# Metodos Estadísticos

Eloísa Díaz Francés
Universidad de Guanajuato
Enero – Junio 2012

Eugenio Daniel Flores Alatorre

## Bibliografía

[1] Kalbfleisch, J. G. (1985). Probability and Statistical Inference. Vol. 2. Springer-Verlag.

[2] Roussas, G. G. (1997). A Course in Mathematical Statistics. Academic Press.

[3] Mood, A. M., Graybill, A. F. y Boes, D. (1974). Introduction to the Theory of Statistics. Mc Graw Hill.

[4] Hogg, R. V. y Craig, A. T. (1978). Introduction to Mathematical Statistics. Collier Mac Millan Internacional Editions.

[5] Bhataccharyya y Johnson (2006). Statistics: Principles and Methods. John Wiley & Sons.

[6] Pawitan, Y. (2001). In All Likelihood. Statistical Modeling and Inference Using Likelihood. Oxford: Oxford Science Publications.

[7] Sagan, Carl (1995). El mundo y sus demonios. La ciencia como una luz en la oscuridad. México: Editorial Planeta.

[8] Sprott, D. A. (2000). Statistical Inference in Science. Springer-Verlag.

[9] Evans, M., Hastings, N. y Peacock, B. (1993). Statistical Distributions. John Wiley & Sons.

[10] Serfling, R. (1980). Approximation Theorems of Mathematical Statistics, Wiley.

[11] Kalbfleisch, J. G. (1985). Probability and Statistical Inference. Vol. 1. Springer-Verlag.

[12] Box, G. E. P. (1980). Sampling and Bayes' Inference in Scientific Modelling and Robustness. JRSS, Series A, V.143, No. 4. pp. 383-430.

## Objetivos del curso

- Dar una cultura general sobre el razonamiento inductivo-deductivo y conceptos básicos de la Inferencia Estadística, destacando el ámbito científico para analizar fenómenos naturales repetibles.
- Describir, discutir y aplicar a diversos problemas reales el proceso de modelación estadística: Sugerir, estimar y validar un modelo estadístico para describir adecuadamente un fenómeno repetible de interés, modificando el modelo si fuese necesario y repitiendo este proceso hasta "converger" al mejor modelo para el fenómeno, (Box, 1980).
- Presentar el problema general de estimación de parámetros de modelos estadísticos paramétricos, desarrollando para ello conceptos básicos relevantes de teoría estadística.
- Distinguir la diferencia entre las situaciones que requieren probar hipótesis y en las que se necesita estimar parámetros en el proceso de modelación estadística.
- Desarrollar las habilidades computacionales del estudiante para fortalecer el conocimiento aprendido en la gran mayoría de los temas tratados en el temario.

# Temario

1. Introducción a la modelación estadística y al razonamiento inductivo. Objetivos de la modelación estadística. Modelos probabilísticos y estadísticos, relación y diferencias; ejemplos. Modelos paramétricos y no paramétricos. Familias de distribuciones más importantes. La ciencia y la importancia de la Estadística en ella. [Box, 1980; Capítulo 1 Sprott; Capítulos 1-3 Sagan]

2. Conceptos fundamentales de estimación en la modelación estadística.

2.1. La función de verosimilitud completa de los parámetros de un modelo estadístico y sus propiedades: Invarianza. El estimador de máxima verosimilitud (EMV). La función de verosimilitud relativa como estandarización de la verosimilitud y su interpretación en términos de simetría, localización, dispersión y rango de valores del parámetro con plausibilidad. Regiones e intervalos de verosimilitud y su interpretación. Exploración analítica de la verosimlitud relativa a través de su expansión en series de Taylor alrededor del EMV. La información observada de Fisher, su relación con esta expansión y con determinar la curvatura de la verosimilitud. La información esperada de Fisher. Aproximaciones normales a la verosimilitud. [Capítulo 9 Kalbfleisch, Pawitan].

2.2 Otros métodos de estimación puntual. Estimadores de momentos. Estimador de mínima Ji-cuadrada (Mood, Graybill y Boes).

2.3 Suficiencia. [1 ½ semanas] Identificación de estadísticas suficientes a partir de la función de verosimilitud. El Teorema de Factorización de Fisher. Propiedades de estadísticas suficientes. Relación del concepto de suficiencia con la familia exponencial y con la familia de localización y escala. [Secciones 15.1 y 15.2, Kalbfleisch]

2.4 Consistencia. Definición e importancia. (Serfling).

2.5 Otras propiedades asintóticas importantes de estimadores: a) Insesgamiento [Kalbfleisch, Sección 11.7], b) Cota para la varianza de estimadores insesgados de Cramer-Rao. [Serfling, y Mood, Graybill y Boes], c) Propiedades asintóticas del EMV como estimador puntual: el Teorema de Máxima Verosimilitud (Serling), d) Propiedades asintóticas del estimador de momentos (Serfling).

2.6 Estimación por intervalos de un parámetro de interés.

2.6.1 Intervalos aleatorios y probabilidades de cobertura.

2.6.2 Intervalos de confianza para un parámetro. Cantidades pivotales y su uso para construir intervalos de confianza.

2.6.3 Intervalos de verosimilitud-confianza para el parámetro de interés. Comparación e importancia de las propiedades de verosimilitud y de confianza de un intervalo de estimación. [Capítulo 11 de Kalbfleisch].

2.6.4. Obtención de intervalos de confianza en el caso de muestras finitas, incluso pequeñas, a partir de cantidades pivotales asintóticas: a) a través de la aproximación normal a la verosimilitud, b) a través de la aproximación Ji-cuadrada para la distribución del negativo del doble de la log razón de verosimilitud.

2.7. Criterios de optimalidad para seleccionar estimadores; ventajas y desventajas del error cuadrático medio como criterio. Propiedades deseables de estimadores: insesgamiento asintótico, consistencia y eficiencia. Intervalos de estimación vs. estimadores puntuales.

2.8. Modelos estadísticos multiparamétricos. El problema de estimación por separado de parámetros de interés en presencia de parámetros de estorbo. La función de verosimilitud perfil. La función de verosimilitud condicional y marginal como otras posibilidades.

3. Introducción a pruebas de hipótesis y de significancia. Conceptos básicos. Uso adecuado de p-valores. El rol importante que juegan en la validación de modelos estadísticos [Capítulo 6, Sprott, Bhattacharyya y Johnson, Capítulo 6.]

4. Comparación y selección de modelos estadísticos paramétricos. La razón de verosimilitud como herramienta de comparación para modelos estadísticos anidados y no anidados. Pruebas de hipótesis asociadas en el caso de modelos anidados. El criterio de Akaike. Ejemplos. [Box, 1980; Capítulo 11 de Kalbfleisch; Pawitan, Capítulo 3].

# Proceso de modelar estadísticamente un fenómeno aleatorio

1) Fenómeno de interés con datos

Un fenómeno aleatorio obtenido de la naturaleza a partir de la observación y la experimentación. Se buscan explicaciones con la ciencia y la religión. Es importante tener los datos o sería imposible cualquier modelación.

