años de su fructífera vida, Kolmogorov los dedicó casi por completo a la obra pedagógica, pero el interés por la educación integral de la joven generación lo acompañó durante toda su vida profesional. Siendo ya profesor de la Universidad de Moscú, Kolmogorov participó con gran entusiasmo en la organización de la Primera Olimpiada Matemática. A partir del año 63 comienzan a organizarse las escuelas de verano y las escuelas internados. Las primeras se organizaban en uno de los meses de vacaciones, con los ganadores de las olimpiadas que se reunían con el fin de trabajar en matemática y descansar activamente. Kolmogorov personalmente participó en al menos 11 de estas escuelas de verano entre los años 63 y 77. Publicó más de 400 trabajos de divulgación científica, entre los que se destacan los 88 artículos que escribió para la segunda edición de la Gran Enciclopedia Soviética, los 72 entregados a la popular revista La Matemática en la Escuela (Matematika v Schkole) y los 44 de carácter periodístico aparecidos en los diarios y revistas de amplia difusión en la URSS.

Desde 1970 Andrei Kolmogorov junto al físico académico Isaac Kikoin, se encargó de la edición de la primera revista científica en el mundo que fuera escrita expresamente para estudiantes de secundaria, la revista Kvant (Cuanto). En 1988 totalizaban más de 400 los exalumnos de la escuela con el grado científico de doctor. Se contabilizan al menos 63 discípulos directos de Kolmogorov, además de los muchos que, sin serlo oficialmente, recibieron su influencia directa. De ellos 15 han sido o son miembros de la Academia de Ciencias.

Fue miembro de la Academia de Ciencias y Ciencias Pedagógicas de la Unión Soviética. Más aún, fue miembro de la Academia de Ciencias de EE.UU., del Instituto Francés, de la Royal Society en Londres. Fue miembro honorífico además de en las anteriores, de la Academia Internacional de Historia y Ciencias, y recibió premios nacionales e internacionales. Escribió muchos libros y más de 200 artículos sobre Teoría de Funciones, Lógica Matemática, Teoría de las Probabilidades y aplicaciones, Problemas de Estacionalidad, Educación e Historia de las matemáticas. Le fueron dadas las 7 órdenes Lenin, el Premio Stalin y el Premio Lenin que le concedió el Estado soviético. También fue electo miembro honorario de numerosas Academias de Ciencias y Sociedades Científicas de mayor prestigio del mundo. Por sus contribuciones a la Humanidad le fue otorgado el Premio Internacional de la Fundación Balzan en 1962, junto al Papa Juan XXIII y al compositor Paul Hindemith y el Premio Internacional de la Fundación Wolf en 1980 por profundos y originales descubrimientos en análisis de Fourier, teoría de probabilidades, teoría ergódica y sistemas dinámicos, compartido con H. Cartan. El premio Balzan recompensa todos los años a cuatro personajes importantes de las artes, de las ciencias o letras. Se instituyó en 1956 por Angela Lina Balzan en

tes de igual ley y media μ , entonces $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\text{c.s.}} \mu$.

Estos resultados se verán con más detalle cuando veamos el estudio del teorema del límite central.

A.12 Propiedades de la convergencia en ley y en probabilidad

Consideremos $(\Omega, \mathfrak{A}, P)$ un espacio de probabilidad y X una variable aleatoria real.

Proposición A.12.1 Sea $(X_n)_n$ una sucesión de variables aleatorias reales tales que $X_n \xrightarrow{\text{Pr}} a$, $a \in \mathbb{R}$, si ϕ es una función medible de $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}})$ en $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}})$, continua en a , entonces $\phi(X_n) \xrightarrow{\text{Pr}} \phi(a)$.

Proposición A.12.2 Si $X_n \rightarrow X$ (X variable aleatoria real finita) y si ϕ es una función medible de $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}})$ en $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}})$, continua en todo punto de \mathbb{R} , entonces $\phi(X_n) \xrightarrow{\text{Pr}} \phi(X)$.

Nota Si $X_n \xrightarrow{\text{Pr}} X$, (X finita), $Y_n \xrightarrow{\text{Pr}} Y$, (Y finita) y si ϕ es una función medible de $(\mathbb{R}^2, \mathfrak{B}_{\mathbb{R}^2})$ en $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}})$, ϕ continua en todo punto, entonces $\phi(X_n, Y_n) \xrightarrow{\text{Pr}} \phi(X, Y)$.

Aplicación $X_n + Y_n \xrightarrow{\text{Pr}} X + Y$.

Proposición A.12.3 Si $X_n \xrightarrow{\text{ley}} X$ y $Y_n \xrightarrow{\text{Pr}} a$, entonces $X_n + Y_n \xrightarrow{\text{Pr}} X + a$.

Nota

- Si $X_n \xrightarrow{\text{ley}} X$ y $Y_n \xrightarrow{\text{ley}} Y \not\Rightarrow X_n + Y_n \xrightarrow{\text{ley}} X + Y$.
- La expresión $X_n \xrightarrow{\text{ley}} X$ quiere decir que la ley de X_n se aproxima a la ley de X , pero como existen otras variables aleatorias con la misma ley que X , este límite no está muy bien determinado.

Propiedad A.12.1 Si $X_n \xrightarrow{\text{ley}} X$, con X variable aleatoria finita y $Y_n \xrightarrow{\text{Pr}} 0$, entonces $X_n Y_n \xrightarrow{\text{Pr}} 0$.

Proposición A.12.4 Si $X_n \xrightarrow{\text{ley}} X$, con X variable aleatoria finita y $Y_n \xrightarrow{\text{Pr}} a$, entonces $X_n Y_n \xrightarrow{\text{ley}} aX$.

memoria de su padre Eugene, director del Diario de Milán "Corriere della Serra". En 1985, 1986 y 1987 Nauka publica tres volúmenes de los trabajos de Kolmogorov (en ruso) con comentarios de él y de sus alumnos. En estos libros se tratan temas tan diversos como teoría de las series trigonométricas, teoría de la medida y conjuntos, teoría de la integración, teoría de la aproximación, construcciones lógicas, topología, teoría de la superposición de funciones, temas de mecánica clásica, teoría ergódica, teoría de la turbulencia, difusión y modelos en la dinámica de la población, artículos sobre los fundamentos de la teoría de la probabilidad, teoremas límites, teoría de procesos estocásticos, estadística matemática, teoría de los algoritmos, teoría de la información,...

Proposición A.12.5 Sea $(X_n)_n$ una sucesión de variables aleatorias reales tales que $E(X_n) = \mu_n \rightarrow \mu$ ($\mu \in \mathbb{R}$) y $\text{var}(X_n) = \sigma_n^2 \rightarrow 0$. Sea ϕ una aplicación medible, continuamente derivable en el punto μ . Si $\frac{X_n - \mu_n}{\sigma_n} \xrightarrow{\text{ley}} Z$ (variable aleatoria real finita), entonces $\frac{\phi(X_n) - \phi(\mu_n)}{\sigma_n} \xrightarrow{\text{ley}} \phi'(\mu)Z$.

Nota Si $\phi'(\mu) \neq 0$ se tiene $\frac{\phi(X_n) - \phi(\mu_n)}{\sigma_n \phi'(\mu)} \xrightarrow{\text{ley}} Z$.

Apéndice B

Anexos

B.1 Notas sobre las funciones eulerianas¹

B.1.1 La función Γ

Se llama *integral euleriana de segunda especie la integral*:

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt.$$

¹**Leonard Euler (1707-1783)** Nace el 15 de abril de 1707 en Basilea, Suiza. Muere el 18 de setiembre 1783 en San Petersburgo, Rusia. Euler, fue hijo de un clérigo, que vivía en los alrededores de Basilea. Su talento natural para las matemáticas se evidenció pronto por el afán y la facilidad con que dominaba los elementos, bajo la tutela de su padre. A una edad temprana fue enviado a la Universidad de Basilea, donde atrajo la atención de Jean Bernoulli. Inspirado por un maestro así, maduró rápidamente, a los 17 años de edad, cuando se graduó de Doctor, provocó grandes aplausos con un discurso probatorio, cuyo tema era una comparación entre los sistemas cartesiano y newtoniano. Fue el matemático más prolífico en la historia. Sus 866 libros y artículos representan aproximadamente una tercera parte de la investigación en las matemáticas, física teórica y la ingeniería mecánica entre 1726 y 1800.

En la matemática pura, integró el cálculo diferencial de Leibniz y el método de Newton de fluxiones dentro del análisis matemático; refinó la noción de función; hizo común muchas notaciones matemáticas, incluso e , i , el símbolo de π y el símbolo de sumatoria Σ . Introdujo los fundamentos de la teoría de funciones especiales, incluyendo las funciones trascendentales beta y gamma. En la física articuló la dinámica de Newton y los fundamentos de la mecánica analítica, en su libro *Teoría de los Movimientos de Cuerpos Rígidos* (1765).

Se instaló en San Petersburgo de Pedro I el Grande (en sustitución de Daniel Bernoulli) y luego en Berlín (1741), bajo el reino de Federico II, donde presidió la Academia de Ciencias hasta en 1766. Lagrange lo sucedió. Hacia el final de su vida, entonces ciego, volvió a San Petersburgo invitado por Catalina II. Euler produjo en los tres ámbitos fundamentales de la ciencia de su tiempo: la astronomía (órbitas planetarias, trayectorias de los cometas), las ciencias físicas (campos magnéticos, hidrodinámica, la óptica, naturaleza ondulante de la luz...) y matemáticas, en todas sus ramas, del aritmética a la geometría diferencial pasando por el análisis numérico y funcional, el cálculo de las variaciones, las curvas y las superficies algebraicas, el cálculo de probabilidades y los primeros aspectos de la teoría de los gráficos y de la topología. Sus hijos Jean Albert (1734-1800), Charles (1740-1790) y Christophe (1743-1812) fueron también matemáticos en San Petersburgo. Fundador de lo que se llama hoy el análisis funcional, publica numerosos tratados, precisa el concepto de función y adopta la notación $f(x)$, también utilizada por Clairaut, para designar la imagen por una función f de un número x , más adaptada que el de Jean Bernoulli que utilizaba la notación f_x . Extiende los trabajos de Bernoulli, precisa el concepto de función derivada, crea el concepto de ecuación en derivadas parciales (1734) y el cálculo de las variaciones (1744): búsqueda de máximos sobre las superficies, una de las ramas más fértiles del análisis. En 1770, Euler publica en alemán (el latín es cada vez más pasado en las publicaciones científicas) una Introducción completa al álgebra (*Vollständige Einleitung zur Algebra*), donde se puede considerar que se dice todo en cuanto a los números negativos y a su estatuto definitivo de verdadero número. De la misma forma, los números complejos (aún llamados imaginarios, el calificativo complejo es de Gauss en la forma $a + bi$), se definen y se estudian sus propiedades. Euler ya concibe el concepto de número trascendente (o incluso de función trascendente), que solo se lo pueden

Esta integral es función del parámetro x , converge en infinito y cero si $x > 0$.

Propiedades

- $\Gamma(x)$ es continua
- $\Gamma(x)$ es derivable y $\Gamma'(x) = \int_0^\infty \frac{\partial}{\partial x} t^{x-1} e^{-t} dt$
- $\Gamma(1) = 1, \Gamma(2) = 1, \Gamma(x+1) = x\Gamma(x)$
- $\Gamma(n+1) = n!,$ si $n \in \mathbb{N}^*$
- El cambio de variable $t = u^2$ transforma la función a $\Gamma(x) = 2 \int_0^\infty u^{2x-1} e^{-u^2} du$ y en particular $\Gamma(\frac{1}{2}) = 2 \int_0^\infty e^{-u^2} du = \sqrt{\pi}$
- Si $n \in \mathbb{N}$ la fórmula de recurrencia permite calcular:

$$\Gamma(x + \frac{1}{2}) = 2 \int_0^\infty u^{2x} e^{-u^2} du = \frac{1 \cdot 3 \cdots (2n-1)}{2^n} \sqrt{\pi} = \frac{(2n)!}{2^{2n} n!} \sqrt{\pi}$$

- $\Gamma(p)\Gamma(1-p) = \frac{\pi}{\operatorname{sen} p\pi}$
- $|\Gamma(ix)|^2 = \frac{\pi}{x \operatorname{sen} \pi x}$
- $\Gamma(2x) \sqrt{\pi} = 2^{2x-1} \Gamma(x) \Gamma(x + \frac{1}{2})$
- $\Gamma(x) \Gamma(x + \frac{1}{m}) \Gamma(x + \frac{2}{m}) \cdots \Gamma(x + \frac{m-1}{m}) = m^{\frac{1}{2}-mx} (2\pi)^{\frac{1}{2}(m-1)} \Gamma(mx)$

obtener por medio series convergentes (como π , e , $\ln 2$ y más generalmente e^x , $\operatorname{sen} x$, $\ln x$). Sin embargo, \mathbb{R} no ha sido construido, los conceptos de límite, diferencial y convergencia siguen siendo aún aproximados. El concepto de continuidad aún no se ha precisado, ni a fortiori, el concepto de continuidad uniforme que asegura, en el caso de una serie convergente de funciones, la suma. Estas preocupaciones aparecerán especialmente en el inicio del siglo XIX con Bolzano, Cauchy, Abel, luego Riemann y Weierstrass. El desarrollo del cálculo diferencial e integral responde a la voluntad de solucionar eficazmente los grandes problemas científicos, al alba del siglo de las luces de la tecnología (mecánica, hidráulica, producción de energía: vapor, electricidad). Se trata de incluir las leyes de la naturaleza, del movimiento (velocidad, aceleración, extremos). En su Tratado Introductio in Analysin infinitorum (1748), desarrolla una extensa síntesis de informaciones en cuanto a funciones trigonométricas, logaritmos y exponenciales (de base cualquiera). Establece entonces sus famosas fórmulas que muestran el vínculo entre la trigonometría, la exponencial y el análisis complejo $e^{ix} = \cos x + i \operatorname{sen} x$ y $e^{i\pi} + 1 = 0$, dónde en la segunda ecuación se encuentran los cinco grandes misterios del análisis real y complejo. Se tiene así:

$$\cos x = \frac{e^{ix} + e^{-ix}}{2}, \quad \operatorname{sen} x = \frac{e^{ix} - e^{-ix}}{2i}.$$

Se debe a Euler el primer método eficaz de resolución aproximada de una ecuación diferencial. El método consiste en discretizar la curva integral, buscando un número finito de puntos sobre un intervalo dado, linealizando (se reemplaza por sus tangentes). Runge y Kutta mejoraron este método a inicios del siglo XX.

- Otras definiciones de la función Γ :

$$\Gamma(x) = \lim_{k \rightarrow \infty} \frac{1 \cdot 2 \cdot 3 \cdots k}{(x+1)(x+2) \cdots (x+k)} k^x$$

$$\frac{1}{\Gamma(x)} = x e^{\gamma x} \prod_{m=1}^{\infty} \left(1 + \frac{x}{m}\right) e^{-x/m}, \quad \gamma \text{ es la constante de Euler}$$

- Derivada de la función $\Gamma(x)$:

$$\Gamma'(1) = \int_0^{\infty} e^{-x} \ln x dx = -\gamma$$

$$\frac{\Gamma'(x)}{\Gamma(x)} = -\gamma + \left(\frac{1}{1} - \frac{1}{x}\right) + \left(\frac{1}{2} - \frac{1}{x+1}\right) + \cdots + \left(\frac{1}{n} - \frac{1}{x+n-1}\right) + \cdots$$

- **Fórmula de Stirling** $\Gamma(x+1) = \sqrt{2\pi x} x^x e^{-x} \left(1 + \frac{1}{12x} + \frac{1}{288x^2} - \frac{139}{51840x^3} + \cdots\right)$

Si $x \gg 0$, $\Gamma(x+1) \approx \sqrt{2\pi x} x^x e^{-x}$ i.e. $n \gg 0$, $n! \approx \sqrt{2\pi n} n^n e^{-n}$.

B.1.2 La función β

La función euleriana β está definida por la integral:

$$\beta(p, q) = \int_0^1 t^{p-1} (1-t)^{q-1} dt.$$

La integral existe si $p > 0$ y $q > 0$.

Propiedades

- $\beta(p, q) = \beta(q, p)$. En efecto, basta efectuar el cambio de variable $u = 1 - t$
- Si $t = \sin^2 \theta$, $\beta(p, q) = 2 \int_0^{\frac{1}{2}\pi} \sin^{2p-1} \theta \cos^{2q-1} \theta d\theta$
- Si $t = \frac{z}{z+1}$, $\int_0^{\infty} z^{p-1} (1+z)^{p+q} dz$
- La función β se reduce a la función Γ :

$$\beta(p, q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}.$$

En efecto, consideremos el producto $\Gamma(p)\Gamma(q)$ como una integral doble, es decir:

$$I =_{\Omega} t^{p-1} e^{-t} u^{q-1} e^{-u} dt du,$$

donde el dominio Ω es el primer cuadrante del plano u, t . Consideremos el cambio de variable

$t + u = \alpha$, $\frac{t}{u} = \beta$, es decir $t = \frac{\alpha\beta}{1+\beta}$, $u = \frac{\alpha}{1+\beta}$ que va de Ω a Ω' , con Ω' el dominio $\alpha > 0, \beta > 0$.

