FACULTAD DE MATEMÁTICAS
PONTIFICIA UNIVERSIDAD
CATÓLICA DE CHILE

Pontificada Universidad Católica

EYP2127

Inferencía Estadística

Autor:
Sebastián Lepe V.

2 de julio de 2025

Índice

Introducción y Motivación

Estadística: Es la disciplina que proporciona una metodología para ahcer inferencia, a partir de datos aleatorios reales sobre parámetros de modelos probabiliístico que se cree generaron tales datos,

Algunos ejemplos donde ocurren aleatoriedad son: En juegos de azar, en ciencias naturales, ingeniería, economía y ciencias sociales, etc.

Ejemplo: Sea $X : \Omega \to \mathbb{R}$ una función que cuenta el número de partículas emitidas por una fuente radiactiva en el próximo minuto.

Entonces podemos observar que X toma valores naturales. X se denomina **variable aleatoria**.

Notación:

- Denotaremos las variables aleatorias por letras mayúsculas.
- Denotaremos los valores tomados por las variables aleatorias por letras minúscula.
- Ω es el espacio muestral del evento que se está trabajando.

Supongamos que X tiene distribución Poisson con parámetro $\theta = 2$, entonces,

$$\mathbb{P}(X = x) = \frac{e^{-2}2^x}{x!}$$

Entonces si $x = 4$, entonces,

$$\mathbb{P}(X = 4) = \frac{e^{-2}2^4}{4!}$$

Supongamos ahora que θ es un parámetro desconocido positivo. Entonces la probabilidad de X tome valor x es,

$$\mathbb{P}(X = x) = \frac{e^{-\theta}\theta^x}{x!}$$

con $x \in \{0, 1, \dots\} = \mathbb{N}_0$. ¿Es posible determinar la probabilidad cuando $x = 4$? Y evidentemente no, puesto que tenemos un parámetro desconocido. Pero nos podemos enfocar en otro punto de vista. ¿Se puede aprender algo de θ? Y en este caso sí. Supongamos que queremos medir cinco períodos de 1 minuto, definiendo las siguientes variables aleatorias,

- X_1 número de partículas emitidas en el primer período.
- X_2 número de partículas emitidas en el segundo período.
- $\dots$
- X_5 número de partículas emitidas en el quinto período.

Definimos el vector aleatorio $\mathbf{X} = (X_1, \dots, X_5)$ que toma posibles valores de la forma $\mathbf{x} = (x_1, \dots, x_5)$.

Supongamos que $\mathbf{x} = (x_1, \ldots, x_n) = (2, 1, 0, 3, 4)$. Entoncea si $X_1, \ldots, X_5$ son variables independientes, se tiene que la probabilidad de que $\mathbf{X} = \mathbf{x}$ es,

$$\begin{aligned}
\mathbb{P}(\mathbf{X} = \mathbf{x}) &= \mathbb{P}(X_1 = x_1, \ldots, X_5 = x_5) \\
&= \prod_{i=1}^{5} \mathbb{P}(X_i = x_i) \\
&= \prod_{i=1}^{5} \frac{e^{-\theta}\theta^{x_i}}{x_i!}
\end{aligned}$$

Acabamos de construir un **modelo de probabilidad para el vector aleatorio $\mathbf{X}$**. Queremos:

- Utilizar el valor de $\mathbf{x}$ del vector $\mathbf{X}$ para aprender algo sobre el parámetro θ.
- Usar el método de verosimilitud como herramienta central para determinar varías técnicas de inferencia estadística.

Probabilidad vs Verosimilitud: Siguiendo con el ejemplo anterior de una variables aleatoria $X \sim \text{Poisson}(\theta)$, determinamos una función que depende de x y del parámetro θ. Si fijamos θ y vamos evaluando solamente x, obtenemos una función de probabilidad,

$$p(x) = \frac{e^{-\theta}\theta^{x}}{x!}$$

Fijando ahora x, digamos que $x = 5$ y variamos el parámetro θ obtenemos la **función de verosimilitud**,

$$L(\theta) := \frac{e^{-\theta}\theta^{5}}{5!}$$

(Solo consideramos un ejemplo con $x = 5$, pero esto puede variar).

De forma general para una función de probabildad/densidad, denotamos,

$$f_\theta(x) := p(x|\theta) = L(\theta|x)$$

La función de verosimilitud nos dice que tan verosimil es usar cierto parámetro θ.

Como objetivo principal es hacer inferencia sobre una población basadas en información contenida en una muestra. Como las poblaciones están descritas por parámetros, necesitamos hacer inferencia para poder estudiar la población.