Mientras que las matemáticas funcionan de manera deductiva, y la probabilidad también, la estadística, como parte de la matemática aplicada, funciona de manera inductiva.

2) Planteo de un modelo $f(x;\theta)$
3) Estimación del parámetro $\theta$
4) Validación del modelo
5) Comparación con otros modelos

**Modelo de probabilidad:** una variable aleatoria con distribución conocida. En un modelo de probabilidad conocemos todos los parámetros de una variable aleatoria, pero desconocemos los datos.

Se representa como una terna $(S,\Omega,P)$ donde el primer elemento es un conjunto de resultados posibles del experimento, el segundo es una sigma-álgebra que contiene los elementos del primer elemento de la terna y el tercero es una función de probabilidad

$$P:\Omega\rightarrow[0,1]$$

donde a cada elemento le asocia su probabilidad. Tradicionalmente, asociamos esta función con la función de distribución de una variable aleatoria:

$$F_x: X\rightarrow\mathbb{R}$$

donde $X$ es una variable aleatoria con los parámetros completamente especificados $f(X:\theta)$. Es decir, todas las probabilidades de los eventos son conocidas –o para conocerlas sólo hace falta usar las funciones de distribución o densidad.

**Ejemplo:**

Una variable aleatoria Poisson con parámetro $\lambda=3$. Si nos preguntamos por $P[X=0]=e^{-\lambda}$, pues sabemos que en una variable aleatoria que distribuye como una Poisson,

$$P[X=x;\lambda]=\frac{\lambda^x e^{-\lambda}}{x!}, x\in\{0,1,\dots\}$$

**Modelo estadístico:** conjunto de modelos de probabilidad. En un modelo estadístico desconocemos los parámetros y la distribución que más adecuadamente pueda describir un fenómeno del cual tenemos un conjunto de datos.

**Ejemplo:**

Una variable aleatoria con distribución Normal y parámetros desconocidos.

## Clasificación de variables

Existen numerosas maneras de clasificar las variables aleatorias según si distribución. Estas clasificaciones se conocen como *familias de distribuciones*. Dos de las más grandes son las siguientes:

| Discretas | Continuas |
|---|---|
| Poisson | Normal |
| Geométrica | Exponencial |
| Bernoulli | Gama |
| Hipergeométrica | Uniforme |
| Multinomial | Ji-cuadrada |
| Binomial Negativa | LogNormal |
| Uniforme | t-Student |
| | Weibull |
| | Gumbell |
| | Fréchet |
| | F-Fisher |
| | Rayleigh |
| | Maxwell |
| | Beta |
| | Logística |
| | LogF |
| | Pareto |
| | Cauchy |

En cuya clasificación caben todas las variables aleatorias conocidas.

Antes de continuar con la clasificación, introducimos un teorema importantísimo para poder determinar si una variable pertenece o no a una familia dada.

## Teorema de cambio de variable

Sea X una variable aleatoria continua con función de densidad $f_X(x;\theta)$. Interesa encontrar la densidad de $Y = g(X)$, donde $g$ es una función continua e invertible, de tal manera que $X = g^{-1}(Y)$. Entonces

$$h_Y(y;\theta) = f_X(g^{-1}(y);\theta)\left|\frac{dx}{dy}\right|$$

## Quantiles de una distribución

Definimos la Función de Quantiles (Inversa Generalizada) de la siguiente manera:

$$Q_\alpha = Q(\alpha) = \inf\{\mathrm{x}|\mathrm{F}(\mathrm{x}) \geq \alpha\}$$

que funciona adecuadamente tanto para distribuciones continuas como para discretas.

## Familias de distribuciones

Las principales familias de distribuciones, además de las continuas y discretas, son:

1) **Familia Exponencial de *k* parámetros** $\theta = (\theta_1, \dots, \theta_k)$

Se dice que una variable aleatoria pertenece a esta familia si su función de probabilidad o de densidad se puede factorizar de la siguiente manera:

$$f(x;\theta) = A(\theta)B(x)\exp\left[\sum_{i=1}^{k} C_i(\theta)D_i(x)\right]$$

donde la función A$(\theta)$ es función únicamente de los parámetros, no necesariamente de los *k*, la función $B(x)$ es función únicamente de la variable y ambas son no negativas.

**Ejemplo 1:** La distribución normal.

Consideremos una variable con distribución normal $N(\mu,\sigma)$. Recordando, si $X{\sim}N(\mu,\sigma)$, entonces

$$f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

Dado que

$$\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] = \exp\left[-\frac{x^2}{2\sigma^2}+\frac{2x\mu}{2\sigma^2}-\frac{\mu^2}{2\sigma^2}\right] = \exp\left(\frac{-\mu^2}{2\sigma^2}\right)\exp\left[-\frac{1}{2\sigma^2}x^2+\frac{\mu}{\sigma^2}x\right]$$

Proponemos la siguiente factorización:

$$\left[\frac{1}{\sigma}\exp\left(\frac{-\mu^2}{2\sigma^2}\right)\right]\left[\frac{1}{\sqrt{2\pi}}\right]\exp\left[-\frac{1}{2\sigma^2}x^2+\frac{\mu}{\sigma^2}x\right]$$

Donde, definiendo

$$A(\theta) = \frac{1}{\sigma}\exp\left(\frac{-\mu^2}{2\sigma^2}\right)$$

$$B(x) = \frac{1}{\sqrt{2\pi}}$$

$$C_1 = -\frac{1}{2\sigma^2}$$
$$D_1 = x^2$$
$$C_2 = \frac{\mu}{\sigma^2}$$
$$D_2 = x$$

concluimos que la distribución normal pertenece a la familia exponencial.

**Ejemplo 2:** La distribución Poisson.

Consideremos una variable aleatoria con distribución Poisson$(\lambda)$. Recordando, si $X \sim Poisson(\lambda)$ entonces

$$f(x;\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$$

Proponemos la siguiente factorización

$$\left[e^{-\lambda}\right]\left[\frac{1}{x!}\right]\exp[x \ln \lambda]$$

donde

$$A(\theta) = e^{-\lambda}$$
$$B(x) = \frac{1}{x!}$$
$$C_1 = \ln \lambda$$
$$D_1 = x$$

que nos permite concluir que la distribución Poisson también pertenece a la familia exponencial.

**Ejemplo 3:** La distribución Binomial $(N,p)$ con N conocida.

Recordamos la función de probabilidad de la distribución Binomial y la factorizamos de tal manera que podamos concluir que pertenece a la familia exponencial

$$f(x;p) = \binom{N}{x} p^x (1-p)^{N-x} = \binom{N}{x}\left(\frac{p}{1-p}\right)^x (1-p)^N = \binom{N}{x}\exp\left(x\, ln\left(\frac{p}{1-p}\right)\right)(1-p)^N$$

Observamos que si ambos parámetros N y p fueran desconocidos, entonces no sería posible encontrar una factorización que nos permita incluirla dentro de esta familia, pues

$$\binom{N}{x} = \frac{N!}{x!\,(N-x)!}$$

no se puede separar en funciones multiplicativas donde cada una dependa sólo del argumento o sólo de los parámetros.