El Jacobiano de la transformación viene dado por $J = \frac{\alpha}{(1+\beta)^2}$, por lo que:

$$I = \int_0^\infty \int_0^\infty \frac{\alpha^{p-1} \beta^{q-1}}{(1+\beta)^{p-1} (1+\beta)^{q-1}} e^{-\alpha} \frac{\alpha}{(1+\beta)^2} d\alpha d\beta = \int_0^\infty \frac{\beta^{p-1}}{(1+\beta)^{p+q}} d\beta \int_0^\infty \alpha^{p+q-1} e^{-\alpha} d\alpha,$$

es decir:

$$\frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)} = \int_0^\infty \frac{\beta^{p-1}}{(1+\beta)^{p+q}} d\beta.$$

Efectuando el cambio de variable $1+\beta = \frac{1}{1-t}$, o sea $\beta = \frac{t}{1-t}$ se tiene:

$$\int_0^\infty \frac{\beta^{p-1}}{(1+\beta)^{p+q}} d\beta = \beta(p, q)$$

- Si $0 < p < 1$ se tiene $\int_0^\infty \frac{\beta^{p-1}}{1+\beta} d\beta = \frac{\pi}{\operatorname{sen} p\pi}$
- $\beta(\frac{1}{2}, \frac{1}{2}) = \pi$.

B.2 Sobre las funciones de densidad

Toda función real f tal que $f(x) \geq 0$, para todo x , con $\int_{-\infty}^{+\infty} f(t)dt = 1$ es una función de densidad i.e. si $f(x) \geq 0$, para todo x , con $\int_{-\infty}^{+\infty} f(t)dt = C$, entonces $f' = f/C$ es una función de densidad.

- **Distribución normal** $\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = 1.$
- **Distribución gama** $\int_0^{+\infty} \frac{a^p}{\Gamma(p)} x^{p-1} e^{-ax} dx = 1.$
- **Distribución χ^2 con ν g.d.l.** $\int_0^{+\infty} \frac{1}{2^{\frac{1}{2}\nu} \Gamma(\frac{1}{2}\nu)} x^{\frac{1}{2}\nu-1} e^{-\frac{1}{2}x} dx = 1.$
- **Distribución de Student con ν g.d.l.** $\int_0^{+\infty} \frac{1}{\sqrt{\nu} \beta(\frac{1}{2}, \frac{1}{2}\nu)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{1}{2}(\nu+1)} dx = 1.$
- **Distribución beta** $\int_0^1 \frac{1}{\beta(p, q)} x^{p-1} (1-x)^{q-1} dx = 1.$
- **Distribución exponencial** $\int_0^{+\infty} \theta e^{-\theta x} dx = 1.$
- **Distribución de Cauchy²** $\int_{-\infty}^{+\infty} \frac{\pi}{1+x^2} dx = 1.$
- **Distribución de Fisher con n_1 y n_2 g.d.l.** $\int_0^{+\infty} \frac{n_1^{\frac{1}{2}n_1} n_2^{\frac{1}{2}n_2}}{\beta(\frac{1}{2}n_1, \frac{1}{2}n_2)} t^{\frac{1}{2}n_1} (n_2 + n_1 t)^{\frac{1}{2}(n_1+n_2)} dt = 1.$

²**Augustín Louis Cauchy (1789-1857)** Nace el 21 de Agosto 1789 en París, Francia. Muere el 23 de Mayo 1857 en Sceaux (cerca de París), Francia. Agustín Louis Cauchy fue pionero en el análisis. Investigó la convergencia y la divergencia de las series infinitas, las ecuaciones diferenciales, los determinantes, la probabilidad y física matemática.

Ocupó diversos puestos en la Facultad de Ciencia de París, el Colegio de Francia y la Escuela Politécnica. En 1814 publicó su trabajo de la integral definida que llegó a ser la base de la teoría de las funciones complejas.

Gracias a Cauchy, el análisis infinitesimal adquiere bases sólidas.

Cauchy, produjo 789 escritos, pero fue desaprobado por la mayoría de sus colegas. Mostró una obstinada rectitud a sí mismo y un agresivo fanatismo religioso. Como un apasionado del realismo pasó algún tiempo en Italia después de rechazar tomar un juramento de lealtad. Dejó París después de la Revolución de 1830. Aceptó una oferta del Rey de Piedmont para dar una cátedra en Turín donde estuvo hasta 1832. En 1833 se marchó a Praga en atención a Charles X y para ser el tutor de su hijo.

Cauchy volvió a París en 1838 y retoma su cargo en la academia pero no su posición de profesor por haber rechazado tomar el juramento de lealtad. Cuando Louis Philippe fue destronado en 1848 Cauchy vuelve a su cátedra en la Sorbonne. Ayudó en los posgrados hasta su muerte.

B.3 Matrices

Recordemos que una matriz $A_{n \times m}$ es un arreglo bidimensional de números reales, en el cual la entrada ij de la matriz, se denota a_{ij} , donde i corresponde a la fila $i \in \{1, \dots, n\}$ del arreglo y j corresponde a la columna $j \in \{1, \dots, m\}$ de ese mismo arreglo.

Se define la suma y el producto de matrices de tamaño conveniente:

$$A + B = C \iff c_{ij} = a_{ij} + b_{ij}.$$

$$A \cdot B = C \iff c_{ij} = \sum_{k=1}^n a_{ik}b_{kj}.$$

B.3.1 Teoremas

$$(A')' = A$$

$$\text{rang}(AB) \leq \inf\{\text{rang } A, \text{rang } B\}$$

$$(A^{-1})^{-1} = A$$

$$\text{rang}(A + B) \leq \text{rang } A + \text{rang } B$$

$$(A')^{-1} = (A^{-1})'$$

$$A_{n \times n} \text{ si } \det A = 0, \text{rang } A < n$$

$$(AB)' = B'A'$$

$$\text{Si } A'A = 0, \text{ entonces } A = 0$$

$$(AB)^{-1} = B^{-1}A^{-1}$$

$$A_{n \times n}, B_{n \times n}, \text{rang}(AB) \geq \text{rang } A + \text{rang } B - n$$

$$A \text{ no singular } \mathbf{y} = A\mathbf{x}, \text{ entonces } \mathbf{x} = A^{-1}\mathbf{y}$$

$$\text{rang}(A'A) = \text{rang}(AA') = \text{rang } A = \text{rang } A'$$

B.4 Formas cuadráticas

Sea $A_{n \times n}$ matriz simétrica, \mathbf{y} vector $n \times 1$, la forma cuadrática $\mathbf{y}'A\mathbf{y}$ se define por $\sum_{i=1}^n \sum_{j=1}^n a_{ij}y_i y_j$.

- El rango de la forma cuadrática $\mathbf{y}'A\mathbf{y}$ es el rango de A .
- La forma cuadrática se dice semidefinida positiva si $\mathbf{y}'A\mathbf{y} \geq 0$ y $\mathbf{y}'A\mathbf{y} = 0$, para algún $\mathbf{y} \neq 0$. En adelante consideramos sólo este tipo de formas cuadráticas.
- La matriz P se dice ortogonal si $P'P = I$ i.e. $P' = P^{-1}$.
- Si P es no singular, A definida positiva, $P'AP$ es definida positiva.
- A simétrica es definida positiva \iff existe P no singular tal que $A = PP'$.

$$- A \text{ es definida positiva } \iff a_{11} > 0, \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0, \dots, \begin{vmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{vmatrix} > 0.$$

- Si C es ortogonal $\mathbf{y} = Cz$, entonces $\mathbf{y}'\mathbf{y} = \mathbf{z}'\mathbf{z}$.
- Sea A matriz, si existe $\mathbf{x} \in \mathbb{R}^n$, $\lambda \in \mathbb{R}$ tal que $A\mathbf{x} = \lambda\mathbf{x}$, \mathbf{x} es un vector propio, λ es un valor propio.
- $|A - \lambda I| = 0$ es un polinomio de grado n en λ y se llama polinomio característico. Los ceros del polinomio son los valores propios.
- El número de valores propios $\neq 0$ es el rango de A .
- Si A es simétrica todos los valores propios son reales.
- Los valores propios de una matriz A definida positiva son positivos.
- Sea A una matriz simétrica, existe una matriz ortogonal P tal que $P'AP = D$ matriz diagonal de los valores propios de A .
- Sean A_1, \dots, A_k matrices simétricas, existe una matriz ortogonal C tal que $C'A_jC$ es diagonal $j = 1, \dots, k \iff A_iA_j$ es simétrica $i, j = 1, \dots, k$. Como las matrices A_i son simétricas entonces A_iA_j es simétrica $\iff A_i$ y A_j conmutan.

B.5 Determinantes

- Sean A, B matrices cuadradas del mismo orden, entonces $|AB| = |BA| = |B||A|$.
- Si A es singular, $|A| = 0$.
- Si C es ortogonal, $|C'AC| = |A|$ y $|C| = \pm 1$.
- Si C es triangular, $|C| = \prod_{i=1}^n c_{ii}$.
- Si A es no singular, $|A^{-1}| = |A|^{-1}$.
- Sea $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$, donde A_{11} y A_{22} son matrices cuadradas. Si la matriz A_{12} o $A_{21} = 0$, entonces $|A| = |A_{11}||A_{22}|$.
- Si las matrices A_1, A_2 son simétricas, A_2 es definida positiva, $A_1 - A_2$ es semi-definida positiva, entonces $|A_1| \geq |A_2|$.
- $\text{tr}(AB) = \text{tr}(BA)$, $\text{tr}(ABC) = \text{tr}(CAB) = \text{tr}(BCA)$.
- $\text{tr}(I) = n$, donde I es la matriz identidad $n \times n$.

- Si C es ortogonal $\text{tr}(C'AC) = \text{tr}(A)$.
- Sea A una matriz simétrica definida positiva tal que $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$ y sea B la matriz $\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$ inversa de A . Si A_{ii}, B_{ii} son de dimensión $m_i \times m_i$, entonces $A_{11}^{-1} = B_{11} - B_{12}B_{22}^{-1}B_{21}$.
- Sea A una matriz cuadrada tal que $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$ y si la sub-matriz A_{22} es no singular, $|A| = |A_{22}||A_{11} - A_{12}A_{22}^{-1}A_{21}|$.
- Sea la ecuación $A\mathbf{x} = \mathbf{y}$, una condición necesaria y suficiente para que la ecuación sea consistente es que el $\text{rang}(A)$ sea igual al de la matriz aumentada $B = (A, \mathbf{y})$.
- Si $\text{rang } A = \text{rang } B = m < n$, existe un único vector \mathbf{x} tal que $A\mathbf{x} = \mathbf{y}$.

B.6 Derivadas de matrices y vectores

- Sea \mathbf{a} un vector $p \times 1$, sea \mathbf{b} un vector $p \times 1$ y sea X matriz $p \times q$. Sea $Z = \mathbf{a}'X\mathbf{b} = \sum_{i=1}^q \sum_{j=1}^p a_i x_{ij} b_j$, entonces $\frac{\partial Z}{\partial X} = \mathbf{a}\mathbf{b}'$.
- Sea \mathbf{x} un vector $p \times 1$, A una matriz $p \times p$ simétrica y sea $Z = \mathbf{x}'A\mathbf{x}$, entonces $\frac{\partial Z}{\partial \mathbf{x}} = 2A\mathbf{x}$.

B.7 Matrices idempotentes

- Una matriz se dice idempotente si $AA = A^2 = A$
- Los ceros de una matriz simétrica idempotente son 0 o 1
- Sea A matriz simétrica idempotente de rango k , existe una matriz ortogonal tal que $P'AP = \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix}$.
- Toda matriz simétrica idempotente A de rango no completo, es semidefinida positiva
- Sea A matriz simétrica idempotente de rango k , $\text{tr}(A) = k$.
- Si A, B son simétricas idempotentes, entonces AB es idempotente si $AB = BA$.
- Si A es idempotente simétrica, P ortogonal entonces PAP' es idempotente.
- Si A es idempotente y $A + B = I$, entonces B es idempotente y $AB = BA = 0$.

- Si A_1, \dots, A_n son matrices $p \times p$ simétricas idempotentes, entonces existe una matriz P ortogonal tal que $P'A_iP$ es diagonal $\iff A_iA_j = A_jA_i, i = 1, \dots, n; j = 1, \dots, n$.

- Si A_1, \dots, A_m son matrices $p \times p$ simétricas, cualesquiera dos condiciones implica la tercera:

(1) A_1, \dots, A_m idempotentes

(2) $A_1 + \dots + A_m = B$ es idempotente

(3) $A_iA_j = 0, i \neq j$.

Si dos de las tres condiciones se dan, $\text{rang}(\sum_{i=1}^m A_i) = \sum_{i=1}^m \text{rang}(A_i)$.

B.8 Máximos, mínimos, Jacobianos³

Sea $\mathbf{y} = f(x_1, \dots, x_n)$ una función de n variables, si las derivadas parciales $\frac{\partial f}{\partial x_i}$ son continuas, entonces f puede tener un máximo o un mínimo en los puntos en donde:

$$\frac{\partial f}{\partial x_1} = \frac{\partial f}{\partial x_2} = \dots = \frac{\partial f}{\partial x_n} = 0.$$

- Si $f(x_1, \dots, x_n)$ es tal que las primeras y segundas derivadas parciales son continuas, entonces en los puntos donde $\frac{\partial f}{\partial x_i} = 0, i = 1, \dots, n$, la función tiene:

(1) un mínimo si la matriz $K = \left(\frac{\partial^2 f}{\partial x_i \partial x_j} \right)$ es definida positiva.

(2) un máximo si la matriz K es definida negativa.

- Consideremos $f(\mathbf{x}), g_j(\mathbf{x}), j = 1, \dots, p$, funciones reales de n variables ($n > p$), con derivadas parciales continuas en $a \in U$ abierto, $U \subset \mathbb{R}^n$, tales que $g_j(a) = 0, j = 1, \dots, p$ y tales que la matriz

³**Karl Gustav Jacob Jacobi (1804-1851)** Nace el 10 de diciembre de 1804 en Potsdam, Prusia (hoy Alemania). Muere el 18 de febrero 1851 en Berlín, Alemania. Jacobi estableció con Abel la Teoría de las funciones elípticas. Demostró la solución de integrales elípticas mediante la aplicación de las funciones, series exponenciales introducidas por él mismo. Desarrolló los determinantes funcionales, llamados después jacobianos y las ecuaciones diferenciales.

El padre de Jacobi era banquero y su familia era muy próspera, fue así como él recibió una buena educación en la Universidad de Berlín. Obtuvo su Doctorado en 1825 y enseñó matemáticas en Königsberg desde 1826 hasta su muerte. Fue nominado para una cátedra en 1832. En 1834 probó que si una función uni-valuada de una variable es doblemente periódica, entonces la razón de los periodos es imaginaria. Este resultado impulsó enormemente el trabajo en esta área, en particular a Liouville y Cauchy.

Jacobi tenía la reputación de ser un excelente maestro, atraía a muchos estudiantes. Introdujo un método de seminario para enseñar a los estudiantes los últimos avances matemáticos.

jacobianade los g_j sea de rango p en a , entonces si f tiene un extremo en a con las restricciones $g_j(\mathbf{x}) = 0, j = 1, \dots, p$, existen $\lambda_1, \dots, \lambda_p$ tales que:

$$\frac{\partial f}{\partial x_i}(a) + \sum_{j=1}^p \lambda_j \frac{\partial g_j}{\partial x_i}(a) = 0, \quad i = 1, \dots, n.$$

Los números λ_j se llaman *multiplicadores de Lagrange*⁴ y la función $L(\mathbf{x}) = f(\mathbf{x}) + \sum_{j=1}^p \lambda_j g_j(\mathbf{x})$ es llamada *función de Lagrange o lagrangiano*. Con esta notación escribimos $\frac{\partial L}{\partial x_i}(a) = 0, i = 1, \dots, n$ o $\nabla L(a) = 0$. Los puntos a que satisfacen esta condición, son *puntos críticos bajo restricciones*.

- Sean $f, g_j, j = 1, \dots, p$ aplicaciones de $U \subset \mathbb{R}^n$ abierto sobre \mathbb{R} , $a \in U$, $p < n$, de clase C^2 en una bola abierta $B_a \subset U$ tal que $g_j(a) = 0, j = 1, \dots, p$ y tales que las condiciones siguientes se verifican:

- i) $\text{rang} \frac{\partial(g_1, \dots, g_p)}{\partial(x_1, \dots, x_n)}(a) = p$,
- ii) existen $\lambda_1, \dots, \lambda_p$ tales que $\frac{\partial f}{\partial x_i}(a) + \sum_{j=1}^p \lambda_j \frac{\partial g_j}{\partial x_i}(a) = 0, i = 1, \dots, n$.

Sea $L(\mathbf{x}) = f(\mathbf{x}) + \sum_{j=1}^p \lambda_j g_j(\mathbf{x})$, entonces se tiene:

- a) Si f tiene un máximo bajo las restricciones $g_j(\mathbf{x}) = 0, j = 1, \dots, p$, se satisface:

$$\sum_{k=1}^n \sum_{r=1}^n \frac{\partial^2 L}{\partial x_k \partial x_r}(a) h_k h_r \leq 0, \quad \forall \mathbf{h} \in \mathbb{R}^n,$$

con la condición $\frac{\partial(g_1, \dots, g_p)}{\partial(x_1, \dots, x_n)}(a) \mathbf{h} = 0$.