Hay tres grandes categirías de inferencia estadística:

- **Frecuentista:** Los parámetros poblaciones son desconocidos y se obtienen estimaciones de ellos en base a una muestra.
- **Bayesiana:** Se considera los datos de una muestra y los parámetros fijos.
- **No-paramétrica:** No importa una forma matemática de los datos, solo los datos mismo.

1. Modelo Estadístico

Ejemplo: Supongamos que una variable aleatoria Y está descrito de la siguiente forma $Y = X\beta + \varepsilon$, donde $\varepsilon \sim \text{Normal}(0, \sigma^2 I)$.

¿Cómo utilizar las observaciones de manera óptima para:

- determinar las variables en X?
- Aprender sobre el valo de β ?
- validar el supuesto de Normalidad?
- Predecir datos?

Son preguntas al estudiar un modelo estadístico.

Definición: *Un modelo estadístico para un conjunto de datos, corresponde a un modelo de probabilidad que representa el proceso aleatorio que los generó.*

Definición: *Un modelo estadístico se dice paramétrico si la distribución de probabilidad subyacente, depende de un parámetro. En caso contrario, se dice no paramétrico.*

Nota: No centramos en modelos estadísticos paramétricos (MEP).

Definición: *Sea $X_1, \dots, X_n$ un conjunto de variables aleatorias. Se define el espacio muestral por el conjunto Ω de todos los valores que puede tomar el vector $(x_1, \dots, x_n)$.*

Abuso de notación: Denotaremos el vector aleatorio por X en vez de $\mathbf{X}$. De esta forma un vector aleatorio $X = (X_1, \dots, X_n)$ es una transformación de Ω a $\mathbb{R}^n$.

Definición: *Sea un modelo estadísitico paramétrico para $X_1, \dots, X_n$. Se define el espacio paramétrico,*

$$\mathcal{P}_\theta := \{f_\theta : \theta \in \Theta\}$$

donde Θ es el conjunto de todos los valores posibles de θ y se denomina espacio paramétrico del modelo.

Definición: *Las variables aleatoria $X_1, \dots, X_n$ se denominan una muestra aleatoria de tamaño n desde la población f_θ si $X_1, \dots, X_n$ son mutuamente independientes y la función de probabilidad/densidad de cada X_i es f_θ. En tal caso denotamos,*

$$X_1, \dots, X_n \overset{iid}{\sim} f$$

Definición: *Sea $X_1, \dots, X_n$ una muestra aleatoria de tamaño n. Consideremos la transformación $T(x_1, \dots, x_n) : D \subseteq \mathbb{R}^n \to \mathbb{R}/\mathbb{R}^n$ donde $\Omega \subseteq D$, entonces $T(X) = T(X_1, \dots, X_n)$ se denomina estadístico. La distribución de T se denomina la distribución muestral de T.*

Ejemplo (Media muestral): Es el promedio aritmético de una muestra aleatoria,

$$\overline{X} := \frac{X_1 + \cdots + X_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Ejemplo (Varianza muestral) Es el estadístico definido por,

$$S^2 := \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

Y se define la desviación estándar muestral por $S = \sqrt{S^2}$.

Proposición: *Sea $X_1, \ldots, X_n$ una muestra aleatoria con población μ y varianza $\sigma^2 < \infty$, entonces,*

(a) $\mathbb{E}(\overline{X}) = \mu$.

(b) $Var(\overline{X}) = \sigma^2/n$.

(c) $\mathbb{E}(S^2) = \sigma^2$.

Demostración: Si $X_1, \ldots, X_n$ es una muestra aleatoria, entonces son mutuamente independientes. Luego por definición se tiene que,

(a)

$$\mathbb{E}(\overline{X}) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i) = \mu$$

(b)

$$\text{Var}(\overline{X}) = \text{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\text{Var}(X_i) = \frac{\sigma^2}{n}$$

(c)

■

Inferencia sobre un Parámetro: Sea $X_1, \ldots, X_n$ una muestra aleatoria que toma valores en Ω. Queremos hacer inferencia sobre un parámetro θ del modelo estadístico. Para ello utilizamos estadísticos: $T(X) = T(X_1, \ldots, X_n)$.

Consideremos el estadístico $T(X) = T(X_1, \ldots, X_n)$ y sea,

$$\tau := \{t \in \mathbb{R}/\mathbb{R}^n : T(x) = t, \text{ para algún } x \in \Omega\}$$

Este estadístico induce una partición del espacio muestral Ω.

$$A_t := \{x \in \Omega : T(x) = t\}$$

con $t \in \tau$.

Si utilizamos $T(X)$ para hacer inferencia, la inferencia debe ser la misma para valores x en el mismo conjunto A_t.