## 2) Localización y escala

Tiene exclusivamente dos parámetros, uno de localización y otro de escala. Los denotamos por $(\mu, \sigma)$ respectivamente.

Reconocemos que una distribución pertenece a esta familia si

a) la densidad se puede escribir

$$f(x;\mu,\sigma) = \frac{1}{\sigma} g\left(\frac{x-\mu}{\sigma}\right)$$

donde $g$ es el miembro estándar de la densidad, y su densidad no depende ni de $\mu$ ni de $\sigma$.

b) si $Y = \frac{X-\mu}{\sigma}$ tiene densidad que no depende de ninguno de los parámetros, entonces X pertenece a la familia de localización y escala.

**Ejemplo 1:** La distribución Normal.

Puesto que $Y = \frac{X-\mu}{\sigma}$, tenemos que $X = \sigma Y + \mu$. Usando el Teorema de Cambio de Variable, conociendo la densidad de la distribución Normal, tenemos que

$$h(y;\mu,\sigma) = f(\sigma y + \mu;\mu,\sigma)|\sigma|$$

y, para concluir que la distribución normal pertenece a esta familia, queremos ver que esa función de densidad no depende de los parámetros de localización o escala.

$$f(\sigma y + \mu;\mu,\sigma)|\sigma| = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{\sigma y + \mu - \mu}{\sigma}\right)^2\right]\sigma$$
$$= \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}y^2\right]\sigma = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}y^2\right]$$

que claramente no depende de los parámetros de localización o escala y, por lo tanto, nos permite concluir que la distribución normal pertenece a la familia de localización y escala.

**Ejemplo 2:** La distribución Lognormal.

Recordemos que una variable aleatoria Z tiene distribucíon Lognormal si su logaritmo distribuye como una Normal con parámetros $\mu, \sigma$. Es decir, si $Z \sim Lognormal$, entonces $W = \log Z \sim N(\mu, \sigma)$, o bien, $Z = \exp{(W)}$.

Usando el Teorema de Cambio de Variable,

$$h(z;\mu,\sigma) = f(\log(z);\mu,\sigma)\left|\frac{dw}{dz}\right| = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{log(z) - \mu}{\sigma}\right)^2\right]\frac{1}{z}$$

que claramente depende de los parámetros y que no es posible simplificar de alguna manera que no dependa.

### 3) Valores extremos

Una aplicación importante del teorema de convergencia aparece en el estudio de las posibles distribuciones límite para máximos normalizados de variables aleatorias independientes, idénticamente distribuidas. Supongamos que $\{X_n, n \geq 1\}$ es una sucesión independiente, idénticamente distribuida con función de distribución común $F$ y sea $M_n = \max\{X_i \ : \ 1 \leq i \leq n\}$.

Supongamos que existen constantes $a_n > 0$ y $b_n \in \mathbb{R}$ para $n \geq 1$ tales que

$$\mathbb{P}\left[\frac{M_n - b_n}{a_n} \leq x\right] = F^n(a_n x + b_n) \to G(x)$$

débilmente cuando $n \to \infty$ donde $G$ es propia ($G(\mathbb{R}) = 1$) y no está concentrada en un punto. Entonces $G$ pertenece a alguna de las siguientes tres familias de distribuciones:

(a) Tipo I: $\Lambda(x) = \exp(-e^{-x})$, $x \in \mathbb{R}$.

(b) Tipo II: $\Phi_\alpha(x) = \begin{cases} 0 & x < 0 \\ \exp(-x^{-\alpha}) & x \geq \alpha \end{cases}$, para algún $\alpha > 0$.

(c) Tipo III: $\psi_\alpha(x) = \begin{cases} \exp(-(-x)^{\alpha}) & x < 0 \\ 1 & x \geq 0 \end{cases}$, para algún $\alpha > 0$.

Este teorema fue originalmente propuesto por Fisher y Tippet en 1928 y demostrado rigurosamente por Gnedenko en 1943. La distribución (a) es la distribución de Gumbel, (b) es la distribución de Fréchet y (c) es la distribución de Weibull.

### 4) Transformaciones de Box y Cox

La familia de transformaciones más utilizada para resolver los problemas de falta de normalidad y de heterocedestacidad es la familia Box-Cox, cuya definición es la siguiente.

Se desea transformar la variable $Y$, cuyos valores muestrales se suponen positivos, en caso contrario se suma una cantidad fija $M$ tal que $Y + M > 0$. La transformación de Box-Cox depende de un parámetro $\lambda$ por determinar y viene dada por

$$Z(\lambda) = \begin{cases} \dfrac{y^\lambda - 1}{\lambda} & si\ \lambda \neq 0 \\ \ln y & si\ \lambda = 0 \end{cases}$$

donde $Y \sim N(\mu, \sigma^2)$ y $g(Y;\ \lambda, \mu, \sigma)$.

Si se quieren transformar los datos para conseguir normalidad, el mejor método para estimar el parámetro es el de máxima verosimilitud.

5) Simétricas (Densidades alrededor de la moda)

Distribución Gama con parámetro $\alpha > 1$

6) Asimétricas

Distribución Gama con parámetro $\alpha \leq 1$

## Teorema Integral de probabilidad

Si X es una variable aleatoria con función de distribución continua $F_X(x)$ entonces $U = F_X(x)$ se distribuye uniformemente en el intervalo $(0,1)$. Inversamente, si $U$ se distribuye uniformemente en el intervalo $(0,1)$ entonces $X = F_X^{-1}(U)$ tiene función de distribución $F_X(x)$.

**Demostración**

Ida:

$$P[U \leq u] = P[F_X(x) \leq u] = P[X \leq F_X^{-1}(u)] = F_X\left(F_X^{-1}(u)\right) = u$$

Regreso:

$$P[X \leq x] = P[F_X^{-1}(U) \leq x] = P[U \leq F_X(x)] = F_X(x)$$

**Ejemplo:**

Utilizando el teorema integral de probabilidad, describe el procedimiento para simular

a) $f(x;\theta) = \frac{1}{\theta}e^{-\frac{x}{\theta}},\ \theta > 0,\ x \geq 0$ (exponencial)

Recordando que

$$F(x,\theta) = \int_0^x \frac{1}{\theta}e^{-\frac{y}{\theta}}dy = 1 - e^{-\frac{x}{\theta}}$$

definimos

$$U = 1 - e^{-\frac{x}{\theta}} \Rightarrow e^{-\frac{x}{\theta}} = 1 - U \Rightarrow -\frac{x}{\theta} = \ln(1-U) \Rightarrow x = -\theta \ln(1-U)$$

b) $F_X(x) = 1 - \exp\left(-\exp\left(\frac{x-\mu}{\sigma}\right)\right)$ (gumbel)

## Función Generadora de Momentos

La función generadora de momentos o función generatriz de momentos de una variable aleatoria $X$ es:

$$M_X(t) := E(e^{tX}),\ \ t \in \mathbb{R}$$

siempre que esa esperanza exista.