- b) Si la condición siguiente se satisface:

$$\sum_{k=1}^n \sum_{r=1}^n \frac{\partial^2 L}{\partial x_k \partial x_r}(a) h_k h_r < 0, \quad \forall \mathbf{h} \in \mathbb{R}^n, \quad \mathbf{h} \neq 0$$

⁴**Joseph-Louis Lagrange (1736-1813)** Nace el 25 de Enero de 1736 en Turín, Sardinia-Piedmont (hoy Italia). Muere el 10 de Abril de 1813 en París, Francia. Lagrange, procedía de una ilustre familia parisiense, que tenía profundo arraigo en Cerdeña y algún rastro de noble linaje italiano. Pasó sus primeros años en Turín, su activa madurez en Berlín y sus últimos años en París, donde logró su mayor fama. En la escuela, sus intereses infantiles eran Homero y Virgilio y cuando una memoria de Halley le cayó en las manos, se alumbró la chispa matemática. A los dieciséis años de edad, fue nombrado profesor de matemáticas en la Escuela Real de Artillería de Turín. A los diecinueve años de edad, obtuvo fama resolviendo el así llamado problema isoperimétrico, que había desconcertado al mundo matemático durante medio siglo. En realidad Lagrange no sólo había resuelto un problema, también había inventado un nuevo método, el cálculo de variaciones, que sería el tema central de la obra de su vida. Este cálculo pertenece a la historia del mínimo esfuerzo, que comenzó en los espejos reflectores de Heron y continuó cuando Descartes reflexionó sobre la curiosa forma de sus lentes ovales. Lagrange podía demostrar que los postulados newtonianos de materia y movimiento, un tanto modificados, se adaptaban al amplio principio de economía de la naturaleza. El principio ha conducido a los resultados aún más fructíferos de Hamilton, Maxwell, en la obra de Einstein y en las últimas fases de la mecánica ondulatoria.

con la condición $\frac{\partial(g_1, \dots, g_p)}{\partial(x_1, \dots, x_n)}(a) \mathbf{h} = 0$, entonces a es un máximo local estricto de la función f con las restricciones $g_j(\mathbf{x}) = 0$, $j = 1, \dots, p$.

Estos resultados son válidos, con las desigualdades invertidas en el caso de un mínimo.

B.9 Combinación lineales de variables aleatorias

<i>Hipótesis</i>	<i>Ley de Y</i>	
$X_i \sim B(n_i, p), i = 1, \dots, k, \text{ independientes}$	$Y = \sum_{i=1}^k X_i$	$B\left(\sum_{i=1}^k n_i, p\right)$
$X_i \sim P(\lambda_i), i = 1, \dots, k, \text{ independientes}$	$Y = \sum_{i=1}^k X_i$	$P\left(\sum_{i=1}^k \lambda_i\right)$
$X_i \sim N(\mu_i, \sigma_i^2), i = 1, \dots, k, \text{ independientes}$	$Y = \sum_{i=1}^k a_i X_i$	$N\left(\sum_{i=1}^k a_i \mu_i, \sum_{i=1}^k a_i^2 \sigma_i^2\right)$
$X_i \sim \Gamma(p_i, a), i = 1, \dots, k, \text{ independientes}$ Caso particular $X_i \sim \chi_{v_i}^2$	$Y = \sum_{i=1}^k X_i$	$\Gamma\left(\sum_{i=1}^k p_i, a\right)$ $\chi_{\sum_{i=1}^k v_i}^2$
$X_i \sim C(v_i, \lambda), i = 1, \dots, k, \text{ independientes}$	$Y = \frac{1}{k} \sum_{i=1}^k X_i$	$C\left(\sum_{i=1}^k v_i, \lambda\right)$
<i>Cambio de variable</i>		<i>Ley de Z</i>
$X_i \sim U\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$	$Z = \text{tg } X$	$C(0, 1)$
$X, Y \text{ independientes} \sim N(0, 1)$	$Z = Y/X$	$C(0, 1)$
$X \sim N(0, 1)$	$Z = X^2$	χ_1^2
$X \sim N(v, \sigma^2)$	$Z = \log X$	<i>Ley log-normal</i>
$X \sim N(0, 1), Y \sim \chi_v^2 \text{ independientes}$	$Z = \frac{X}{\sqrt{Y/v}}$	<i>Student a v g.d.l.</i>
$X \sim \Gamma(p, a), Y \sim \Gamma(q, a) \text{ independientes}$	$Z = \frac{X}{X+Y}$	$\beta(p, q)$
$X_1 \sim \chi_{v_1}^2, X_2 \sim \chi_{v_2}^2 \text{ independientes}$	$Z = \frac{X_1/v_1}{X_2/v_2}$	$F(v_1, v_2)$
<i>T ley de Student v g.d.l.</i>	$Z = T^2$	$F(1, v)$

B.10 Características de las variables aleatorias conocidas

Denominación	Densidad o ley de probabilidad	Media	Varianza	Función generatriz
Binomial $B(n, p)$, $n \in \mathbb{N}^*$, $0 < p < 1$	$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$, $x = 0, 1, \dots, n$	np	$np(1-p)$	$(1-p + pe^t)^n$
Multinomial $M(n, p_1, \dots, p_k)$, $n \in \mathbb{N}^*$, $0 < p_i < 1$, $\sum_{i=1}^k p_i = 1$	$P(X_1 = x_1, \dots, X_k = x_k) = \frac{n!}{\prod_{i=1}^k x_i!} \prod_{i=1}^k p_i^{x_i}$, $\sum_{i=1}^k x_i = n$, $0 \leq x_i \leq n$	$E(X_i) = np_i$	$\text{var}(X_i) = np_i(1-p_i)$, $\text{cov}(X_i, X_j) = np_i(\delta_{ij} - p_j)$	$\left(\sum_{i=1}^k p_i e^{t_i} \right)^n$
Hipergeométrica $H(n, M, N)$	$P(X = x) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}}$, $p = \frac{M}{N}$, $x = 0, 1, \dots, n$	$E(X) = np$	$\text{var}(X) = \frac{N-n}{N-1} np(1-p)$	
Multihipergeométrica $H(n, N_1, \dots, N_k, N)$ $\sum_{i=1}^k N_i = N$	$P(X_1 = x_1, \dots, X_k = x_k) = \frac{\binom{N_1}{x_1} \dots \binom{N_k}{x_k}}{\binom{N}{n}}$, $\sum_{i=1}^k x_i = n$, $p_i = \frac{N_i}{N}$, $0 \leq x_i \leq n$	$E(X_i) = np_i$	$\text{var}(X_i) = \frac{N-n}{N-1} np_i(1-p_i)$, $\text{cov}(X_i, X_j) = -n \frac{N-n}{N-1} p_i p_j$	
Binomial negativa $n \in \mathbb{N}^*$, $0 < p < 1$ $n = 1$: ley geométrica	$P(X = x) = \binom{n-1+x}{n-1} p^n (1-p)^x$, $x = 0, 1, 2, \dots$	$\frac{n(1-p)}{p}$	$\frac{n(1-p)}{p^2}$	$\left(\frac{p}{1-(1-p)e^t} \right)^n$
Ley de Poisson de parámetro $\lambda > 0$, $P(\lambda)$	$P(X = x) = e^{-\lambda} \frac{\lambda^x}{x!}$	λ	λ	$e^{\lambda(e^t-1)}$
Series logarítmicas $L(p)$, $0 < p < 1$, $\alpha = \frac{-1}{\ln(1-p)}$	$P(X = x) = \frac{1}{\ln(1-p)} \frac{p^x}{x}$, $x = 1, 2, \dots$	$\frac{-\alpha p}{(1-p)}$	$\frac{-\alpha p(1+\alpha p)}{(1-p)^2}$	$\alpha \ln(1-pe^t)$
Uniforme sobre $[a, b]$, $U[a, b]$	$f(x) = \begin{cases} 0 & x \notin [a, b] \\ \frac{1}{b-a} & x \in [a, b] \end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{bt} - e^{at}}{t(b-a)}$
Normal $N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma > 0$	$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$	μ	σ^2	$e^{\mu t + \sigma^2 t^2 / 2}$

Denominación	Densidad o ley de probabilidad	Media	Varianza	Función generatriz
Ley gama $\Gamma(p, a)$, $a > 0$, $p > 0$	$f(x) = \begin{cases} 0 & x \leq 0 \\ \frac{a^p}{\Gamma(p)} x^{p-1} e^{-ax} & x > 0 \end{cases}$	$\frac{p}{a}$	$\frac{p}{a^2}$	$(1 - \frac{t}{a})^{-p}$, $t < a$
Ley χ^2 a ν grados de libertad	$f(x) = \begin{cases} 0 & x \leq 0 \\ \frac{e^{-x/2} x^{\nu/2-1}}{2^{\nu/2} \Gamma(\nu/2)} & x > 0 \end{cases}$	ν	2ν	$\frac{1}{(1-2t)^{\nu/2}}$
Ley beta $\beta(p, q)$, $p > 0$, $q > 0$	$f(x) = \begin{cases} 0 & x \notin [0, 1] \\ \frac{x^{p-1}(1-x)^{q-1}}{\beta(p, q)} & x \in [0, 1] \end{cases}$	$\frac{p}{p+q}$	$\frac{pq}{(p+q)^2(p+q+1)}$	
Student a ν g.d.l. $\nu > 0$ $\nu \in \mathbb{N}^*$ en la práctica	$f(x) = \frac{\left(1 + \frac{x^2}{\nu}\right)^{-\frac{1}{2}(\nu+1)}}{\sqrt{\nu} \beta\left(\frac{1}{2}, \frac{\nu}{2}\right)}$	0 , $\nu \geq 2$	$\frac{\nu}{\nu-2}$, $\nu \geq 3$	
Laplace $L(\mu, \sigma)$, $\mu > 0$, $\sigma > 0$	$f(x) = \frac{1}{\sqrt{2}\sigma} e^{-\frac{\sqrt{2} x-\mu }{\sigma}}$	μ	σ^2	$\frac{e^{\mu t}}{1 - \frac{1}{2}\sigma^2 t^2}$
Ley de Fisher de parámetros ν_1 y ν_2 $\nu_1 > 0$, $\nu_2 > 0$, $F(\nu_1, \nu_2)$	$f(x) = \begin{cases} 0 & x < 0 \\ \frac{\frac{1}{2} \nu_1 \frac{1}{2} \nu_2 x^{\frac{1}{2} \nu_1 - 1}}{\beta\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right) (1 + \frac{\nu_1 x}{\nu_2})^{\frac{\nu_1 + \nu_2}{2}}} & x \geq 0 \end{cases}$	$\frac{\nu_2}{\nu_2 - 2}$, $\nu_2 \geq 3$	$\frac{2\nu_2^2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 2)^2(\nu_2 - 4)}$, $\nu_2 \geq 5$	
Ley exponencial $\text{Exp}(\theta) = \Gamma(1, \theta)$ $0 < \theta < \infty$	$f(x) = \begin{cases} 0 & x \leq 0 \\ \theta e^{-\theta x} & x > 0 \end{cases}$	$\frac{1}{\theta}$	$\frac{1}{\theta^2}$	$\frac{\theta}{\theta - t}$
Ley de Cauchy $C(\mu, \lambda)$	$f(x) = \frac{\lambda}{\pi(\lambda^2 + (x - \mu)^2)}$			Función característica $\phi(t) = e^{i\mu t - t \lambda}$
Multinormal $N(\mu, \Sigma)$, $\mu \in \mathbb{R}^n$ Σ definida positiva	$f(\mathbf{x}) = \frac{e^{-\frac{1}{2}(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)}}{\sqrt{(2\pi)^n \det \Sigma}}$	μ	Σ	$\lambda = 1$, $\mu = 0$, $\phi(t) = e^{- t }$ $e^{\mu' t + \frac{1}{2} t' \Sigma t}$

B.11 Teorema de Rao-Cramer

Teorema B.11.1 Rao-Cramer

Si se supone la existencia y la regularidad de la matriz de información de Fisher $I_\theta = \|\text{cov}(\frac{\partial}{\partial \theta_i} \log L, \frac{\partial}{\partial \theta_j} \log L)\|$, $i, j = 1, \dots, r$, la existencia y regularidad de la matriz V_T^θ de covarianza de un estimador insesgado T de θ , así como la existencia de $E(T_i \frac{\partial \log L}{\partial \theta_j})$, $i, j = 1, \dots, r$, entonces:

$$\forall z \in \mathbb{R}^r, \quad z' I_\theta^{-1} z \leq z' V_T^\theta z.$$

Demostración La prueba del teorema supone, que se puede derivar bajo el signo de integral. Sea P una matriz ortogonal tal que $PI_\theta^{-1}P' = H$ es diagonal. Así:

$$P(V_T^\theta - I_\theta^{-1})P' = PV_T^\theta P' - PI_\theta^{-1}P' = B^* - H_\theta.$$

Sea A la matriz de covarianza de $(T_1, \dots, T_r, \frac{\partial \log L}{\partial \theta_1}, \dots, \frac{\partial \log L}{\partial \theta_r})$, entonces:

$$A = \begin{bmatrix} V_T^\theta & E \\ E & I \end{bmatrix},$$

con $E = I$. Tomemos las $k(\leq r)$ primeras componentes de la matriz PT y de la matriz PL , con $L' = \left(\frac{\partial \log L}{\partial \theta_1}, \dots, \frac{\partial \log L}{\partial \theta_r} \right)$, su matriz de covarianza es $A_1 = \begin{bmatrix} B_{kk}^* & E_{kk} \\ E_{kk} & H_{kk} \end{bmatrix}$ y $|A_1| \geq 0$. Como la matriz H es diagonal $H_{kk}^{-1} = (H^{-1})_{kk}$. Definimos la matriz auxiliar N por:

$$N = \begin{bmatrix} E_{kk} & -H_{kk}^{-1} \\ 0 & H_{kk}^{-1} \end{bmatrix}, \quad \text{i.e.} \quad |N| = |H_{kk}^{-1}| > 0.$$

Por otro lado:

$$NA_1 = \begin{bmatrix} B_{kk}^* - H_{kk}^{-1} & 0 \\ H_{kk}^{-1} & E_{kk} \end{bmatrix},$$

entonces $|NA_1| = |B_{kk}^* - H_{kk}^{-1}| \geq 0$, o sea $V_T^\theta - I_\theta^{-1}$ es semidefinida positiva.

B.12 Convergencia de un proceso aleatorio

Una clase importante de problemas en estadística matemática es la determinación del límite de funciones de distribución de ciertas funciones de n variables aleatorias, cuando $n \rightarrow \infty$, es decir si (X_1, \dots, X_n) es una variable aleatoria n -dimensional y $g_n(X_1, \dots, X_n)$ es función de (X_1, \dots, X_n) , la cual es ella misma una variable aleatoria (si g es medible), el problema es determinar el límite de la función de repartición de $g_n(X_1, \dots, X_n)$, cuando $n \rightarrow \infty$, o al menos algunas propiedades de la función de repartición, si existe. Así, se hace necesario tratar con variables aleatorias con un número infinito de componentes. Tales variables aleatorias son llamadas procesos estocásticos. Un proceso estocástico con componentes numerable, se puede referir a una sucesión de variables aleatorias.

Más precisamente, un proceso estocástico o proceso aleatorio es una familia de variables aleatorias $\{X_t: t \in A\}$, donde A es en general un conjunto de \mathbb{R} .

El espacio $(\mathbb{R}^\infty, \mathfrak{B}_{\mathbb{R}^\infty}, P)$, es un espacio de probabilidad, donde los elementos de \mathbb{R}^∞ son sucesiones de números reales, $\mathfrak{B}_{\mathbb{R}^\infty}$ es la tribu generada por los cilindros $\prod_{i \in \mathbb{N}} B_i$, $B_i \in \mathfrak{B}_{\mathbb{R}}$ y $B_i = \mathbb{R}$ salvo en un conjunto finito de índices y P es tal que $P(\prod_{i \in \mathbb{N}} B_i) = \prod_{\alpha \in I} P_\alpha(B_\alpha)$, con I finito.

Consideremos un proceso estocástico (X, X_1, X_2, \dots) tal que:

$$\lim_{n \rightarrow \infty} P(|X_n - X| < \epsilon) = 1,$$

se dice que (X, X_1, X_2, \dots) convergen en probabilidad a X y se escribe $X_n \xrightarrow{\text{Pr}} X$.

Propiedades

1. Sea el proceso estocástico (X, X_1, X_2, \dots) , tal que $X_n \xrightarrow{\text{Pr}} X$ y sea $g: \mathbb{R} \rightarrow \mathbb{R}$ continua, entonces

$$g(X_n) \xrightarrow{\text{Pr}} g(X).$$

$$\text{Si } X_n \xrightarrow{\text{Pr}} a, \text{ entonces } g(X_n) \xrightarrow{\text{Pr}} g(a).$$

Una manera más general de la propiedad, es considerar que g es continua en un intervalo I cerrado, con $P(I) = 1$.

2. Si el proceso estocástico bidimensional $((X, Y), (X_1, Y_1), (X_2, Y_2), \dots)$ es tal que se tiene $(X_n, Y_n) \xrightarrow{\text{Pr}} (X, Y)$ y si $f(x, y)$ es continua, entonces $g(X_n, Y_n) \xrightarrow{\text{Pr}} g(X, Y)$.