La función generadora de momentos se llama así porque, si existe un entorno de $t = 0$, permite generar los momentos de la distribución de probabilidad:

$$E(X^n) = M_X^{(n)}(0) = \frac{d^n M_X}{dt^n}(0).$$

## Función de distribución empírica

Sean $X_1, \ldots, X_n$ variables aleatorias independientes idénticamente distribuidas. Entonces

$$F_n(X) = \sum_{i=1}^{n} \mathbb{I}_{(-\infty,x]}(x_i) = \frac{\# x_i \leq x}{n}$$

además $\mathbb{I}_{(-\infty,x]}(x_i) \sim Bernoulli(p)$. Luego, $F_n(x)$ es un promedio de variables aleatorias con distribución $Bernoulli(p)$ independientes idénticamente distribuidas donde $p = F_X(x)$, así, por la Ley fuerte de los grandes números:

$$\mathbb{P}\left[\lim_{n\to\infty} F_n(x) = F_X(x)\right] = 1$$

o bien,

$$F_n(x) \xrightarrow{c.p.1} F_X(x)$$

De manera habitual trabajaremos con

$$\alpha_i' s \coloneqq \frac{1}{n+1}, \frac{2}{n+2}, \ldots, \frac{i}{n+i}, \ldots, \frac{n}{n+1}$$

para que

$$0 < \frac{i}{n+i} < 1$$

**Ejemplo:**

Gumbel de máximos, $X \sim Gum(\mu, \sigma)$.

$$F(x) = \exp\left[-exp\left(-\frac{x-\mu}{\sigma}\right)\right]$$

entonces

$$Q_\alpha = F^{-1}(\alpha) = \mu - \sigma \ln(-ln(\alpha)).$$

Veamos que podemos calcular los cuantiles de la siguiente manera:

$$\alpha = F(x;\ \mu, \sigma) = G\left(\frac{x-\mu}{\sigma}\right) \Rightarrow G^{-1}(\alpha) = \frac{x-\mu}{\sigma}$$
$$\Rightarrow Q_\alpha = X = \mu + \sigma G^{-1}(\alpha)$$

donde $\alpha \sim U(0,1)$ y $G$ es nuestra variable con distribución estándar.

En ocasiones alguna variable de localización y escala podemos reescribir la distribución en términos de un cuantil y el parámetro de escala:

$$(\mu, \sigma) \leftrightarrow (Q_\alpha, \sigma)$$

y se logra gracias al resultado anterior:

$$Q_\alpha = \mu + k\sigma \Longrightarrow \mu = Q_\alpha - k\sigma$$

# Modos de convergencia

## Convergencia casi segura

Se dice que la sucesión $\{x_n\}_{n\geq 1}$ converge casi seguramente o con probabilidad 1 o fuertemente a $X$, denotando esto por

$$\{x_n\}_{n\geq 1} \underset{n \to \infty}{\overset{c.p.1}{\longrightarrow}} X$$

si existe un evento $A$ de medida cero $(P(A) = 0)$ tal que para cada $s \in A^c \subset \mathcal{S}$ se tiene que

$$\lim_{n\to\infty} X_n(s) = X(s);$$

$$\mathbb{P}\left[\left\{s \middle| \lim_{n\to\infty} X_n(s) = X(s)\right\}\right] = 1;$$

$$P(A^c) = 1.$$

## Convergencia en probabilidad

Se dice que la sucesión $\{x_n\}_{n\geq 1}$ converge en probabilidad a $X$, denotando esto por

$$\{x_n\}_{n\geq 1} \overset{P}{\to} X$$

si para $\varepsilon > 0$ se tiene que

$$\lim_{n\to\infty} \mathbb{P}[\{s||x_n(s) - x(s)| < \varepsilon\}] = 1$$

o, equivalentemente,

$$\lim_{n\to\infty} \mathbb{P}[\{s||x_n(s) - x(s)| > \varepsilon\}] = 0$$

## Convergencia en distribución

Se dice que la sucesión $\{x_n\}_{n\geq 1}$ converge en distribución o en ley o débilmente, denotado por

$$\{x_n\}_{n\geq 1} \underset{n \to \infty}{\overset{d}{\longrightarrow}} X$$

o, equivalentemente,

$$F_n \underset{n \to \infty}{\overset{d}{\longrightarrow}} F_x$$

si se tiene que

$$\lim_{n\to\infty} F_n(x) = F_x(x)$$

para toda $x$ que sea punto de continuidad de $F_x$.

**Ley fuerte de los grandes números**

Sea $\{X_n\}_{n\geq 1}$ una sucesión de variables aleatorias independientes idénticamente distribuidas con media $\mathbb{E}(X_n) = \mu$ y considérese la sucesión de los promedios aritméticos $\bar{X}_n = \frac{X_1+\cdots+X_n}{n}$, entonces se cumple

$$\{\overline{X_n}\}_{n\geq 1} \xrightarrow{c.p.1} \mu$$

**Ley débil de los grandes números**

Sea $\{X_n\}_{n\geq 1}$ una sucesión de variables aleatorias independientes idénticamente distribuidas con media $\mathbb{E}(X_n) = \mu$ y considérese la sucesión de los promedios aritméticos $\bar{X}_n = \frac{X_1+\cdots+X_n}{n}$, entonces se cumple

$$\bar{X}_n \xrightarrow{P} \mu$$

**Demostración:** P.D. $\lim_{n\to\infty} \mathbb{P}[|\bar{X}_n - \mu| \geq \varepsilon] = 0$

Por el **Teorema Chebyshev**,

$$\mathbb{P}[|x-\mu| \geq \varepsilon] \leq \frac{\mathbb{E}(x-\mu)^2}{\varepsilon^2} = \frac{var(X)}{\varepsilon^2}$$

aplicado para $\bar{X}_n$:

$$0 \leq \mathbb{P}[|\bar{X}_n - \mu| \geq \varepsilon] \leq \frac{var(\bar{X}_n)}{\varepsilon} = \frac{\sigma^2}{n\varepsilon}$$

pues

$$var(\bar{X}_n) = \frac{1}{n^2}\sum_{i=1}^{n} var(X_i) = \frac{\sigma^2}{n}$$

entonces

$$\lim_{n\to\infty} 0 \leq \lim_{n\to\infty} \mathbb{P}[|\bar{X}_n - \mu| \geq \varepsilon] \leq \lim_{n\to\infty} \frac{\sigma^2}{n\varepsilon} = 0$$

## Teorema de Glivenko-Cantelli

El Teorema de Glivenko-Cantelli es consecuencia de la Ley fuerte de los grandes números.