Este resultado se generaliza fácilmente a p dimensiones.

3. Sea $(X_1^{(1)}, \dots, X_1^{(k)}, Y_1^{(1)}, \dots, Y_1^{(k)}, X_2^{(1)}, \dots, Y_2^{(1)}, \dots, Y_2^{(k)}, \dots)$ un proceso estocástico multivariado tal que:

$$|Y_n^{(i)} - c^{(i)}| \leq |X_n^{(i)} - c^{(i)}|, \quad i = 1, \dots, k, \quad n = 1, 2, \dots$$

y $(X_n^{(1)}, \dots, X_n^{(k)}) \xrightarrow{\text{Pr}} (c^{(1)}, \dots, c^{(k)})$, entonces $(Y^{(1)}, \dots, Y^{(k)}) \xrightarrow{\text{Pr}} (c^{(1)}, \dots, c^{(k)})$.

Definición B.12.1 Sea $(g(X, \theta), g(X_1, \theta), g(X_2, \theta), \dots)$ un proceso estocástico, para $\theta \in \Theta_0$ intervalo abierto, si $\forall \epsilon > 0, \forall \theta_0 \in \Theta$ se tiene:

$$P(|g(X_n, \theta) - g(X, \theta)| > \epsilon) \rightarrow 0,$$

cuando $n \rightarrow \infty$, se dice que $g(X_n, \theta)$ converge en probabilidad a $g(X, \theta)$, uniformemente con respecto a $\theta \in \Theta_0$.

4. Si $g(X, \theta) = h(\theta)$ se dice que $g(X_n, \theta) \xrightarrow{\text{Pr}} h(\theta)$ uniformemente con respecto a θ .

Teorema B.12.1 Sea (X_1, X_2, \dots) un proceso estocástico dependiente de un parámetro $\theta \in \Theta_0$, sean $f_n(X_1, \dots, X_n, \theta), n = 1, 2, \dots$ y $\theta_n^*(X_1, \dots, X_n), n = 1, 2, \dots$ sucesiones de variables aleatorias convergiendo en probabilidad a $g(\theta)$ y θ respectivamente, uniformemente con respecto a $\theta \in \Theta_0$, donde $g(\theta)$ es continua en Θ_0 , entonces

$$f_n(X_1, \dots, X_n, \theta_n^*) \xrightarrow{\text{Pr}} g(\theta).$$

Demostración Sea (X_1, X_2, \dots) un proceso estocástico y sea $\theta_0 \in \Theta_0$, por hipótesis tenemos $\theta_n^*(X_1, \dots, X_n) \xrightarrow{\text{Pr}} \theta_0$, es decir $\forall \epsilon > 0$ y $\forall \eta > 0$, existe $n_{\epsilon\eta}$ tal que $\forall n \geq n_{\epsilon\eta}$:

$$P(E'_n) > 1 - \frac{\epsilon}{2},$$

donde $E'_n = \{w \in \mathbb{R}^\infty : |\theta_n^* - \theta_0| < \eta\}$. Además $\forall \epsilon > 0$ y $\forall \eta' > 0$, existe $n_{\epsilon\eta'}$ tal que:

$$\forall n \geq n_{\epsilon\eta'}, \quad P(E''_n) > 1 - \frac{\epsilon}{2},$$

donde $E''_n = \{w \in \mathbb{R}^\infty : |f_n(\cdot, \theta) - g(\theta)| < \frac{\eta'}{2}\}, \forall \theta \in \Theta_0$.

Dado $\eta' > 0$, existe $\delta > 0$ tal que si $|\theta - \theta_0| < \delta_{\eta'} \implies |g(\theta) - g(\theta_0)| < \frac{\eta'}{2}$; así tenemos que el conjunto

$E_n^* = \{\omega \in \mathbb{R}^\infty : |f_n(\cdot, \theta) - g(\theta_0)| < \eta'\}$ es tal que:

$$P(E_n^*) > 1 - \frac{\epsilon}{2}, \quad \text{si } |\theta - \theta_0| < \delta_{\eta'},$$

para todo $n \geq n_{\epsilon\eta'}$, pues $E_n'' \subset E_n^*$.

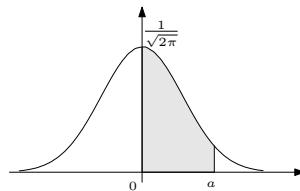
Así, $\forall \epsilon > 0, \forall \eta' > 0$, si $n \geq \max\{n_{\epsilon\eta'}, n_{\epsilon\delta_{\eta'}}\}$ tenemos que:

$$P(E_n' \cap E_n^*) = P(\{\omega \in \mathbb{R}^\infty : |f_n(\cdot, \theta_n^*) - g(\theta_0)| < \eta'\}) > 1 - \epsilon.$$

B.13 Tablas estadísticas

Tabla de la ley normal reducida $N(0, 1)$

Para cada valor de a , la tabla da el área de la región sombreada del gráfico $P(0 < x \leq a) = \frac{1}{\sqrt{2\pi}} \int_0^a e^{-\frac{1}{2}x^2} dx$, es decir la probabilidad de tener la desviación entre 0 y a .



Valor de a	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.000	0.004	0.008	0.012	0.016	0.020	0.024	0.028	0.032	0.036
0.1	0.040	0.044	0.048	0.052	0.056	0.060	0.064	0.068	0.071	0.075
0.2	0.079	0.083	0.087	0.091	0.095	0.099	0.103	0.106	0.110	0.114
0.3	0.118	0.122	0.126	0.129	0.133	0.137	0.141	0.144	0.148	0.152
0.4	0.155	0.159	0.163	0.166	0.170	0.174	0.177	0.181	0.184	0.188
0.5	0.192	0.195	0.199	0.202	0.205	0.209	0.212	0.216	0.219	0.222
0.6	0.226	0.229	0.232	0.236	0.239	0.242	0.245	0.249	0.252	0.255
0.7	0.258	0.261	0.264	0.267	0.270	0.273	0.276	0.279	0.282	0.285
0.8	0.288	0.291	0.294	0.297	0.300	0.302	0.305	0.308	0.311	0.313
0.9	0.316	0.319	0.321	0.324	0.326	0.329	0.332	0.334	0.337	0.339
1.0	0.341	0.344	0.346	0.349	0.351	0.353	0.355	0.358	0.360	0.362
1.1	0.364	0.367	0.369	0.371	0.373	0.375	0.377	0.379	0.381	0.383
1.2	0.385	0.387	0.389	0.391	0.393	0.394	0.396	0.398	0.400	0.402
1.3	0.403	0.405	0.407	0.408	0.410	0.412	0.413	0.415	0.416	0.418
1.4	0.419	0.421	0.422	0.424	0.425	0.427	0.428	0.429	0.431	0.432
1.5	0.433	0.435	0.436	0.437	0.438	0.439	0.441	0.442	0.443	0.444
1.6	0.445	0.446	0.447	0.448	0.450	0.451	0.452	0.453	0.454	0.455
1.7	0.455	0.456	0.457	0.458	0.459	0.460	0.461	0.462	0.463	0.463
1.8	0.464	0.465	0.466	0.466	0.467	0.468	0.469	0.469	0.470	0.471
1.9	0.471	0.472	0.473	0.473	0.474	0.474	0.475	0.476	0.476	0.477
2.0	0.477	0.478	0.478	0.479	0.479	0.480	0.480	0.481	0.481	0.482
2.1	0.482	0.483	0.483	0.483	0.484	0.484	0.485	0.485	0.485	0.486
2.2	0.486	0.486	0.487	0.487	0.488	0.488	0.488	0.488	0.488	0.489
2.3	0.489	0.490	0.490	0.490	0.490	0.491	0.491	0.491	0.491	0.492
2.4	0.492	0.492	0.492	0.493	0.493	0.493	0.493	0.493	0.493	0.494
2.5	0.494	0.494	0.494	0.494	0.495	0.495	0.495	0.495	0.495	0.495
2.6	0.495	0.496	0.496	0.496	0.496	0.496	0.496	0.496	0.496	0.496
2.7	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497	0.497
2.8	0.497	0.498	0.498	0.498	0.498	0.498	0.498	0.498	0.498	0.498
2.9	0.498	0.498	0.498	0.498	0.498	0.498	0.499	0.499	0.499	0.499
3.0	0.499	0.499	0.499	0.499	0.499	0.499	0.499	0.499	0.499	0.499
3.1	0.499	0.499	0.499	0.499	0.499	0.499	0.499	0.499	0.499	0.499

Ejemplo: para $a = 0.403$, $P(0 < x \leq a) = 0.166$.

Para leer la probabilidad para 0.43, se busca 0.4 en la primera columna y 0.03 en la quinta columna.

Ordenadas de la curva normal reducida

	0.0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.39894	0.39892	0.39886	0.39876	0.39862	0.39844	0.39822	0.39797	0.39767	0.39733
0.1	0.39695	0.39654	0.39608	0.39559	0.39505	0.39448	0.39387	0.39322	0.39253	0.39181
0.2	0.39104	0.39024	0.38940	0.38853	0.38762	0.38667	0.38568	0.38466	0.38361	0.38251
0.3	0.38139	0.38023	0.37903	0.37780	0.37654	0.37524	0.37391	0.37255	0.37115	0.36973
0.4	0.36827	0.35678	0.36526	0.36371	0.36213	0.36053	0.35889	0.35723	0.35553	0.35381
0.5	0.35207	0.35029	0.34849	0.34667	0.34482	0.34294	0.34105	0.33912	0.33718	0.33521
0.6	0.33322	0.33121	0.32918	0.32713	0.32506	0.32297	0.32086	0.3187	0.31659	0.31443
0.7	0.31225	0.31006	0.30785	0.30563	0.30339	0.30114	0.29887	0.29658	0.29430	0.29200
0.8	0.28969	0.28737	0.28504	0.28269	0.28034	0.27798	0.27562	0.27324	0.27086	0.26848
0.9	0.26609	0.26369	0.26129	0.25888	0.25647	0.25406	0.25164	0.24923	0.24681	0.24439
1.0	0.24197	0.23955	0.23713	0.23471	0.23230	0.22988	0.22747	0.22506	0.22265	0.22025
1.1	0.21785	0.21546	0.21307	0.21069	0.20831	0.20594	0.20357	0.20121	0.19886	0.19652
1.2	0.19419	0.19186	0.18954	0.18724	0.18494	0.18265	0.18037	0.17810	0.17585	0.17360
1.3	0.17137	0.16915	0.16694	0.16474	0.16256	0.16038	0.15822	0.15608	0.15395	0.15183
1.4	0.14973	0.14764	0.14556	0.1435	0.14146	0.13943	0.13742	0.13542	0.13344	0.13147
1.5	0.12952	0.12758	0.12566	0.12376	0.12188	0.12001	0.11816	0.11632	0.11450	0.11270
1.6	0.11092	0.10915	0.10741	0.10567	0.10396	0.10226	0.10059	0.09893	0.09728	0.09566
1.7	0.09405	0.09246	0.09089	0.08933	0.08780	0.08628	0.08478	0.08329	0.08183	0.08038
1.8	0.07895	0.07754	0.07614	0.07477	0.07341	0.07206	0.07074	0.06943	0.06814	0.06687
1.9	0.06562	0.06438	0.06316	0.06195	0.06077	0.05959	0.05844	0.05730	0.05618	0.05508
2.0	0.05399	0.05292	0.05186	0.05082	0.04980	0.04879	0.04780	0.04682	0.04586	0.04491
2.1	0.04398	0.04307	0.04217	0.04128	0.04041	0.03955	0.03871	0.03788	0.03706	0.03626
2.2	0.03547	0.03470	0.03394	0.03319	0.03246	0.03174	0.03103	0.03034	0.02965	0.02898
2.3	0.02833	0.02768	0.02705	0.02643	0.02582	0.02522	0.02463	0.02406	0.02349	0.02294
2.4	0.02239	0.02186	0.02134	0.02083	0.02033	0.01984	0.01936	0.01888	0.01842	0.01797
2.5	0.01753	0.01709	0.01667	0.01625	0.01585	0.01545	0.01506	0.01468	0.01431	0.01394
2.6	0.01358	0.01323	0.01289	0.01256	0.01223	0.01191	0.01160	0.01130	0.01100	0.01071
2.7	0.01042	0.01014	0.00987	0.00961	0.00935	0.00909	0.00885	0.00861	0.00837	0.00814
2.8	0.00792	0.00770	0.00748	0.00727	0.00707	0.00687	0.00668	0.00649	0.00631	0.00613
2.9	0.00595	0.00578	0.00562	0.00545	0.00530	0.00514	0.00499	0.00485	0.00470	0.00457
3.0	0.00443									
3.5	0.00087									
4.0	0.00013									
4.5	0.00002									
5.0	0.00000									

Para cada valor de u , la tabla da la ordenada $y = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}$ de la curva.

Ejemplo: para $u = 0.43$, $y = 0.36371$.

Distribución normal - Transformación de Probits

%	probits	%	probits	%	probits	%	probits
0.5	2.42	25.5	4.34	50.5	5.01	75.5	5.69
1.0	2.67	26.0	4.35	51.0	5.03	75.0	5.71
1.5	2.83	25.5	4.37	51.5	5.04	76.5	5.72
2.0	2.95	27.0	4.39	52.0	5.05	77.0	5.74
2.5	3.04	27.5	4.40	52.5	5.05	77.5	5.76
3.0	3.12	28.0	4.42	53.0	5.07	78.0	5.77
3.5	3.19	28.5	4.43	53.5	5.09	78.5	5.79
4.0	3.25	29.0	4.45	54.0	5.10	79.0	5.81
4.5	3.30	29.5	4.46	54.5	5.11	79.5	5.82
5.0	3.35	30.0	4.48	55.0	5.13	80.0	5.84
5.5	3.40	30.5	4.49	55.5	5.14	80.5	5.86
6.0	3.45	31.0	4.50	56.0	5.15	81.0	5.88
6.5	3.49	31.5	4.52	56.5	5.16	81.5	5.90
7.0	3.52	32.0	4.53	57.0	5.18	82.0	5.92
7.5	3.56	32.5	4.55	57.5	5.19	82.5	5.93
8.0	3.59	33.0	4.56	58.0	5.20	83.0	5.95
8.5	3.53	33.5	4.57	58.5	5.21	83.5	5.97
9.0	3.56	34.0	4.59	59.0	5.23	84.0	5.99
9.5	3.69	34.5	4.60	59.5	5.24	84.5	6.02
10.0	3.72	35.0	4.61	60.0	5.25	85.0	6.04
10.5	3.75	35.5	4.63	60.5	5.27	85.5	6.06
11.0	3.77	36.0	4.54	61.0	5.28	85.0	6.08
11.5	3.8	36.5	4.55	61.5	5.29	86.5	6.10
12.0	3.83	37.0	4.67	62.0	5.31	87.0	6.13
12.5	3.85	37.5	4.68	62.5	5.32	87.5	6.15
13.0	3.87	38.0	4.69	63.0	5.33	88.0	6.18
13.5	3.9	38.5	4.71	63.5	5.35	88.5	6.20
14.0	3.92	39.0	4.72	64.0	5.36	89.0	6.23
14.5	3.94	39.5	4.73	64.5	5.37	89.5	6.25
15.0	3.96	40.0	4.75	65.0	5.39	90.0	6.28
15.5	3.98	40.5	4.76	65.5	5.40	90.5	6.31
16.0	4.01	41.0	4.77	66.0	5.41	91.0	6.34
16.5	4.03	41.5	4.79	66.5	5.43	91.5	6.37
17.0	4.05	42.0	4.80	67.0	5.44	92.0	6.41
17.5	4.07	42.5	4.81	67.5	5.45	92.5	6.44
18.0	4.08	43.0	4.82	68.0	5.47	93.0	6.48
18.5	4.1	43.5	4.84	68.5	5.48	93.5	6.51
19.0	4.12	44.0	4.85	69.0	5.50	94.0	6.55
19.5	4.14	44.5	4.86	69.5	5.51	94.5	6.60
20.0	4.15	45.0	4.87	70.0	5.52	95.0	6.64
20.5	4.18	45.5	4.89	70.5	5.54	95.5	6.70
21.0	4.19	46.0	4.90	71.0	5.55	95.0	6.75
21.5	4.21	46.5	4.91	71.5	5.57	95.5	6.81
22.0	4.23	47.0	4.92	72.0	5.58	97.0	6.88
22.5	4.24	47.5	4.94	72.5	5.60	97.5	6.95
23.0	4.25	48.0	4.95	73.0	5.51	98.0	7.05
23.5	4.28	48.5	4.96	73.5	5.63	98.5	7.16
24.0	4.29	49.0	4.97	74.0	5.64	99.0	7.33
24.5	4.31	49.5	4.99	74.5	5.65	99.5	7.58
25.0	4.32	50.0	5.00	75.0	5.67		

Dada X una variable aleatoria normal reducida, si $P(x \leq a) = p$, se tiene $\text{Probit}(p) = a + 5$.

Observe que $\text{Probit}(p) + \text{Probit}(1 - p) = 10$.

Ejemplo: $\text{Probit}(72.5\%) = 5.60$, $\text{Probit}(27.5\%) = 4.40$.