Sea $\{X_n\}$ una sucesión de variables aleatorias independientes, idénticamente distribuidas. Luego la función de distribución empírica de los primeros $n$ términos converge casi seguro a la función de distribución de $X_1$.

Demostración: Por la ley fuerte de los grandes números, es obvio que para cada $x$ tenemos la convergencia casi segura:

$$\lim_{n\to\infty} sup\left|F_n(x_j) - F_{X_1}(x_j)\right| = 0.$$

Es fácil ver, usando la monotonía de $F_{X_1}$ y $F_n$ que esto implica el teorema.

$$sup_{x\in\mathbb{R}}|F_n(x) - F_X(x)| \xrightarrow{c.p.1} 0$$

## Función Indicadora

Definimos la función indicadora de un conjunto como

$$\mathbb{I}_A(x) = \begin{Bmatrix} 1 & si\ x \in A \\ 0 & si\ x \notin A \end{Bmatrix}$$

En la mayoría de los casos, las densidades o funciones de probabilidad de una distribución pueden expresarse en términos de funciones indicadoras o, al menos, pueden incluirlas como un factor, para asegurar que los parámetros cumplen las condiciones fijadas de inicio.

## Invarianza

Hay dos aspectos de invarianza necesarios: mediciones y parámetros.

## Cantidad Pivotal

Sea $X$ una variable aleatoria cuya distribución depende de un parámetro desconocido $\theta$ y sea $X_1, \dots, X_n$ una muestra aleatoria de $X$.

La cantidad pivotal –o simplemente, el pivote- es una función $Q(X_1, X_2, \dots, X_n;\ \theta)$ de la muestra aleatoria y de un parámetro de interés que tiene distribución completamente especificada que no depende de $\theta$ ni de ningún otro parámetro desconocido.

Si $Q(X_1, \dots, X_n;\ \theta)$ es una función monótona de $\theta$, se obtiene un intervalo de confianza para $\theta$, encontrando dos valores $K_1, K_2$ tales que, fijando un nivel de confianza de $1 - \alpha$,

$$\mathbb{P}(K_1 \leq Q(X_1, \dots, X_n;\ \theta) \leq K_2) = 1 - \alpha$$

donde $K_1, K_2$ se obtienen a partir de la distribución del pivote, que es conocida.

Una vez determinados los valores de $K_1, K_2$, la expresión del intervalo se obtiene despejando $\theta$ de la ecuación resultante.

## Métodos de estimación de parámetros

Para $X_1, \dots, X_n$ variables aleatorias independientes, idénticamente distribuidas con $F(x;\ \theta)$ y $\theta$ desconocido, existen varios métodos para estimar el parámetro.

1) Método de momentos (propuesto por Karl Pearson).
2) Método de mínima Ji-cuadrada (propuesto por Karl Pearson).
3) Método de máxima verosimilitud (propuesto por Ronald Fisher).

### Método de momentos

Para estimar el parámetro $\theta$ donde $\dim(\theta) = k$, se igualan los primeros momentos (centrales o no centrales) empíricos con los teóricos.

**Momentos Empíricos**:

No centrales:

$$M_r = \frac{1}{n}\sum_{i=1}^{n} x_i^r$$

para $r = 1, 2, \dots$

Centrales:

$$MC_r = \frac{1}{n}\sum_{i=1}^{n}[(X_i - \bar{X})^n]$$

**Momentos Teóricos**:

No centrales:

$$\mathbb{E}(X^r)$$

para $r = 1, 2, \dots$

Centrales:

$$\mathbb{E}\left[\left(X - \mathbb{E}(X)\right)^r\right]$$

### Método de máxima verosimilitud

Sea $X_1, \dots, X_n$ un conjunto de variables aleatorias independientes, idénticamente distribuidas como $f(x;\theta)$ y sean $x_1, \dots x_n$ una realización. La función de máxima verosimilitud se define proporcional a la probabilidad conjunta de las variables aleatorias observadas.

1) Caso discreto:

$$\mathcal{L}(\theta; x_1, \dots, x_n) = c(x_1, \dots x_n)\mathbb{P}[X_1 = x_1, \dots, X_n = x_n; \theta]$$

donde $c(x_1, \dots x_n) \geq 0$.

2) Caso continuo:

$$\mathcal{L}(\theta; x_1, \dots, x_n) = c(x_1, \dots x_n)\prod_{i=1}^{n}[F(x_i + h; \theta) - F(x_i - h; \theta)]$$

donde $2h$ es la precisión del instrumento de medición.

Generalmente, usaremos para el caso continuo la aproximación continua a la función de verosimilitud, es decir:

$$\mathcal{L}(\theta; x_1, \dots, x_n) = c\prod_{i=1}^{n}[F(x_i; \theta)]$$

donde $c$ es una función constante con respecto al parámetro. Normalmente escogemos esta función de modo que nos quede la verosimilitud en función únicamente del parámetro de interés –en la medida de lo posible.

Será común utilizar la función de log verosimilitud:

$$l(\theta; x_1, \dots, x_n) = \ln \mathcal{L}(\theta; x_1, \dots, x_n)$$

Tanto para la función verosimilitud como para la log verosimilitud, la elección de la función constante $c$ no afecta el punto de maximización. Existen casos en los que es posible maximizar cualquiera de estas dos funciones analíticamente sin problemas. Cuando no, consideramos la función marcador (score) definida como:

$$S(\theta) = \frac{\partial l(\theta)}{\partial \theta}$$

y encontramos el máximo despejando $\theta$ de $S(\theta) = 0$. Es necesario verificar que sea un máximo e, idealmente, también que sea un máximo global. Para ello, consideramos la función $I$ de la información de Fisher como

$$I(\theta) = -\frac{\partial^2 l(\theta)}{\partial^2 \theta}$$

menos la segunda derivada con respecto al parámetro. Habremos encontrado un máximo cuando $I(\theta) \geq 0$.

## Teorema de Sluisky

Sea $\{X_n\}_{n\geq 1}$ una muestra de variables aleatorias independientes, idénticamente distribuidas y $g$ una función continua. Si

$$\{X_n\}_{n\geq 1} \xrightarrow{d} X$$

entonces

$$\{g(X_n)\}_{n\geq 1} \xrightarrow{d} g(X)$$

es decir, convergen en distribución.

## Propiedades deseables de un estimador

1) Consistencia.
2) Invarianza frente a reparametrizaciones uno a uno de $\theta$ y frente a las unidades de medición de la muestra.
3) Insesgamiento asintótico sobre insesgamiento para $n$ fija:

$$\lim_{n\to\infty} \mathbb{E}(\tilde{\theta}_n) = \theta_0$$

$$\mathbb{E}(\tilde{\theta}_n) = \theta_0$$

### Consistencia

Sea $X_1, \dots, X_n$ una muestra de variables aleatorias independientes, idénticamente distribuidas como $F_X(x;\theta)$ con $\theta$ desconocida. Sea $\widetilde{\theta_n} = g(X_1, \dots, X_n)$ un estimador de $\theta$. Se dice que

1) $\widetilde{\theta_n}$ es consistente fuerte si

$$\left\{\widetilde{\theta_n}\right\}_{n\geq 1} \xrightarrow{c.p.1} \theta$$

2) $\widetilde{\theta_n}$ es consistente débil (o estadístico o simple) si

$$\left\{\widetilde{\theta_n}\right\}_{n\geq 1} \xrightarrow{p} \theta$$

3) $\widetilde{\theta_n}$ es consistente en media $r$-ésima si

$$\left\{\widetilde{\theta_n}\right\}_{n\geq 1} \xrightarrow{r} \theta$$

# Tareas

## Tarea 1

**3.** Encuentra la relación que hay entre las siguientes dos variables aleatorias normales,

$$X \sim N(\mu, \sigma^2) \qquad Y \sim N(0, 1)$$

Queremos probar que $Y = \frac{X-\mu}{\sigma}$. Es decir, $X = \sigma Y + \mu$. Usaremos el Teorema del Cambio de Variable. Conocemos

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

y usando el Teorema del Cambio de Variable, suponiendo que $Y = \frac{X-\mu}{\sigma}$

$$h(y; \mu, \sigma) = f(\sigma y + \mu; \mu, \sigma)|\sigma|$$

$$f(\sigma y + \mu; \mu, \sigma)|\sigma| = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{\sigma y + \mu - \mu}{\sigma}\right)^2\right]\sigma$$
$$= \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}y^2\right]\sigma = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}y^2\right]$$

que es la distribución de $Y \sim N(0, 1)$.

**4.** Considera la distribución Gumbel. Da una expresión para el cuantil $Q_\alpha$ de probabilidad $\alpha$. En particular, encuentra el cuantil de probabilidad 0.25 para una Gumbel con parámetro de localización 5 y de escala unitaria.

Recordemos que si $X \sim Gum(\mu, \sigma)$, entonces

$$F(x) = \exp\left[-exp\left(-\frac{x-\mu}{\sigma}\right)\right]$$

de donde podemos concluir, aplicando repetidamente la respectiva función inversa, que

$$Q_\alpha = F^{-1}(\alpha) = \mu - \sigma \ln(-ln(\alpha)).$$

Ahora, si $\mu = 5$ y $\sigma = 1$ y buscamos $\alpha = {}^{3}/_{4}$, tenemos:

$$Q_{.25} = F^{-1}(.25) = 5 - \ln\left(-ln\left(\frac{3}{4}\right)\right) = 5 - \ln\left(\ln\left(\frac{4}{3}\right)\right) \approx 6.2459$$

**5.** Si $X$ es una exponencial con parámetro de escala $\theta$, encuentra cuál es la densidad de $Y = X/\theta$. Nota que esta densidad es también una Gama e identifica cuáles son sus parámetros. Encuentra cuál es la distribución de $Z = 2Y = 2X/\theta$.

Recordemos que si $X \sim Exp(\theta)$, entonces

$$F(x;\ \theta) = \theta e^{\theta(-x)}.$$

Así que si $Y = X/\theta$, usando el Teorema del Cambio de Variable,

$$h(y;\ \theta) = F\left(\frac{x}{\theta};\ \theta\right)\left|\frac{1}{\theta}\right| = \theta e^{\theta\left(-\frac{x}{\theta}\right)}\frac{1}{\theta} = e^{-x}$$

que distribuye igual que una variable aleatoria Gama, $G(1,1)$.

Análogamente,

$$g\left(\frac{y}{2}, \theta\right)\left|\frac{1}{2}\right| = \frac{1}{2}e^{-\frac{x}{2}}$$

que es una ji cuadrada con 1 grado de libertad.

**6.** Si $W$es una variable Poisson con parámetro de intensidad 50 y $Z = W/10$, ¿cuál es la probabilidad de cero conteos, $P[Z = 0]$?

Recordando, si $X \sim Poisson(\lambda)$ entonces

$$f(x; \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$$

Puesto que $\lambda = 50$ y sustituyendo por $x = 10y$, tenemos

$$\frac{50^{10y}e^{-50}}{(10y)!}(10)$$

que para $P[Y = 0]$, nos queda

$$10e^{-50} \approx 1.9287 \times 10^{-21}.$$

# TAREA 2

# TAREA 3

# GUÍA PRIMER PARCIAL

## 1. La diferencia entre un modelo estadístico y uno probabilístico.

Modelo de probabilidad: una variable aleatoria con distribución conocida. En un modelo de probabilidad conocemos todos los parámetros de una variable aleatoria, pero desconocemos los datos. Ejemplo: Una variable aleatoria Poisson con parámetro $\lambda = 3$. Si nos preguntamos por $P[X = 0] = e^{-\lambda}$, pues sabemos que en una variable aleatoria que distribuye como una Poisson,

$$P[X = x; \lambda] = \frac{\lambda^x e^{-\lambda}}{x!}, x \in \{0,1, \dots\}$$

Modelo estadístico: conjunto de modelos de probabilidad. En un modelo estadístico desconocemos los parámetros y la distribución que más adecuadamente pueda describir un fenómeno del cual tenemos un conjunto de datos. Ejemplo: Una variable aleatoria con distribución Normal y parámetros desconocidos.

## 2. Los pasos del proceso de modelar estadísticamente un modelo aleatorio de interés.

1) Fenómeno de interés con datos
2) Planteo de un modelo $f(x; \theta)$
3) Estimación del parámetro $\theta$
4) Validación del modelo
5) Comparación con otros modelos

## 3. Las familias de distribuciones principales y su definición.

1) Familia Exponencial de *k* parámetros $\theta = (\theta_1, \dots, \theta_k)$

Se dice que una variable aleatoria pertenece a esta familia si su función de probabilidad o de densidad se puede factorizar de la siguiente manera:

$$f(x; \theta) = A(\theta)B(x) \exp\left[\sum_{i=1}^{k} C_i(\theta)D_i(x)\right]$$

donde la función A$(\theta)$ es función únicamente de los parámetros, no necesariamente de los *k*, la función $B(x)$ es función únicamente de la variable y ambas son no negativas.

**Ejemplo 1:** La distribución normal.

2) Localización y escala

Tiene exclusivamente dos parámetros, uno de localización y otro de escala. Los denotamos por $(\mu, \sigma)$ respectivamente.

Reconocemos que una distribución pertenece a esta familia si

c) la densidad se puede escribir

$$f(x;\mu,\sigma) = \frac{1}{\sigma} g\left(\frac{x-\mu}{\sigma}\right)$$

donde $g$ es el miembro estándar de la densidad, y su densidad no depende ni de $\mu$ ni de $\sigma$.

d) si $Y = \frac{X-\mu}{\sigma}$ tiene densidad que no depende de ninguno de los parámetros, entonces X pertenece a la familia de localización y escala.