Probits (para los valores de p de 0 a 50%)

	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0	...	1.9098	2.1218	2.2522	2.3479	2.4242	2.4879	2.5427	2.5911	2.6344
1	2.6737	2.7096	2.7429	2.7738	2.8027	2.8299	2.8256	2.8799	2.5911	2.9251
2	2.9463	2.9665	2.9859	3.0046	3.0226	3.0400	3.0569	3.0732	3.0890	3.1043
3	3.1192	3.1337	3.1478	3.1616	3.175	3.1881	3.2009	3.2134	3.2256	3.2376
4	3.2493	3.2608	3.2721	3.2831	3.294	3.3046	3.3151	3.3253	3.3354	3.3454
5	3.3551	3.3648	3.3742	3.3836	3.3928	3.4018	3.4107	3.4195	3.4282	3.4368
6	3.4452	3.4536	3.4618	3.4699	3.4780	3.4859	3.4937	3.5015	3.5091	3.5167
7	3.5242	3.5316	3.5389	3.5462	3.5534	3.5605	3.5675	3.5745	3.5813	3.5882
8	3.5949	3.6016	3.6083	3.6148	3.6213	3.6278	3.6342	3.6405	3.6468	3.6531
9	3.6592	3.6654	3.6715	3.6775	3.6835	3.6894	3.6953	3.7012	3.7070	3.7127
10	3.7184	3.7241	3.7298	3.7354	3.7409	3.7464	3.7519	3.7574	3.7628	3.7681
11	3.7735	3.7788	3.7840	3.7893	3.7945	3.7996	3.8048	3.8099	3.8150	3.8200
12	3.8250	3.8300	3.8350	3.8399	3.8448	3.8497	3.8545	3.8593	3.8641	3.8689
13	3.8736	3.8783	3.8830	3.8877	3.8923	3.8969	3.9015	3.9061	3.9107	3.9152
14	3.9197	3.9242	3.9286	3.9331	3.9375	3.9419	3.9463	3.9506	3.9550	3.9593
15	3.9636	3.9678	3.9721	3.9763	3.9806	3.9848	3.989	3.9931	3.9973	4.0014
16	4.0055	4.0096	4.0137	4.0178	4.0218	4.0259	4.0299	4.0339	4.0379	4.0419
17	4.0458	4.0498	4.0537	4.0576	4.0615	4.0654	4.0693	4.0731	4.077	4.0808
18	4.0846	4.0884	4.0922	4.096	4.0998	4.1035	4.1073	4.111	4.1147	4.1184
19	4.1221	4.1258	4.1295	4.1331	4.1367	4.1404	4.144	4.1476	4.1512	4.1548
20	4.1584	4.1619	4.1655	4.1690	4.1726	4.1761	4.1796	4.1831	4.1866	4.1901
21	4.1936	4.197	4.2005	4.2039	4.2074	4.2108	4.2142	4.2176	4.2210	4.2244
22	4.2278	4.2312	4.2345	4.2379	4.2412	4.2446	4.2479	4.2512	4.2546	4.2579
23	4.2612	4.2644	4.2677	4.2710	4.2743	4.2775	4.2808	4.2840	4.2872	4.2905
24	4.2937	4.2969	4.3001	4.3033	4.3065	4.3097	4.3129	4.3160	4.3192	4.3224
25	4.3255	4.3287	4.3318	4.3349	4.3380	4.3412	4.3443	4.3474	4.3505	4.3536
26	4.3567	4.3597	4.3628	4.3659	4.3689	4.3720	4.3750	4.3781	4.3811	4.3842
27	4.3872	4.3902	4.3932	4.3962	4.3992	4.4022	4.4052	4.4082	4.4112	4.4142
28	4.4172	4.4201	4.4231	4.4260	4.4290	4.4319	4.4349	4.4378	4.4408	4.4437
29	4.4466	4.4495	4.4524	4.4554	4.4583	4.4612	4.4641	4.4670	4.4698	4.4727
30	4.4756	4.4785	4.4813	4.4842	4.4871	4.4899	4.4928	4.4956	4.4985	4.5013
31	4.5041	4.5070	4.5098	4.5126	4.5155	4.5183	4.5211	4.5239	4.5267	4.5295
32	4.5323	4.5351	4.5379	4.5407	4.5435	4.5462	4.5490	4.5518	4.5546	4.5573
33	4.5601	4.5628	4.5656	4.5684	4.5711	4.5739	4.5766	4.5793	4.5821	4.5848
34	4.5875	4.5903	4.5930	4.5957	4.5984	4.6011	4.6039	4.6066	4.6093	4.6120
35	4.6147	4.6174	4.6201	4.6228	4.6255	4.6281	4.6308	4.6335	4.6362	4.6389
36	4.6415	4.6442	4.6469	4.6495	4.6522	4.6549	4.6574	4.6602	4.6628	4.6655
37	4.6681	4.6708	4.6734	4.6761	4.6787	4.6814	4.6840	4.6866	4.6893	4.6919
38	4.6945	4.6971	4.6998	4.7024	4.7050	4.7076	4.7102	4.7129	4.7155	4.7181
39	4.7207	4.7233	4.7259	4.7285	4.7311	4.7337	4.7363	4.7389	4.7415	4.7441
40	4.7467	4.7492	4.7518	4.7544	4.7570	4.7596	4.7822	4.7647	4.7673	4.7699
41	4.7725	4.7750	4.7776	4.7802	4.7827	4.7853	4.7879	4.7904	4.7930	4.7955
42	4.7981	4.8007	4.8032	4.8058	4.8083	4.8109	4.8134	4.8160	4.8185	4.8211
43	4.8236	4.8262	4.8287	4.8313	4.8338	4.8363	4.8389	4.8414	4.8440	4.8465
44	4.8490	4.8516	4.8541	4.8566	4.8592	4.8617	4.8642	4.8668	4.8693	4.8718
45	4.8743	4.8769	4.8794	4.8819	4.8844	4.8879	4.8895	4.8920	4.8945	4.8970
46	4.8996	4.9021	4.9046	4.9071	4.9096	4.9122	4.9147	4.9172	4.9197	4.9222
47	4.9247	4.9272	4.9298	4.9323	4.9348	4.9373	4.9398	4.9423	4.9448	4.9473
48	4.9498	4.9524	4.9549	4.9574	4.9599	4.9624	4.9649	4.9674	4.9699	4.9724
49	4.9749	4.9774	4.9799	4.9825	4.9850	4.9875	4.9900	4.9925	4.9950	4.9975

Distribución de Poisson

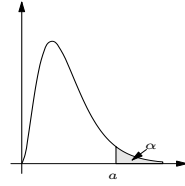
$$P_{\lambda}(k) = e^{-\lambda} \frac{\lambda^k}{k!}, \text{ para diferentes valores de } \lambda.$$

$k \ \lambda$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
0	0.9048	0.8187	0.7408	0.6703	0.6065	0.5488	0.4966	0.4493	0.4066	
1	0.0905	0.1634	0.2222	0.2681	0.3033	0.3293	0.3476	0.3595	0.3659	
2	0.0045	0.0164	0.0333	0.0536	0.0758	0.0988	0.1217	0.1438	0.1647	
3	0.0002	0.0011	0.0033	0.0071	0.0126	0.0198	0.0284	0.0383	0.0494	
4		0.0001	0.0003	0.0007	0.0016	0.0030	0.0050	0.0077	0.0111	
5			0.0000	0.0001	0.0002	0.0004	0.0007	0.0012	0.0020	
6							0.0001	0.0002	0.0003	
$k \ \lambda$	1	2	3	4	5	6	7	8	9	10
0	0.3679	0.1353	0.0498	0.0183	0.0067	0.0025	0.0009	0.0003	0.0001	0.00005
1	0.3679	0.2707	0.1494	0.0733	0.0337	0.0149	0.0064	0.0027	0.0005	0.0002
2	0.1839	0.2707	0.2240	0.1465	0.0842	0.0446	0.0223	0.0050	0.0023	0.0010
3	0.0613	0.1804	0.2240	0.1954	0.1404	0.0892	0.0521	0.0286	0.0150	0.0037
4	0.0153	0.0902	0.1680	0.1954	0.1755	0.1339	0.0912	0.0572	0.0337	0.0189
5	0.0031	0.0361	0.1008	0.1563	0.1755	0.1606	0.1277	0.0916	0.0607	0.0378
6	0.0005	0.0120	0.0504	0.1042	0.1462	0.1606	0.1490	0.1221	0.0911	0.0631
7	0.0001	0.0034	0.0216	0.0595	0.1044	0.1377	0.1490	0.1396	0.1171	0.0901
8		0.0009	0.0081	0.0298	0.0653	0.1033	0.1304	0.1396	0.1318	0.1126
9		0.0002	0.0027	0.0132	0.0363	0.0688	0.1014	0.1241	0.1318	0.1251
10			0.0008	0.0053	0.0181	0.0413	0.0710	0.0993	0.1186	0.1251
11			0.0002	0.0019	0.0082	0.0225	0.0452	0.0722	0.0970	0.1137
12			0.0001	0.0006	0.0034	0.0113	0.0263	0.0481	0.0728	0.0948
13				0.0002	0.0013	0.0052	0.0142	0.0296	0.0504	0.0729
14				0.0001	0.0005	0.0022	0.0071	0.0169	0.0324	0.0521
15					0.0002	0.0009	0.0033	0.0090	0.0194	0.0347
16					0.0001	0.0003	0.0014	0.0045	0.0109	0.0217
17						0.0001	0.0006	0.0021	0.0058	0.0128
18							0.0002	0.0009	0.0029	0.0071
19							0.0001	0.0004	0.0014	0.0037
20								0.0002	0.0006	0.0019
21								0.0001	0.0003	0.0009
22									0.0001	0.0006
23										0.0002

Ejemplo: Para una variable aleatoria X de ley de Poisson de parámetro $\lambda = 3$ se tiene $P(X_{\lambda} = 5) = 0.1008$.

Tabla de los valores χ_n^2

Valor a de una χ_n^2 teniendo la probabilidad α de ser sobrepasado. $P(\chi_n^2 \geq a) = \alpha$.



g.d.l. α	0.90	0.50	0.30	0.20	0.10	0.05	0.02	0.01	0.005	0.001
1	0.0158	0.455	1.074	1.642	2.706	3.841	5.412	6.635	7.879	10.827
2	0.211	1.386	2.408	3.219	4.605	5.991	7.824	9.210	10.597	13.815
3	0.584	2.366	3.665	4.642	6.251	7.815	9.837	11.345	12.838	16.266
4	1.064	3.357	4.878	5.989	7.779	9.488	11.668	13.277	14.860	18.467
5	1.610	4.351	6.064	7.289	9.236	11.070	13.388	15.086	16.750	20.515
6	2.204	5.348	7.231	8.558	10.645	12.592	15.033	15.812	18.548	22.457
7	2.833	6.346	8.383	9.803	12.017	14.067	16.622	18.475	20.278	24.322
8	3.490	7.344	9.524	11.030	13.362	15.507	18.162	20.090	21.955	26.125
9	4.168	8.343	10.656	12.242	14.684	16.919	19.679	21.666	23.589	27.877
10	4.865	9.342	11.781	13.442	15.987	18.307	21.161	23.209	25.188	29.588
11	5.578	10.341	12.899	14.631	17.275	19.675	22.618	24.725	26.757	31.264
12	6.304	11.340	14.011	15.812	18.549	21.026	24.054	26.217	28.300	32.909
13	7.042	12.340	15.119	16.985	19.812	22.362	25.472	27.688	29.819	34.528
14	7.790	13.339	16.222	18.151	21.064	23.685	26.873	29.141	31.319	36.123
15	8.547	14.339	17.322	19.311	22.307	24.996	28.259	30.578	32.801	37.697
16	9.312	15.338	18.418	20.465	23.542	26.296	29.633	32.000	34.267	39.252
17	10.085	16.338	19.511	21.615	24.769	27.587	30.995	33.409	35.718	40.790
18	10.265	17.338	20.601	22.760	25.989	28.869	32.346	34.805	37.156	42.312
19	11.651	18.338	21.689	23.900	27.204	30.144	33.687	36.191	38.582	43.820
20	12.443	19.337	22.775	25.038	28.412	31.410	35.020	37.566	39.997	45.315
21	13.240	20.337	23.858	26.171	29.615	32.671	36.343	38.932	41.401	46.797
22	14.041	21.337	24.939	27.301	30.813	33.924	37.659	40.289	42.796	48.268
23	14.848	22.337	26.018	28.429	32.007	35.172	38.968	41.638	44.181	49.728
24	15.659	23.337	27.096	29.553	33.196	36.415	40.270	42.980	45.558	51.179
25	16.473	24.337	28.172	30.675	34.382	37.652	41.566	44.314	46.928	52.620
26	17.292	25.336	29.246	31.795	35.563	38.885	42.855	45.642	48.290	54.052
27	18.114	26.336	30.319	32.912	36.741	40.113	44.140	46.963	49.645	55.476
28	18.939	27.336	31.391	34.027	37.916	41.337	45.419	48.278	50.993	56.893
29	19.768	28.336	32.461	35.139	39.087	42.557	46.693	49.588	52.336	58.302
30	20.599	29.336	33.530	36.250	40.256	43.773	47.962	50.892	53.672	59.703

Ejemplo: para 11 g.d.l. se tiene $P(\chi^2 \geq 17.275) = 0.10$.

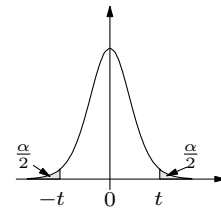
Ley de probabilidad t_n de Student

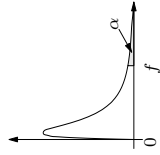
n	α	90%	80%	70%	60%	50%	40%	30%	20%	10%	5%	2%	1%	0.10%
1		0.158	0.325	0.510	0.727	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.857	636.619
2		0.142	0.289	0.445	0.617	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	31.598
3		0.137	0.277	0.424	0.584	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	12.941
4		0.134	0.271	0.414	0.569	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	8.610
5		0.132	0.267	0.408	0.559	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	6.589
6		0.131	0.265	0.404	0.553	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.959
7		0.130	0.263	0.402	0.549	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	5.405
8		0.130	0.262	0.399	0.546	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	5.041
9		0.129	0.261	0.398	0.543	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.781
10		0.129	0.260	0.397	0.542	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.587
11		0.129	0.260	0.396	0.540	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.437
12		0.128	0.259	0.395	0.539	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	4.318
13		0.128	0.259	0.394	0.538	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	4.221
14		0.128	0.258	0.393	0.537	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	4.140
15		0.128	0.258	0.393	0.536	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	4.073
16		0.128	0.258	0.392	0.535	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	4.015
17		0.128	0.257	0.392	0.534	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.965
18		0.127	0.257	0.392	0.534	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.922
19		0.127	0.257	0.391	0.533	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.883
20		0.127	0.257	0.391	0.533	0.687	0.860	1.064	1.325	1.725	2.086	2.520	2.845	3.850
21		0.127	0.257	0.391	0.532	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.819
22		0.127	0.256	0.390	0.532	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.792
23		0.127	0.256	0.390	0.532	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.767
24		0.127	0.256	0.390	0.531	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.745
25		0.127	0.256	0.390	0.531	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.725
26		0.127	0.256	0.390	0.531	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.707
27		0.127	0.256	0.389	0.531	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.690
28		0.127	0.256	0.389	0.530	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.674
29		0.127	0.256	0.389	0.530	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.659
30		0.127	0.256	0.389	0.530	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.646
40		0.126	0.255	0.388	0.529	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.551
60		0.126	0.254	0.387	0.527	0.679	0.848	1.046	1.296	1.671	2.000	2.390	2.660	3.460
120		0.126	0.254	0.386	0.526	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	3.373
∞		0.126	0.253	0.385	0.524	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.291

Los niveles dados se refieren a pruebas bilaterales. En el caso de una prueba unilateral, se divide el nivel por 2.

n es el número de grados de libertad, $P(|t_n| > t) = \alpha$.

Ejemplo: para $n = 11$, $P(|t_{11}| > 0.540) = 0.60$.

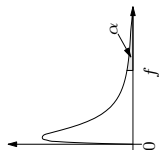




Valores críticos de la Estadística F , ($\alpha = 0.25$).

$v_1 \rightarrow$	Número de grados de libertad del numerador																	∞		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	48	60	120	∞
1	5.828	7.500	8.200	8.581	8.820	8.983	9.102	9.192	9.263	9.320	9.406	9.493	9.581	9.625	9.670	9.714	9.737	9.759	9.804	9.849
2	2.571	3.000	3.153	3.232	3.280	3.312	3.335	3.353	3.366	3.377	3.393	3.410	3.426	3.434	3.443	3.451	3.455	3.459	3.468	3.476
3	2.024	2.280	2.355	2.390	2.409	2.422	2.430	2.436	2.441	2.445	2.450	2.455	2.460	2.463	2.465	2.467	2.469	2.470	2.472	2.474
4	1.807	2.000	2.047	2.064	2.072	2.077	2.079	2.080	2.081	2.082	2.083	2.083	2.083	2.083	2.082	2.082	2.082	2.082	2.081	2.081
5	1.692	1.853	1.884	1.893	1.895	1.894	1.893	1.892	1.891	1.890	1.888	1.885	1.882	1.880	1.878	1.876	1.875	1.874	1.872	1.869
6	1.621	1.762	1.784	1.787	1.785	1.782	1.779	1.776	1.773	1.771	1.767	1.762	1.757	1.754	1.751	1.748	1.746	1.744	1.741	1.737
7	1.573	1.701	1.717	1.716	1.711	1.706	1.701	1.697	1.693	1.690	1.684	1.678	1.671	1.667	1.663	1.659	1.657	1.655	1.650	1.645
8	1.538	1.657	1.668	1.664	1.657	1.651	1.645	1.640	1.635	1.631	1.624	1.617	1.609	1.604	1.600	1.594	1.592	1.589	1.584	1.578
9	1.512	1.624	1.631	1.625	1.617	1.609	1.602	1.596	1.591	1.586	1.579	1.570	1.561	1.556	1.551	1.545	1.542	1.539	1.532	1.526
10	1.491	1.597	1.603	1.595	1.585	1.576	1.569	1.562	1.556	1.551	1.543	1.534	1.523	1.518	1.512	1.506	1.502	1.499	1.492	1.484
11	1.475	1.577	1.580	1.570	1.560	1.550	1.542	1.535	1.528	1.523	1.514	1.504	1.493	1.487	1.480	1.474	1.470	1.466	1.459	1.450
12	1.461	1.559	1.561	1.550	1.539	1.529	1.520	1.512	1.505	1.500	1.490	1.480	1.468	1.461	1.454	1.447	1.443	1.439	1.431	1.422
13	1.450	1.545	1.545	1.534	1.521	1.510	1.501	1.493	1.486	1.480	1.470	1.459	1.446	1.440	1.432	1.425	1.421	1.416	1.407	1.398
14	1.440	1.533	1.532	1.519	1.507	1.495	1.485	1.477	1.470	1.463	1.453	1.441	1.428	1.421	1.414	1.405	1.401	1.397	1.387	1.377
15	1.432	1.523	1.520	1.507	1.494	1.482	1.472	1.463	1.456	1.449	1.438	1.426	1.413	1.405	1.397	1.389	1.384	1.380	1.370	1.359
16	1.425	1.514	1.510	1.496	1.483	1.470	1.460	1.451	1.443	1.437	1.425	1.413	1.399	1.391	1.383	1.374	1.369	1.365	1.354	1.343
17	1.419	1.506	1.501	1.487	1.473	1.460	1.450	1.440	1.432	1.426	1.414	1.401	1.387	1.379	1.370	1.361	1.356	1.351	1.341	1.329
18	1.413	1.499	1.494	1.479	1.464	1.452	1.441	1.431	1.423	1.416	1.404	1.391	1.376	1.368	1.359	1.350	1.345	1.339	1.328	1.316
19	1.408	1.492	1.487	1.472	1.457	1.444	1.432	1.423	1.414	1.407	1.395	1.382	1.367	1.358	1.349	1.339	1.334	1.329	1.317	1.305
20	1.404	1.487	1.481	1.465	1.450	1.437	1.425	1.415	1.407	1.399	1.387	1.374	1.358	1.349	1.340	1.330	1.325	1.319	1.307	1.294
21	1.400	1.482	1.475	1.459	1.444	1.430	1.419	1.409	1.400	1.392	1.380	1.366	1.350	1.341	1.332	1.322	1.316	1.310	1.298	1.285
22	1.396	1.477	1.470	1.454	1.438	1.424	1.413	1.402	1.394	1.386	1.373	1.359	1.343	1.334	1.324	1.314	1.308	1.302	1.290	1.276
23	1.393	1.473	1.466	1.449	1.433	1.419	1.407	1.397	1.388	1.380	1.367	1.353	1.337	1.327	1.318	1.307	1.301	1.295	1.282	1.268
24	1.390	1.469	1.461	1.445	1.428	1.414	1.402	1.392	1.383	1.375	1.362	1.347	1.331	1.321	1.311	1.300	1.294	1.288	1.275	1.261
25	1.387	1.466	1.458	1.441	1.424	1.410	1.398	1.387	1.378	1.370	1.357	1.342	1.325	1.316	1.306	1.294	1.288	1.282	1.269	1.254
30	1.376	1.452	1.443	1.424	1.407	1.392	1.379	1.368	1.359	1.351	1.337	1.321	1.303	1.293	1.282	1.270	1.264	1.257	1.242	1.226
40	1.363	1.435	1.424	1.404	1.386	1.371	1.357	1.345	1.335	1.327	1.312	1.295	1.276	1.265	1.253	1.240	1.232	1.225	1.208	1.188
60	1.349	1.419	1.405	1.385	1.366	1.349	1.335	1.323	1.312	1.303	1.287	1.269	1.248	1.236	1.223	1.208	1.200	1.191	1.171	1.147
80	1.343	1.411	1.396	1.375	1.355	1.338	1.324	1.311	1.300	1.291	1.275	1.256	1.234	1.221	1.207	1.192	1.182	1.173	1.151	1.123
120	1.336	1.402	1.387	1.365	1.345	1.328	1.313	1.300	1.289	1.279	1.262	1.243	1.220	1.207	1.192	1.175	1.165	1.155	1.131	1.099
∞	1.323	1.386	1.369	1.346	1.325	1.307	1.291	1.277	1.265	1.255	1.237	1.216	1.191	1.177	1.160	1.140	1.128	1.116	1.084	1.000

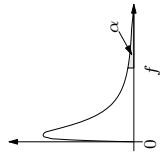
Ejemplo: para $v_1 = 10$, $v_2 = 17$ se tiene $P(F \geq 1.426) = 0.25$



Valores críticos de la Estadística de Fisher F , ($\alpha = 0.10$).

$v_1 \longrightarrow$ Número de grados de libertad del numerador	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	48	60	120	∞
1	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86	60.19	60.70	61.22	61.74	62.00	62.26	62.53	62.66	62.79	63.06	63.33
2	8.526	9.000	9.162	9.243	9.293	9.325	9.349	9.367	9.380	9.392	9.408	9.425	9.441	9.450	9.458	9.466	9.470	9.475	9.483	9.491
3	5.538	5.462	5.391	5.343	5.309	5.285	5.266	5.252	5.240	5.230	5.216	5.200	5.184	5.176	5.168	5.160	5.155	5.151	5.142	5.134
4	4.545	4.325	4.191	4.107	4.051	4.010	3.979	3.955	3.936	3.920	3.905	3.869	3.844	3.831	3.817	3.804	3.797	3.790	3.775	3.761
5	4.060	3.780	3.619	3.520	3.453	3.404	3.368	3.339	3.316	3.297	3.268	3.238	3.207	3.190	3.174	3.157	3.149	3.140	3.123	3.105
6	3.776	3.463	3.289	3.181	3.107	3.055	3.014	2.983	2.958	2.937	2.905	2.871	2.836	2.818	2.800	2.781	2.772	2.762	2.742	2.722
7	3.589	3.257	3.074	2.960	2.883	2.827	2.785	2.752	2.725	2.702	2.668	2.632	2.595	2.575	2.555	2.535	2.525	2.514	2.493	2.471
8	3.458	3.113	2.924	2.806	2.726	2.668	2.624	2.589	2.561	2.538	2.502	2.464	2.425	2.404	2.383	2.361	2.350	2.339	2.316	2.293
9	3.360	3.006	2.813	2.693	2.611	2.551	2.505	2.469	2.440	2.416	2.379	2.340	2.298	2.277	2.255	2.232	2.220	2.208	2.184	2.159
10	3.285	2.924	2.728	2.605	2.522	2.461	2.414	2.377	2.347	2.323	2.284	2.243	2.201	2.178	2.155	2.132	2.119	2.107	2.082	2.055
11	3.225	2.859	2.660	2.536	2.451	2.389	2.342	2.304	2.273	2.248	2.209	2.167	2.123	2.100	2.076	2.052	2.039	2.026	2.000	1.972
12	3.176	2.807	2.605	2.480	2.394	2.331	2.283	2.245	2.213	2.188	2.147	2.105	2.060	2.036	2.011	1.986	1.973	1.960	1.932	1.904
13	3.136	2.763	2.560	2.434	2.347	2.283	2.234	2.195	2.164	2.138	2.097	2.053	2.007	1.983	1.958	1.931	1.918	1.904	1.876	1.846
14	3.102	2.726	2.522	2.395	2.307	2.243	2.193	2.154	2.122	2.095	2.054	2.009	1.962	1.938	1.912	1.885	1.871	1.857	1.828	1.797
15	3.073	2.695	2.490	2.361	2.273	2.208	2.158	2.118	2.086	2.059	2.017	1.972	1.924	1.899	1.873	1.845	1.831	1.817	1.787	1.755
16	3.048	2.668	2.462	2.333	2.244	2.178	2.128	2.088	2.055	2.028	1.985	1.940	1.891	1.866	1.839	1.811	1.796	1.782	1.751	1.718
17	3.026	2.645	2.437	2.308	2.218	2.152	2.102	2.061	2.028	2.001	1.958	1.912	1.862	1.836	1.809	1.780	1.766	1.751	1.719	1.686
18	3.007	2.624	2.416	2.286	2.196	2.130	2.078	2.038	2.005	1.977	1.933	1.887	1.837	1.810	1.783	1.754	1.738	1.723	1.691	1.657
19	2.990	2.606	2.397	2.266	2.176	2.109	2.058	2.017	1.984	1.956	1.912	1.865	1.814	1.787	1.759	1.730	1.714	1.699	1.666	1.631
20	2.975	2.589	2.380	2.249	2.158	2.091	2.040	1.998	1.965	1.937	1.892	1.845	1.794	1.767	1.738	1.708	1.693	1.677	1.643	1.607
21	2.961	2.575	2.365	2.233	2.142	2.075	2.023	1.982	1.948	1.920	1.875	1.827	1.776	1.748	1.719	1.689	1.673	1.657	1.623	1.586
22	2.949	2.561	2.351	2.219	2.128	2.060	2.008	1.967	1.933	1.904	1.859	1.811	1.759	1.731	1.702	1.671	1.655	1.639	1.604	1.567
23	2.937	2.549	2.339	2.206	2.115	2.047	1.995	1.953	1.919	1.890	1.845	1.796	1.744	1.716	1.686	1.655	1.639	1.622	1.587	1.549
24	2.927	2.538	2.327	2.195	2.103	2.035	1.983	1.941	1.906	1.877	1.832	1.783	1.730	1.702	1.672	1.641	1.624	1.607	1.571	1.533
25	2.918	2.528	2.317	2.184	2.092	2.024	1.971	1.929	1.895	1.866	1.820	1.771	1.717	1.689	1.659	1.627	1.610	1.593	1.557	1.518
30	2.881	2.489	2.276	2.142	2.049	1.980	1.927	1.884	1.849	1.819	1.773	1.722	1.667	1.638	1.606	1.573	1.555	1.538	1.499	1.456
40	2.835	2.440	2.226	2.091	1.997	1.927	1.872	1.829	1.793	1.763	1.715	1.662	1.605	1.574	1.541	1.506	1.486	1.467	1.425	1.377
60	2.791	2.393	2.177	2.041	1.946	1.875	1.819	1.775	1.738	1.707	1.657	1.603	1.543	1.511	1.475	1.437	1.416	1.395	1.348	1.291
80	2.770	2.370	2.154	2.017	1.921	1.849	1.793	1.748	1.711	1.680	1.629	1.574	1.513	1.479	1.442	1.402	1.380	1.358	1.306	1.242
120	2.748	2.347	2.130	1.992	1.896	1.824	1.767	1.722	1.684	1.652	1.601	1.545	1.482	1.447	1.409	1.368	1.344	1.320	1.265	1.193
∞	2.705	2.303	2.084	1.945	1.847	1.774	1.717	1.670	1.631	1.599	1.546	1.487	1.421	1.383	1.342	1.295	1.268	1.240	1.169	1.000

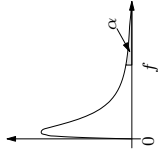
Ejemplo: para $v_1 = 10, v_2 = 17$ se tiene $P(F \geq 2.001) = 0.10$



Valores críticos de la Estadística F , ($\alpha = 0.05$).

$v_1 \rightarrow$	Número de grados de libertad del numerador																	∞		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	48	60	120	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	239.9	240.5	241.9	243.9	245.9	248.0	249.0	250.1	251.1	251.7	252.2	253.2	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.47	19.48	19.49	19.50
3	10.13	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.785	8.745	8.703	8.660	8.638	8.617	8.594	8.583	8.572	8.549	8.526
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964	5.912	5.858	5.802	5.774	5.746	5.717	5.702	5.688	5.658	5.628
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735	4.678	4.619	4.558	4.527	4.496	4.464	4.448	4.431	4.398	4.365
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060	4.000	3.938	3.874	3.841	3.808	3.774	3.757	3.740	3.705	3.669
7	5.591	4.737	4.347	4.120	3.971	3.866	3.787	3.726	3.677	3.636	3.575	3.511	3.444	3.410	3.376	3.340	3.322	3.304	3.267	3.230
8	5.318	4.459	4.066	3.838	3.687	3.581	3.500	3.438	3.388	3.347	3.284	3.218	3.150	3.115	3.079	3.043	3.024	3.005	2.967	2.928
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137	3.073	3.006	2.936	2.900	2.864	2.826	2.807	2.787	2.747	2.707
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978	2.913	2.845	2.774	2.737	2.700	2.661	2.641	2.621	2.580	2.538
11	4.844	3.982	3.587	3.357	3.204	3.095	3.012	2.948	2.896	2.854	2.788	2.719	2.646	2.609	2.570	2.531	2.510	2.490	2.448	2.404
12	4.747	3.885	3.490	3.259	3.106	2.996	2.913	2.849	2.796	2.753	2.687	2.617	2.544	2.505	2.466	2.426	2.405	2.384	2.341	2.296
13	4.667	3.806	3.410	3.179	3.025	2.915	2.832	2.767	2.714	2.671	2.604	2.533	2.459	2.420	2.380	2.339	2.318	2.297	2.252	2.206
14	4.600	3.739	3.340	3.112	2.958	2.848	2.764	2.699	2.646	2.602	2.534	2.463	2.388	2.349	2.308	2.266	2.245	2.223	2.178	2.131
15	4.543	3.682	3.287	3.056	2.901	2.790	2.707	2.641	2.588	2.544	2.475	2.403	2.327	2.288	2.247	2.204	2.182	2.160	2.114	2.066
16	4.494	3.634	3.239	3.007	2.852	2.741	2.657	2.591	2.538	2.493	2.425	2.352	2.276	2.235	2.194	2.151	2.128	2.106	2.059	2.010
17	4.451	3.591	3.197	2.965	2.810	2.699	2.614	2.548	2.494	2.449	2.381	2.308	2.230	2.190	2.148	2.104	2.081	2.058	2.011	1.960
18	4.414	3.555	3.160	2.928	2.773	2.661	2.577	2.510	2.456	2.412	2.342	2.269	2.191	2.150	2.107	2.063	2.040	2.017	1.968	1.917
19	4.381	3.522	3.127	2.895	2.740	2.628	2.543	2.477	2.423	2.378	2.308	2.234	2.155	2.114	2.071	2.026	2.003	1.980	1.930	1.878
20	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.447	2.393	2.348	2.278	2.203	2.124	2.082	2.039	1.994	1.970	1.946	1.896	1.843
21	4.325	3.467	3.072	2.840	2.685	2.573	2.488	2.420	2.366	2.321	2.250	2.176	2.096	2.054	2.010	1.964	1.940	1.916	1.866	1.812
22	4.301	3.443	3.049	2.817	2.661	2.549	2.464	2.396	2.342	2.297	2.226	2.151	2.071	2.028	1.984	1.938	1.914	1.889	1.838	1.783
23	4.279	3.422	3.028	2.795	2.640	2.528	2.442	2.375	2.320	2.275	2.204	2.128	2.048	2.005	1.960	1.914	1.889	1.865	1.813	1.757
24	4.260	3.403	3.009	2.776	2.621	2.508	2.423	2.355	2.300	2.255	2.183	2.108	2.027	1.984	1.939	1.892	1.867	1.842	1.790	1.733
25	4.242	3.385	2.991	2.759	2.603	2.490	2.405	2.337	2.282	2.236	2.165	2.089	2.007	1.964	1.919	1.872	1.847	1.822	1.768	1.711
30	4.171	3.316	2.922	2.690	2.534	2.420	2.334	2.266	2.211	2.165	2.092	2.015	1.932	1.887	1.841	1.792	1.766	1.740	1.683	1.622
40	4.085	3.232	2.839	2.606	2.449	2.336	2.249	2.180	2.124	2.077	2.003	1.924	1.839	1.793	1.744	1.693	1.665	1.637	1.579	1.509
60	4.001	3.150	2.758	2.525	2.360	2.254	2.166	2.097	2.040	1.993	1.917	1.836	1.748	1.700	1.649	1.594	1.564	1.534	1.467	1.389
80	3.961	3.111	2.719	2.486	2.329	2.214	2.127	2.057	1.999	1.952	1.876	1.793	1.703	1.654	1.602	1.545	1.513	1.482	1.410	1.322
120	3.920	3.072	2.680	2.447	2.290	2.175	2.087	2.016	1.959	1.910	1.834	1.750	1.659	1.608	1.554	1.495	1.462	1.429	1.352	1.254
∞	3.841	2.996	2.605	2.372	2.214	2.099	2.010	1.938	1.880	1.831	1.752	1.666	1.570	1.517	1.459	1.394	1.356	1.318	1.221	1.000