**Ejemplo 1:** La distribución Normal.

3) Valores extremos

Una aplicación importante del teorema de convergencia aparece en el estudio de las posibles distribuciones límite para máximos normalizados de variables aleatorias independientes, idénticamente distribuidas. Supongamos que $\{X_n, n \geq 1\}$ es una sucesión independiente, idénticamente distribuida con función de distribución común $F$ y sea $M_n = \max\{X_i \ : \ 1 \leq i \leq n\}$.

Supongamos que existen constantes $a_n > 0$ y $b_n \in \mathbb{R}$ para $n \geq 1$ tales que

$$\mathbb{P}\left[\frac{M_n - b_n}{a_n} \leq x\right] = F^n(a_n x + b_n) \to G(x)$$

débilmente cuando $n \to \infty$ donde $G$ es propia ($G(\mathbb{R}) = 1$) y no está concentrada en un punto. Entonces $G$ pertenece a alguna de las siguientes tres familias de distribuciones:

(a) Tipo I: $\Lambda(x) = \exp(-e^{-x})$, $x \in \mathbb{R}$.

(b) Tipo II: $\Phi_\alpha(x) = \begin{Bmatrix} 0 & x < 0 \\ \exp(-x^{-\alpha}) & x \geq \alpha \end{Bmatrix}$, para algún $\alpha > 0$.

(c) Tipo III: $\psi_\alpha(x) = \begin{Bmatrix} \exp(-(-x)^{\alpha}) & x < 0 \\ 1 & x \geq 0 \end{Bmatrix}$, para algún $\alpha > 0$.

Este teorema fue originalmente propuesto por Fisher y Tippet en 1928 y demostrado rigurosamente por Gnedenko en 1943. La distribución (a) es la distribución de Gumbel, (b) es la distribución de Fréchet y (c) es la distribución de Weibull.

4) Transformaciones de Box y Cox

La familia de transformaciones más utilizada para resolver los problemas de falta de normalidad y de heterocedestacidad es la familia Box-Cox, cuya definición es la siguiente.

Se desea transformar la variable $Y$, cuyos valores muestrales se suponen positivos, en caso contrario se suma una cantidad fija $M$ tal que $Y + M > 0$. La transformación de Box-Cox depende de un parámetro $\lambda$ por determinar y viene dada por

$$Z(\lambda) = \begin{cases} \dfrac{y^{\lambda} - 1}{\lambda} & si\ \lambda \neq 0 \\ \ln y & si\ \lambda = 0 \end{cases}$$

donde $Y \sim N(\mu, \sigma^2)$ y $g(Y;\ \lambda, \mu, \sigma)$.

Si se quieren transformar los datos para conseguir normalidad, el mejor método para estimar el parámetro es el de máxima verosimilitud.

5) Simétricas (Densidades alrededor de la moda)

Distribución Gama con parámetro $\alpha > 1$

6) Asimétricas

Distribución Gama con parámetro $\alpha \leq 1$

## 4. Relaciones entre algunas distribuciones.

## 5. La definición de cuantiles para variables discretas y continuas.

Definimos la Función de Quantiles (Inversa Generalizada) de la siguiente manera:

$$Q_{\alpha} = Q(\alpha) = \inf\{\mathrm{x} | \mathrm{F}(\mathrm{x}) \geq \alpha\}$$

que funciona adecuadamente tanto para distribuciones continuas como para discretas.

## 6. La función de distribución empírica y su gráfica y la relación que guarda con los cuantiles empíricos. Mostrar que los momentos empíricos son los momentos de esta distribución.

$$F_n(X) = \sum_{i=1}^{n} \mathbb{I}_{(-\infty, x]}(x_i) = \frac{\# x_i \leq x}{n}$$

7. La gráfica cuantil-cuantil, QQ, para comparar los cuantiles empíricos con los teóricos.

## 8. Las leyes de los grandes números, el teorema de Glivenko-Cantelli y su importancia.

Ley fuerte de los grandes números. Sea $\{X_n\}_{n \geq 1}$ una sucesión de variables aleatorias independientes idénticamente distribuidas con media $\mathbb{E}(X_n) = \mu$ y considérese la sucesión de los promedios aritméticos $\bar{X}_n = \frac{X_1 + \cdots + X_n}{n}$, entonces se cumple

$$\{\overline{X_n}\}_{n\geq 1} \xrightarrow{c.p.1} \mu$$

Ley débil de los grandes números. Sea $\{X_n\}_{n\geq 1}$ una sucesión de variables aleatorias independientes idénticamente distribuidas con media $\mathbb{E}(X_n) = \mu$ y considérese la sucesión de los promedios aritméticos $\bar{X}_n = \frac{X_1+\cdots+X_n}{n}$, entonces se cumple

Teorema de Glivenko-Cantelli. Sea $\{X_n\}$ una sucesión de variables aleatorias independientes, idénticamente distribuidas. Luego la función de distribución empírica de los primeros $n$ términos converge casi seguro a la función de distribución de $X_1$.

Demostración: Por la ley fuerte de los grandes números, es obvio que para cada $x$ tenemos la convergencia casi segura:

$$\lim_{n\to\infty} sup\left|F_n(x_j) - F_{X_1}(x_j)\right| = 0.$$

Es fácil ver, usando la monotonía de $F_{X_1}$ y $F_n$ que esto implica el teorema.

$$sup_{x\in\mathbb{R}}|F_n(x) - F_X(x)| \xrightarrow{c.p.1} 0$$

## 9. El estimador de momentos y el de máxima verosimilitud.

Para estimar el parámetro $\theta$ donde $\dim(\theta) = k$, se igualan los primeros momentos (centrales o no centrales) empíricos con los teóricos. Normalmente se hace esto con $\bar{X} = \mathbb{E}(X)$ igualando el primer momento empírico no central con el primer momento teórico no central (nos apoyamos de la ley débil de los grandes números).

Generalmente, usaremos para el caso continuo la aproximación continua a la función de verosimilitud, es decir:

$$\mathcal{L}(\theta; x_1, \dots, x_n) = c\prod_{i=1}^{n}[F(x_i;\theta)]$$

donde $c$ es una función constante con respecto al parámetro. Normalmente escogemos esta función de modo que nos quede la verosimilitud en función únicamente del parámetro de interés –en la medida de lo posible.