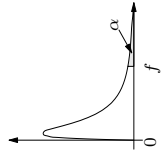
Ejemplo: para $v_1 = 10$, $v_2 = 17$ se tiene $P(F \geq 2.450) = 0.05$



Valores críticos de la Estadística F , ($\alpha = 0.025$),

$v_1 \rightarrow$	Número de grados de libertad del numerador																	∞		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	48	60	120	∞
1	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.7	963.3	968.7	976.7	984.9	993.1	997.2	1001.0	1006.0	1008.0	1010.0	1014.0	1018.0
2	38.510	39.000	39.160	39.250	39.300	39.330	39.350	39.370	39.390	39.400	39.410	39.430	39.450	39.460	39.460	39.470	39.480	39.480	39.490	39.500
3	17.440	16.040	15.440	15.100	14.880	14.730	14.620	14.540	14.470	14.420	14.340	14.250	14.170	14.120	14.080	14.040	14.010	13.990	13.950	13.900
4	12.220	10.650	9.980	9.604	9.364	9.197	9.074	8.980	8.905	8.844	8.751	8.656	8.560	8.511	8.461	8.411	8.386	8.360	8.309	8.257
5	10.010	8.434	7.764	7.388	7.146	6.978	6.853	6.757	6.681	6.619	6.525	6.428	6.328	6.278	6.227	6.175	6.149	6.122	6.069	6.015
6	8.813	7.260	6.599	6.227	5.988	5.820	5.695	5.600	5.523	5.461	5.366	5.269	5.168	5.117	5.065	5.012	4.986	4.959	4.904	4.849
7	8.073	6.541	5.890	5.523	5.285	5.119	4.995	4.899	4.823	4.761	4.666	4.568	4.467	4.415	4.362	4.309	4.282	4.254	4.199	4.142
8	7.571	6.059	5.416	5.053	4.817	4.652	4.529	4.433	4.357	4.295	4.200	4.101	3.999	3.947	3.894	3.840	3.812	3.784	3.728	3.670
9	7.209	5.715	5.078	4.718	4.484	4.320	4.197	4.102	4.026	3.964	3.868	3.769	3.667	3.614	3.560	3.505	3.477	3.449	3.392	3.333
10	6.937	5.456	4.826	4.468	4.236	4.072	3.950	3.855	3.779	3.717	3.621	3.522	3.419	3.365	3.311	3.255	3.227	3.198	3.140	3.080
11	6.724	5.256	4.630	4.275	4.044	3.881	3.759	3.664	3.588	3.526	3.430	3.330	3.226	3.172	3.118	3.061	3.032	3.003	2.944	2.883
12	6.554	5.096	4.474	4.121	3.891	3.728	3.606	3.512	3.436	3.374	3.277	3.177	3.073	3.019	2.963	2.906	2.877	2.848	2.787	2.725
13	6.414	4.965	4.347	3.996	3.767	3.604	3.483	3.388	3.312	3.250	3.153	3.053	2.948	2.893	2.837	2.780	2.750	2.720	2.659	2.595
14	6.298	4.857	4.242	3.892	3.663	3.501	3.380	3.285	3.209	3.147	3.050	2.949	2.844	2.789	2.732	2.674	2.644	2.614	2.552	2.487
15	6.199	4.765	4.153	3.804	3.576	3.415	3.293	3.199	3.123	3.060	2.963	2.862	2.756	2.701	2.644	2.585	2.555	2.524	2.461	2.395
16	6.115	4.687	4.077	3.729	3.502	3.341	3.219	3.125	3.049	2.986	2.889	2.787	2.681	2.625	2.568	2.508	2.478	2.447	2.383	2.316
17	6.042	4.619	4.011	3.665	3.438	3.277	3.156	3.061	2.985	2.922	2.825	2.723	2.616	2.560	2.502	2.442	2.411	2.380	2.315	2.247
18	5.978	4.560	3.954	3.608	3.382	3.221	3.100	3.005	2.929	2.866	2.769	2.667	2.559	2.503	2.444	2.384	2.353	2.321	2.256	2.187
19	5.922	4.507	3.903	3.559	3.333	3.172	3.051	2.956	2.880	2.817	2.720	2.617	2.509	2.452	2.394	2.333	2.301	2.269	2.203	2.133
20	5.871	4.461	3.859	3.515	3.289	3.128	3.007	2.913	2.836	2.774	2.676	2.573	2.464	2.408	2.349	2.287	2.255	2.223	2.156	2.085
21	5.827	4.420	3.819	3.475	3.250	3.089	2.969	2.874	2.798	2.735	2.637	2.534	2.425	2.367	2.308	2.246	2.214	2.182	2.114	2.042
22	5.786	4.383	3.783	3.440	3.215	3.055	2.934	2.839	2.763	2.700	2.602	2.498	2.389	2.331	2.272	2.210	2.177	2.145	2.076	2.003
23	5.750	4.349	3.750	3.408	3.183	3.023	2.902	2.808	2.731	2.668	2.570	2.466	2.357	2.299	2.239	2.176	2.143	2.111	2.041	1.968
24	5.717	4.319	3.721	3.379	3.155	2.995	2.874	2.779	2.703	2.640	2.541	2.437	2.327	2.269	2.209	2.146	2.113	2.080	2.010	1.935
25	5.686	4.291	3.694	3.353	3.129	2.968	2.848	2.753	2.677	2.613	2.515	2.411	2.300	2.242	2.182	2.118	2.085	2.052	1.981	1.905
30	5.567	4.182	3.589	3.250	3.026	2.867	2.746	2.651	2.575	2.511	2.412	2.307	2.195	2.136	2.074	2.009	1.974	1.940	1.866	1.787
40	5.424	4.051	3.463	3.126	2.904	2.744	2.624	2.529	2.452	2.388	2.288	2.182	2.068	2.007	1.943	1.875	1.839	1.803	1.724	1.637
60	5.286	3.925	3.342	3.008	2.786	2.627	2.507	2.412	2.334	2.270	2.169	2.061	1.944	1.882	1.815	1.744	1.705	1.667	1.581	1.482
80	5.219	3.865	3.285	2.951	2.730	2.571	2.451	2.356	2.278	2.214	2.112	2.003	1.885	1.821	1.753	1.679	1.639	1.598	1.507	1.396
120	5.152	3.805	3.227	2.894	2.674	2.515	2.395	2.299	2.222	2.157	2.055	1.945	1.825	1.760	1.690	1.614	1.572	1.530	1.433	1.310
∞	5.024	3.689	3.116	2.786	2.566	2.408	2.287	2.192	2.114	2.048	1.945	1.833	1.708	1.640	1.566	1.483	1.436	1.388	1.288	1.000

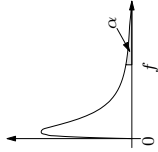
Ejemplo: para $v_1 = 10$, $v_2 = 17$ se tiene $P(F \geq 2.922) = 0.025$



Valores críticos de la Estadística F , ($\alpha = 0.01$).

$v_1 \rightarrow$ Número de grados de libertad del numerador	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	48	60	120	∞
1	4052.0	4999.0	5403.0	5625.0	5764.0	5859.0	5928.0	5981.0	6022.0	6056.0	6106.0	6157.0	6209.0	6235.0	6261.0	6287.0	6300.0	6313.0	6339.0	6366.0
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.48	99.49	99.50
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.34	27.23	27.05	26.87	26.69	26.60	26.50	26.41	26.36	26.32	26.22	26.12
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.20	14.02	13.93	13.84	13.74	13.70	13.65	13.56	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.888	9.722	9.553	9.466	9.379	9.291	9.247	9.202	9.112	9.020
6	13.74	10.92	9.779	9.148	8.746	8.466	8.260	8.102	7.976	7.874	7.718	7.559	7.396	7.313	7.228	7.143	7.100	7.057	6.969	6.880
7	12.25	9.547	8.451	7.847	7.460	7.191	6.993	6.840	6.719	6.620	6.469	6.314	6.155	6.074	5.992	5.908	5.866	5.824	5.737	5.649
8	11.26	8.649	7.591	7.006	6.632	6.371	6.178	6.029	5.911	5.814	5.667	5.515	5.359	5.279	5.198	5.116	5.074	5.032	4.946	4.859
9	10.56	8.021	6.992	6.422	6.057	5.802	5.613	5.467	5.351	5.256	5.111	4.962	4.808	4.729	4.649	4.567	4.525	4.483	4.398	4.310
10	10.04	7.559	6.552	5.994	5.636	5.386	5.200	5.057	4.942	4.849	4.706	4.558	4.405	4.327	4.247	4.165	4.124	4.082	3.996	3.909
11	9.646	7.206	6.217	5.668	5.316	5.069	4.886	4.744	4.631	4.539	4.397	4.251	4.099	4.021	3.941	3.860	3.818	3.776	3.690	3.602
12	9.330	6.927	5.953	5.412	5.064	4.821	4.639	4.499	4.387	4.296	4.155	4.010	3.858	3.780	3.701	3.619	3.577	3.535	3.449	3.361
13	9.074	6.701	5.739	5.205	4.862	4.620	4.441	4.302	4.191	4.100	3.960	3.815	3.663	3.587	3.507	3.425	3.383	3.341	3.255	3.165
14	8.862	6.515	5.564	5.035	4.695	4.456	4.278	4.140	4.030	3.939	3.800	3.656	3.505	3.427	3.348	3.266	3.223	3.181	3.094	3.004
15	8.683	6.359	5.417	4.893	4.556	4.318	4.141	4.004	3.895	3.805	3.666	3.522	3.372	3.294	3.214	3.132	3.089	3.047	2.959	2.868
16	8.531	6.226	5.292	4.773	4.437	4.202	4.026	3.890	3.780	3.691	3.553	3.409	3.259	3.181	3.101	3.018	2.976	2.933	2.845	2.753
17	8.400	6.112	5.185	4.669	4.336	4.101	3.927	3.791	3.682	3.593	3.455	3.312	3.161	3.083	3.003	2.920	2.878	2.835	2.746	2.653
18	8.285	6.013	5.092	4.579	4.248	4.015	3.841	3.705	3.597	3.508	3.371	3.227	3.077	2.999	2.918	2.835	2.792	2.749	2.660	2.566
19	8.185	5.926	5.010	4.500	4.171	3.939	3.765	3.630	3.522	3.434	3.296	3.153	3.003	2.925	2.844	2.761	2.717	2.674	2.584	2.489
20	8.096	5.849	4.938	4.431	4.103	3.871	3.699	3.564	3.457	3.368	3.231	3.088	2.938	2.859	2.778	2.695	2.651	2.608	2.517	2.421
21	8.017	5.780	4.874	4.369	4.042	3.812	3.640	3.506	3.398	3.310	3.173	3.030	2.880	2.801	2.720	2.636	2.592	2.548	2.457	2.360
22	7.945	5.719	4.817	4.313	3.988	3.758	3.587	3.453	3.346	3.258	3.121	2.978	2.827	2.749	2.667	2.583	2.539	2.495	2.403	2.305
23	7.881	5.664	4.765	4.263	3.939	3.710	3.539	3.406	3.299	3.211	3.074	2.931	2.780	2.702	2.620	2.535	2.491	2.447	2.354	2.256
24	7.823	5.614	4.718	4.218	3.895	3.667	3.496	3.363	3.256	3.168	3.032	2.889	2.738	2.659	2.577	2.492	2.448	2.403	2.310	2.211
25	7.770	5.560	4.675	4.177	3.855	3.627	3.457	3.324	3.217	3.129	2.993	2.850	2.699	2.620	2.538	2.453	2.408	2.364	2.269	2.169
30	7.562	5.390	4.510	4.018	3.699	3.473	3.304	3.173	3.066	2.979	2.843	2.700	2.549	2.469	2.386	2.299	2.254	2.208	2.111	2.006
40	7.314	5.178	4.313	3.828	3.514	3.291	3.124	2.993	2.886	2.800	2.665	2.522	2.369	2.288	2.203	2.114	2.067	2.019	1.917	1.805
60	7.077	4.977	4.126	3.649	3.339	3.119	2.953	2.823	2.718	2.632	2.496	2.352	2.198	2.115	2.028	1.936	1.886	1.836	1.726	1.601
80	6.964	4.882	4.037	3.564	3.256	3.037	2.872	2.743	2.639	2.552	2.416	2.272	2.116	2.033	1.944	1.849	1.798	1.746	1.630	1.491
120	6.851	4.786	3.949	3.480	3.173	2.956	2.792	2.663	2.559	2.472	2.336	2.191	2.035	1.950	1.860	1.763	1.709	1.656	1.533	1.380
∞	6.635	4.605	3.782	3.319	3.017	2.802	2.639	2.511	2.407	2.321	2.185	2.038	1.878	1.791	1.696	1.592	1.533	1.473	1.325	1.000

Ejemplo: para $v_1 = 10$, $v_2 = 17$ se tiene $P(F \geq 3.593) = 0.01$



Valores críticos de la Estadística F , ($\alpha = 0.005$).

$v_1 \longrightarrow$ Número de grados de libertad del numerador

		1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	48	60	120	∞
1	16211	20000	21615	22500	23056	23437	23715	23925	24091	24224	24426	24630	24836	24940	25044	25148	25201	25253	25359	25465	
2	198.50	199.00	199.20	199.30	199.30	199.30	199.40	199.40	199.40	199.4	199.4	199.4	199.4	199.4	199.5	199.5	199.5	199.5	199.5	199.5	
3	55.55	49.80	47.47	46.19	45.39	44.84	44.43	44.13	43.88	43.69	43.39	43.08	42.78	42.62	42.47	42.31	42.23	42.15	41.99	41.83	
4	31.33	26.28	24.26	23.15	22.46	21.97	21.62	21.35	21.14	20.97	20.70	20.44	20.17	20.03	19.89	19.75	19.68	19.61	19.47	19.32	
5	22.78	18.31	16.53	15.56	14.94	14.51	14.20	13.96	13.77	13.62	13.38	13.15	12.90	12.78	12.66	12.53	12.47	12.40	12.27	12.14	
6	18.63	14.54	12.92	12.03	11.46	11.07	10.79	10.57	10.39	10.25	10.03	9.814	9.589	9.474	9.358	9.241	9.181	9.122	9.001	8.879	
7	16.24	12.40	10.88	10.05	9.522	9.155	8.885	8.678	8.514	8.380	8.176	7.968	7.754	7.645	7.534	7.422	7.366	7.309	7.193	7.076	
8	14.69	11.04	9.596	8.805	8.302	7.952	7.694	7.496	7.339	7.211	7.015	6.814	6.608	6.503	6.396	6.287	6.232	6.177	6.065	5.950	
9	13.61	10.11	8.717	7.956	7.471	7.134	6.885	6.693	6.541	6.417	6.227	6.032	5.832	5.729	5.625	5.519	5.464	5.410	5.300	5.187	
10	12.83	9.427	8.081	7.343	6.872	6.545	6.302	6.116	5.968	5.847	5.661	5.471	5.274	5.173	5.070	4.966	4.913	4.859	4.750	4.638	
11	12.23	8.912	7.600	6.881	6.422	6.101	5.865	5.682	5.537	5.418	5.236	5.049	4.855	4.756	4.654	4.551	4.498	4.445	4.337	4.226	
12	11.75	8.510	7.226	6.521	6.071	5.757	5.524	5.345	5.202	5.085	4.906	4.721	4.530	4.431	4.331	4.228	4.176	4.123	4.015	3.904	
13	11.37	8.186	6.926	6.233	5.791	5.482	5.253	5.076	4.935	4.820	4.643	4.460	4.270	4.173	4.073	3.970	3.918	3.865	3.758	3.646	
14	11.06	7.922	6.630	5.998	5.562	5.257	5.031	4.857	4.717	4.603	4.428	4.247	4.058	3.961	3.862	3.760	3.708	3.655	3.547	3.436	
15	10.80	7.701	6.476	5.803	5.372	5.071	4.847	4.674	4.536	4.424	4.250	4.070	3.883	3.786	3.687	3.585	3.533	3.480	3.372	3.260	
16	10.57	7.514	6.303	5.638	5.212	4.913	4.692	4.521	4.384	4.272	4.099	3.920	3.734	3.638	3.539	3.437	3.385	3.332	3.224	3.111	
17	10.38	7.354	6.156	5.497	5.075	4.779	4.559	4.389	4.253	4.142	3.971	3.793	3.607	3.511	3.412	3.311	3.258	3.206	3.097	2.984	
18	10.22	7.215	6.028	5.375	4.956	4.663	4.445	4.276	4.141	4.030	3.860	3.683	3.498	3.402	3.303	3.201	3.149	3.096	2.987	2.873	
19	10.07	7.093	5.916	5.268	4.853	4.561	4.345	4.177	4.043	3.933	3.763	3.587	3.402	3.306	3.207	3.106	3.053	3.000	2.891	2.776	
20	9.944	6.986	5.818	5.174	4.762	4.472	4.257	4.090	3.956	3.847	3.678	3.502	3.318	3.222	3.123	3.021	2.969	2.916	2.806	2.690	
21	9.829	6.891	5.730	5.091	4.681	4.393	4.179	4.013	3.880	3.771	3.602	3.427	3.243	3.147	3.049	2.947	2.894	2.841	2.730	2.614	
22	9.727	6.906	5.652	5.017	4.609	4.322	4.109	3.944	3.812	3.703	3.535	3.360	3.176	3.081	2.982	2.880	2.827	2.774	2.662	2.545	
23	9.635	6.730	5.582	4.950	4.544	4.259	4.047	3.882	3.750	3.642	3.474	3.300	3.116	3.021	2.922	2.820	2.766	2.713	2.602	2.484	
24	9.551	6.661	5.519	4.890	4.486	4.202	3.990	3.826	3.695	3.587	3.420	3.246	3.062	2.917	2.868	2.765	2.712	2.658	2.546	2.428	
25	9.475	6.598	5.461	4.835	4.433	4.150	3.939	3.776	3.645	3.537	3.370	3.196	3.013	2.918	2.819	2.716	2.662	2.609	2.496	2.376	
30	9.180	6.355	5.239	4.623	4.228	3.949	3.742	3.580	3.450	3.344	3.179	3.006	2.823	2.727	2.628	2.524	2.470	2.415	2.300	2.176	
40	8.828	6.066	4.976	4.374	3.986	3.713	3.509	3.350	3.222	3.117	2.953	2.781	2.598	2.502	2.401	2.296	2.240	2.184	2.063	1.932	
60	8.495	5.795	4.729	4.140	3.760	3.492	3.291	3.134	3.008	2.904	2.742	2.570	2.387	2.290	2.187	2.079	2.021	1.962	1.834	1.688	
80	8.337	5.667	4.613	4.030	3.654	3.388	3.189	3.034	2.908	2.805	2.643	2.472	2.288	2.189	2.086	1.975	1.915	1.855	1.720	1.560	
120	3.179	5.539	4.497	3.921	3.548	3.285	3.087	2.933	2.808	2.705	2.544	2.373	2.188	2.089	1.984	1.871	1.809	1.747	1.605	1.431	
∞	7.879	5.299	4.279	3.715	3.350	3.091	2.897	2.744	2.621	2.519	2.358	2.187	2.000	1.898	1.789	1.669	1.601	1.532	1.364	1.000	

$v_2 \longrightarrow$ Número de grados de libertad del denominador

Ejemplo: para $v_1 = 10$, $v_2 = 17$ se tiene $P(F \geq 4.142) = 0.005$

Valores límites del coeficiente de correlación r

ν	Nivel α			
	0.1	0.05	0.02	0.01
1	0.9877	0.9969	0.9995	0.9999
2	0.9000	0.9500	0.9800	0.9900
3	0.8054	0.8783	0.9343	0.9587
4	0.7293	0.8114	0.8822	0.9172
5	0.6694	0.7545	0.8329	0.8745
6	0.6215	0.7067	0.7887	0.8343
7	0.5822	0.6664	0.7498	0.7977
8	0.5494	0.6319	0.7155	0.7646
9	0.5214	0.6021	0.6851	0.7348
10	0.4973	0.5760	0.6581	0.7079
11	0.4762	0.5529	0.6339	0.6835
12	0.4575	0.5324	0.6120	0.6614
13	0.4409	0.5139	0.5923	0.6411
14	0.4259	0.4973	0.5742	0.6226
15	0.4124	0.4821	0.5577	0.6055
16	0.4000	0.4683	0.5425	0.5897
17	0.3887	0.4555	0.5285	0.5751
18	0.3783	0.4438	0.5155	0.5614
19	0.3687	0.4329	0.5034	0.5487
20	0.3598	0.4227	0.4921	0.5368
25	0.3233	0.3809	0.4451	0.4869
30	0.2960	0.3494	0.4093	0.4487
35	0.2746	0.3246	0.3810	0.4182
40	0.2573	0.3044	0.3578	0.3932
45	0.2428	0.2875	0.3384	0.3721
50	0.2306	0.2732	0.3218	0.3541
60	0.2108	0.2500	0.2948	0.3248
70	0.1954	0.2319	0.2737	0.3017
80	0.1829	0.2172	0.2565	0.283
90	0.1726	0.205	0.2422	0.2673
100	0.1638	0.1946	0.2301	0.2540

Ejemplo: para $\nu = 12$, $P(|r| \geq 0.5324) = 0.05$

Para un coeficiente de correlación total ν es igual al número de pares de observaciones menos dos.

Para un coeficiente de correlación parcial ν es igual a $n - 2$ menos el número de variables eliminadas.

Transformación de r en $z = \frac{1}{2} \log \frac{1+r}{1-r}$

r	z	r	z	r	z
0.00	0.0000	0.45	0.4847	0.900	1.4722
0.01	0.0100	0.46	0.4973	0.910	1.5275
0.02	0.0200	0.47	0.5101	0.920	1.5890
0.03	0.0300	0.48	0.5230	0.930	1.6584
0.04	0.0400	0.49	0.5361	0.940	1.7380
0.05	0.0500	0.50	0.5493	0.950	1.8318
0.06	0.0601	0.51	0.5627	0.960	1.9459
0.07	0.0701	0.52	0.5763	0.961	1.9588
0.08	0.0802	0.53	0.5901	0.962	1.9721
0.09	0.0902	0.54	0.6042	0.963	1.9857
0.10	0.1003	0.55	0.6184	0.964	1.9996
0.11	0.1104	0.56	0.6328	0.965	2.0139
0.12	0.1206	0.57	0.6475	0.966	2.0287
0.13	0.1307	0.58	0.6625	0.967	2.0439
0.14	0.1409	0.59	0.6777	0.968	2.0595
0.15	0.1511	0.60	0.6931	0.969	2.0756
0.16	0.1614	0.61	0.7089	0.970	2.0923
0.17	0.1717	0.62	0.7250	0.971	2.1095
0.18	0.1820	0.63	0.7414	0.972	2.1273
0.19	0.1923	0.64	0.7582	0.973	2.1457
0.20	0.2027	0.65	0.7753	0.974	2.1649
0.21	0.2132	0.66	0.7928	0.975	2.1847
0.22	0.2237	0.67	0.8107	0.976	2.2054
0.23	0.2342	0.68	0.8291	0.977	2.2269
0.24	0.2448	0.69	0.8480	0.978	2.2494
0.25	0.2554	0.70	0.8673	0.979	2.2729
0.26	0.2661	0.71	0.8872	0.980	2.2976
0.27	0.2769	0.72	0.9076	0.981	2.3235
0.28	0.2877	0.73	0.9287	0.982	2.3507
0.29	0.2986	0.74	0.9505	0.983	2.3796
0.30	0.3095	0.75	0.9730	0.984	2.4101
0.31	0.3205	0.76	0.9962	0.985	2.4427
0.32	0.3316	0.77	1.0203	0.986	2.4774
0.33	0.3428	0.78	1.0454	0.987	2.5147
0.34	0.3541	0.79	1.0714	0.988	2.5550
0.35	0.3654	0.80	1.0986	0.989	2.5987
0.36	0.3769	0.81	1.1270	0.990	2.6467
0.37	0.3884	0.82	1.1568	0.991	2.6996
0.38	0.4001	0.83	1.1881	0.992	2.7587
0.39	0.4118	0.84	1.2212	0.993	2.8257
0.40	0.4236	0.85	1.2562	0.994	2.9031
0.41	0.4356	0.86	1.2933	0.995	2.9945
0.42	0.4477	0.87	1.3331	0.996	3.1063
0.43	0.4599	0.88	1.3758	0.997	3.2504
0.44	0.4722	0.89	1.4219	0.998	3.4534

Valores críticos del coeficiente de correlación ρ_s de Spearman

La tabla da en función de N (número de pares) y del nivel α , el valor ρ_s .

N	α	0.05	0.01
4		1.00	
5		0.90	1.00
6		0.83	0.94
7		0.71	0.89
8		0.64	0.83
9		0.60	0.78
10		0.56	0.75
12		0.51	0.71
14		0.46	0.64
16		0.43	0.60
18		0.40	0.56
20		0.38	0.53
22		0.36	0.51
24		0.34	0.48
26		0.33	0.46
28		0.32	0.45
30		0.31	0.43

Ejemplo: para $N = 16$ (pares), $P(\rho_s > 0.60) = 0.01$.

Valor crítico de T de Wilcoxon (muestras pareadas - caso de pequeñas muestras)

α	0.05	0.02	0.01	$\left\{ \begin{array}{l} \text{nivel de significancia para} \\ \text{pruebas bilaterales} \\ \text{nivel de significancia} \\ \text{para pruebas unilaterales} \end{array} \right.$
N	α^*	0.025	0.01	
6		0	-	-
7		2	-	-
8		4	2	0
9		6	3	2
10		8	5	3
11		11	7	5
12		14	10	7
13		17	13	10
14		21	16	13
15		25	20	16
16		30	24	20
17		35	28	23
18		40	33	28
19		46	38	32
20		52	43	38
21		59	49	43
22		66	56	49
23		75	62	55
24		81	69	61
25		89	77	68

N es el número de diferencias no nulas.

Ejemplo: En el caso de 2 muestras pareadas, donde $N = 12$, se tiene:

$P(T \leq 14) \leq 0.05 \implies$ que se rechaza al nivel del 5%, si el T observado es ≤ 14 , con:

H_0 : las dos muestras tienen idéntica distribución

H_1 : las dos muestras tienen distribuciones diferentes (Prueba bilateral).

Tabla para la prueba de Mann-Whitney (nivel 5%)

$n_2 \ n_1$	2	3	4	5	6	7	8	9	10	11
4	-	-	10 - 26							
5	-	6 - 21	11 - 29	17 - 38						
6	-	7 - 23	12 - 32	18 - 42	26 - 52					
7	-	7 - 26	13 - 35	20 - 45	27 - 57	36 - 69				
8	3 - 19	8 - 28	14 - 38	21 - 49	29 - 61	38 - 74	49 - 87			
9	3 - 21	8 - 31	14 - 42	22 - 53	31 - 65	40 - 79	51 - 93	62 - 109		
10	3 - 23	9 - 33	15 - 45	23 - 57	32 - 70	42 - 84	53 - 99	65 - 115	78 - 132	
11	3 - 25	9 - 36	16 - 48	24 - 61	34 - 74	44 - 89	55 - 105	68 - 121	81 - 139	96 - 157
12	4 - 26	10 - 38	17 - 51	26 - 64	35 - 79	46 - 94	58 - 110	71 - 127	84 - 146	99 - 165
13	4 - 28	10 - 41	18 - 54	27 - 68	37 - 83	48 - 99	60 - 116	73 - 134	88 - 152	103 - 172
14	4 - 30	11 - 43	19 - 57	28 - 72	38 - 88	50 - 104	62 - 122	76 - 140	91 - 159	106 - 180
15	4 - 32	11 - 46	20 - 60	29 - 76	40 - 92	52 - 109	65 - 127	79 - 146	94 - 166	110 - 187
16	4 - 34	12 - 48	21 - 63	30 - 80	42 - 96	54 - 114	67 - 133	82 - 152	97 - 173	113 - 195
17	5 - 35	12 - 51	21 - 67	32 - 83	43 - 101	56 - 119	70 - 138	84 - 159	100 - 180	117 - 202
18	5 - 37	13 - 53	22 - 70	33 - 87	45 - 105	58 - 124	72 - 144	87 - 165	103 - 187	121 - 209
19	5 - 39	13 - 56	23 - 73	34 - 91	46 - 110	60 - 129	74 - 150	90 - 171	107 - 193	124 - 217
20	5 - 41	14 - 58	24 - 76	35 - 95	48 - 114	62 - 134	77 - 155	93 - 177	110 - 200	128 - 224

$n_2 \ n_1$	12	13	14	15	16	17	18	19	20
12	115 - 185								
13	119 - 193	136 - 215							
14	123 - 201	141 - 223	160 - 246						
15	127 - 209	145 - 232	164 - 256	184 - 281					
16	131 - 217	150 - 240	169 - 265	190 - 290	211 - 317				
17	135 - 225	154 - 249	174 - 274	195 - 300	217 - 327	240 - 355			
18	139 - 233	158 - 258	179 - 283	200 - 310	222 - 338	246 - 366	270 - 396		
19	114 - 241	163 - 266	183 - 293	205 - 320	228 - 348	252 - 377	277 - 407	303 - 433	
20	147 - 249	167 - 275	188 - 302	210 - 330	234 - 358	258 - 388	283 - 419	309 - 451	337 - 483

n_1 : tamaño de la muestra más pequeña, n_2 : tamaño de la muestra más grande.

Con un nivel α , la región crítica está dada por los valores de S exteriores al intervalo I_α indicadas por $P(S \notin I_\alpha) \leq \alpha$.

Ejemplo: para $n_1 = 5$ y $n_2 = 16$, se tiene $P(S \notin [30, 80]) \leq 0.05$.

Si se tiene $S \leq 30$ o $S \geq 80$, se rechaza la hipótesis al nivel 5%.

Tabla para la prueba de Mann-Whitney (nivel 1%)

$n_2 \ n_1$	2	3	4	5	6	7	8	9	10	11
5	-	-	-	15 - 40						
6	-	-	10 - 34	16 - 44	23 - 55					
7	-	-	10 - 38	16 - 49	24 - 55	32 - 73				
8	-	-	11 - 41	17 - 53	25 - 65	34 - 78	43 - 93			
9	-	6 - 33	11 - 45	18 - 57	26 - 70	35 - 84	45 - 99	56 - 115		
10	-	6 - 36	12 - 48	19 - 61	27 - 75	37 - 89	47 - 105	58 - 122	71 - 139	
11	-	7 - 38	13 - 51	20 - 65	28 - 80	38 - 95	49 - 111	61 - 128	73 - 147	87 - 166
12	-	7 - 41	13 - 55	21 - 69	30 - 84	40 - 100	51 - 117	63 - 135	76 - 154	90 - 174
13	-	7 - 44	14 - 58	22 - 73	31 - 89	41 - 106	53 - 123	65 - 142	79 - 161	93 - 182
14	-	7 - 47	14 - 62	22 - 78	32 - 94	43 - 111	54 - 130	67 - 149	81 - 169	96 - 190
15	-	8 - 49	15 - 65	23 - 82	33 - 99	44 - 117	56 - 136	69 - 156	84 - 176	99 - 198
16	-	8 - 52	15 - 69	24 - 86	34 - 104	46 - 122	58 - 142	72 - 162	86 - 184	102 - 206
17	-	8 - 55	16 - 72	25 - 90	36 - 108	47 - 128	60 - 148	74 - 169	89 - 191	105 - 214
18	3 - 39	9 - 57	17 - 75	26 - 94	37 - 113	49 - 133	62 - 154	76 - 176	92 - 198	108 - 222
19	3 - 41	9 - 60	17 - 79	27 - 98	38 - 118	50 - 139	64 - 160	78 - 183	94 - 206	111 - 230
20	3 - 43	9 - 63	18 - 82	28 - 102	39 - 123	52 - 144	66 - 166	81 - 189	97 - 213	114 - 238

$n_2 \ n_1$	12	13	14	15	16	17	18	19	20
12	105 - 195								
13	109 - 203	125 - 226							
14	112 - 212	129 - 235	147 - 259						
15	115 - 221	133 - 244	151 - 269	171 - 294					
16	119 - 229	136 - 254	155 - 279	175 - 305	196 - 332				
17	122 - 238	140 - 263	160 - 288	180 - 315	201 - 343	223 - 372			
18	125 - 247	144 - 272	164 - 298	184 - 326	206 - 354	228 - 384	252 - 414		
19	129 - 255	148 - 281	168 - 308	189 - 336	211 - 365	234 - 395	258 - 426	283 - 458	
20	132 - 264	152 - 290	172 - 318	193 - 347	216 - 376	239 - 407	263 - 439	289 - 471	315 - 505

Ejemplo: para $n_1 = 5$ y $n_2 = 16$, se tiene $P(S \notin [24, 86]) \leq 0.01$.

Si se tiene $S \leq 24$ o $S \geq 86$, se rechaza la hipótesis al nivel 1%.

Bibliografía

- [1] T.W. Anderson (1958) An introduction to multivariate statistical analysis *J. Wiley, New York.*
- [2] J. Barra (1970) Notions fondamentales de statistique mathématique *Dunod, Paris.*
- [3] J. Coursol (1980) Techniques statistiques des modèles lineaires *CIMPA.*
- [4] A. Fuchs–C. Fourgeaud (1972) Statistique *Dunod, Paris.*
- [5] F. A. Graybill (1961) An introduction to linear statistical models *MacGraw-Hill, New York.*
- [6] M. Loeve (1961) Probability theory *Van Nostrand C.*
- [7] M. Metivier (1968) Théorie des probabilités *Dunod, Paris.*
- [8] H. L. Royden (1963) Real analysis *MacMillan C, New York.*
- [9] H. Seal (1966) Multivariate statical analysis for biologists *Neuhrn and Co. Ltd, London.*
- [10] A. Wald (1949b) Note on the consistency of the maximun likelihood estimate
Ann. Math. Stat. Vol. 20 p.595-601.
- [11] S. R. Searle (1971) Linear models *J. Wiley, New York.*
- [12] D. L. Wallace (1958) Asymptotic approximations to distribution *Ann. Math. Stat. Vol. 29 p.635-654.*
- [13] S. Wilks (1963) Mathematical Statistics *J. Wiley, New York.*