Será común utilizar la función de log verosimilitud:

$$l(\theta; x_1, \dots, x_n) = \ln \mathcal{L}(\theta; x_1, \dots, x_n)$$

Tanto para la función verosimilitud como para la log verosimilitud, la elección de la función constante $c$ no afecta el punto de maximización. Existen casos en los que es posible maximizar cualquiera de estas dos funciones analíticamente sin problemas. Cuando no, consideramos la función marcador (score) definida como:

$$S(\theta) = \frac{\partial l(\theta)}{\partial \theta}$$

y encontramos el máximo despejando $\theta$ de $S(\theta) = 0$. Es necesario verificar que sea un máximo e, idealmente, también que sea un máximo global. Para ello, consideramos la función $I$ de la información de Fisher como

$$I(\theta) = -\frac{\partial^2 l(\theta)}{\partial^2 \theta}$$

menos la segunda derivada con respecto al parámetro. Habremos encontrado un máximo cuando $I(\theta) \geq 0$.

**10. Identificación de estadísticas suficientes a partir de la función de verosimilitud.**

# TAREA

**1. Demuestra que la variable Bernoulli $(\boldsymbol{p})$ pertenece a la familia exponencial de un parámetro.**

Ofrecemos la factorización:

$$f(x;p) = (1-p)\exp\left[x\ln\left(\frac{p}{1-p}\right)\right]$$

**2. Demuestra que la densidad Gumbel de máximos con parámetros $(\boldsymbol{a},\boldsymbol{b})$,**

$$\boldsymbol{f(x;a,b) = \frac{1}{b}\exp\left[-\frac{x-a}{b}\right]\exp\left\{-\exp\left[-\frac{x-a}{b}\right]\right\}I_{(-\infty,\infty)}(x)},$$

**pertenece a la familia de localización y escala. Di cuál es el parámetro de localización y cuál el de escala.**

Observando la función de distribución, vemos que $x$ aparece siempre en la estructura $\frac{x-a}{b}$ y que, además, la expresión está multiplicada por $\frac{1}{b}$. Esto nos dice que efectivamente es de localización y escala con parámetro de localización $a$ y de escala $b$.

Podemos confirmar esto usando el teorema de cambio de variable $Y = \frac{X-a}{b}$, de donde $X = Yb + a$ y el valor absoluto de la derivada es $b$. Haciendo la sustitución, veremos que no depende de ninguno de los parámetros.

**3. Para la densidad exponencial con tiempo de vida garantizado $\boldsymbol{\theta}$,**

$$f(x;\ \theta) = e^{-(x-\theta)} I_{[\theta,\infty]}(x),$$

**calcula la mediana en términos del parámetro $\theta$. Supón que tienes una muestra de variables aleatorias $X_1, \dots, X_n$ independientes e idénticamente distribuidas con esta densidad. Encuentra un estimador de momentos para $\theta$. Da el estimador de máxima verosimilitud para $\theta$ y da la expresión de la función de verosimilitud relativa.**

Considerando la función de quantiles como la inversa generalizada:

$$Q_{50} = \inf\{x: F(x) \geq \alpha\} = \inf\{x: F(x) = .50\}$$

como es continua

$$\int_{\theta}^{Q_{50}} e^{-(x-\theta)}\, dx = \frac{1}{2}$$

de donde

$$1 - e^{\theta - Q_{50}} = \frac{1}{2} \Leftrightarrow e^{Q_{50}} = 2e\theta \Leftrightarrow Q_{50} = \theta + \ln(2).$$

Para encontrar un estimador de momentos, consideramos

$$\bar{X} = \mathbb{E}(X) = \int_{\theta}^{\infty} xe^{-(x-\theta)} dx = \theta + 1$$

de donde

$$\hat{\theta} = \bar{X} - 1$$

Finalmente, consideramos

$$\prod_{i=1}^{n} e^{-(x_i-\theta)} = e^{-(\sum x_i - n\theta)} \propto L(\theta; x_1, \dots, x_n)$$

Puesto que $0 \leq e^{-(\sum x_i - n\theta)} \leq 1$ por ser una probabilidad, el valor máximo se encuentra cuando $e^{-(\sum x_i - n\theta)} = 1$, es decir, cuando $n\theta - \sum x_i = 0$ de donde $\hat{\theta} = \frac{1}{n}\sum x_i$ como era de esperarse.

**4. Supón que tienes una muestra de variables aleatorias $X_1, \dots, X_n$ independientes e idénticamente distribuidas como normales con media $\mu$ y varianza $\sigma^2$. Demuestra que el estimador de máxima verosimilitud de $\sigma^2$ es insesgado asintóticamente y que también es consistente débil.**

Consideramos la aproximación continua a la función de verosimilitud como

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

considerando la función de log verosimilitud

$$l(\sigma^2;\ \mu, x) = \ln\left(\frac{1}{\sigma^n}\right) - \frac{\sum(x_i-\mu)^2}{2\sigma^2}$$

y derivando con respecto a $\sigma^2$ encontramos

$$S(\sigma^2) = \frac{\sum(x_i-\mu)^2}{\sigma^3} - \frac{n}{\sigma} = 0$$

de donde

$$\widehat{\sigma^2} = \frac{\sum(x_i-\mu)^2}{n}$$

Los estimadores de máxima verosimilitud siempre son consistentes.

Para mostrar que es insesgado asintótico, queremos mostrar que

$$\lim_{n\to\infty} \mathbb{E}\left(\widehat{\sigma^2}\right) = \sigma^2$$

y usaremos que $\mathbb{E}((x_i-\mu)^2) = \frac{n-1}{n}Var(X)$ y el hecho de que $\frac{n-1}{n} \to 1$ cuando $n \to \infty$.

**5. Supón que tienes una muestra de variables aleatorias $\boldsymbol{X_1, \dots, X_n}$ independientes e idénticamente distribuidas como una uniforme continua en el intervalo $(\boldsymbol{\theta}, \boldsymbol{2\theta})$. Da una expresión para la función de verosimilitud relativa y grafícala a mano. Utiliza para ello a las estadísticas de orden**

$$\boldsymbol{X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}},$$

**y considera la relación que guarda en particular el mínimo y el máximo con respecto a $\boldsymbol{\theta}$ y a $2\boldsymbol{\theta}$. Utiliza la aproximación continua a la función de verosimilitud.**

**6. Grafica a mano la función de distribución empírica para una muestra con las siguientes cuatro observaciones ordenadas:**

$$\boldsymbol{x_{(1)} = 2, x_{(2)} = 5, x_{(3)} = 6, x_{(4)} = 10}.$$

**Calcula usando tu gráfica ¿cuál es el cuantil empírico de probabilidad 0.8?**

La gráfica da 4 “saltos”, es la función continua 0 desde $-\infty$ hasta $x_{(1)}$ no inclusive; vale $0.25$ de $x_{(1)}$ hasta $x_{(2)}$ abierto a la derecha; vale $0.50$ de $x_{(2)}$ hasta $x_{(3)}$abierto a la derecha; vale $0.75$ desde $x_{(3)}$ hasta $x_{(4)}$ abierto a la derecha; y finalmente, vale $1$ desde $x_{(4)}$.

Usando la definición de inversa generalizada, el cuantil empírico de probabilidad 0.8 es $x_{(4)}$ pues para todo $x < x_{(4)}$ se tiene que $x \leq 0.75$.