1.1. Suficiencia

Definición: *Un estadístico $T(X)$ es suficiente para θ si y sólo si la distribución condicional de la muestra X dado el valor de $T(X)$ no depende de θ.*

Principio de Suficiencia: Si $T(X)$ es un estadístico suficiente para θ, la inferencia sobre este parámetro debe ser hecho sólo sobre la base de $T(X)$. Es decir, toda la información que tienen los datos sobre θ, está contenida en el estadístico $T(X)$, es suficiente.

Ejemplo: Consideremos dos casos.

(1) Sea una muestra de tamaño $n = 3$ con una observación $X = (3, 1, 2)$. Se sabe que proviene de una distribución poisson(θ). Se desea intuir θ.

(2) Se sabe que $T(X) = \sum_{i=1} 3X_i = 6$ y se sabe que $X = (x_1, x_2, x_3)$ sobre una distrivución poisson(θ). Se desea inferir θ.

En ambos casos nos dan información, pero en la segunda es más general que la primera, y aun así nos entrega igual o más información que la primera.

Teorema: *Sea f_θ una función densidad conjunta de $X_1, \ldots, X_n$. El estadístico $T(X)$ tiene una distribución q_θ es suficiente para θ si y sólo si,*

$$\frac{f_\theta(x)}{q_\theta(T(x))}$$

es constante como función con respecto de θ, para todo $x \in \Omega$.

Demostración :¿?

Este teorema caracteriza los estadístico suficientes para un parámetro θ. Por lo que para no estudiar complemetamente $T(X)$, se estudiar las funciones densidades de X y $T(X)$.

Ejemplo: Sea $X_1, \ldots, X_n$ una muestra aleatoria con distribución Bernoulli(θ). Como tenemos una muestra aleatoria, entonces la función densidad conjunta es,

$$\begin{aligned} f_\theta(x) &= \theta^{x_1}(1-\theta)^{1-x_1}\theta^{x_2}(1-\theta)^{1-x_2}\ldots(1-\theta)^{1-x_n} \\ &= \theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i} \end{aligned}$$

Consideremos el estadístico $T(X) = X_1 + \cdots + X_n$. Entonces $T(X)$ tiene distribución Binomial(n, θ). Entonces,

$$q_\theta(T(x)) = \binom{n}{T(x)}\theta^{T(x)}(1-\theta)^{n-T(x)}$$

Luego,

$$\frac{f_\theta(x)}{q_\theta(T(x))} = \frac{1}{\binom{n}{\sum_{i=1}^n x_i}}$$

Que es una constante con respecto a θ, para todo $x \in \Omega$. Por lo tanto, $T(X)$ es un estadístico suficiente.

Ejemplo: Sea $X_1, \dots, X_n$ una muestra aleatoria con distribución $N(\theta, \sigma^2)$. Sea el estadístico $T(X) = \overline{X}$. Entonces la función densidad conjunto es,

$$\begin{aligned} f_\theta(x) &= \prod_{i=1}^{n} \left(\frac{1}{2\pi\sigma^2}\right)^{1/2} \exp\left(-\frac{1}{2}\frac{(x_i-\theta)^2}{\sigma^2}\right) \\ &= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2\right) \\ &= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(x_i-\overline{x})^2 + n(\overline{x}-\theta)^2\right)\right) \end{aligned}$$

donde $\overline{x} = (\sum_{i=1}^{n} x_i)/n$. Notemos que $T(X)$ tiene distribución $N(\theta, \sigma^2/n)$, es decir,

$$q_\theta(y) = \left(\frac{1}{2\pi\sigma^2/n}\right)^{1/2} \exp\left(-\frac{n(y-\theta)^2}{2\sigma^2}\right)$$

Por tanto,

$$\frac{f_\theta(x)}{q_\theta(T(x))} = n^{1/2}\left(\frac{1}{2\pi\sigma^2}\right)^{(n-1)/2} \exp\left(\sum_{i=1}^{n}(x_i-\overline{x})^2\right)$$

Donde claramente no depende de θ, por tanto $T(X)$ es suficiente.

Teorema (Teorema de factorización (Halmos y Savage)): *Sea f_θ la función de densidad conjunta de la muestra X. Un estadístico $T(X)$ es suficiente para θ si y sólo si existen funciones g_θ y h, tales que, para todo $x \in \Omega$ y $\theta \in \Theta$,*

$$f_\theta(x) = g_\theta(T(x))h(x)$$

Demostración: Sea $T(X)$ un estadístico suficiente. Esto significa que la razón,

$$\frac{f_\theta(x)}{q_\theta(T(x))} = C$$

es constante para todo $x \in \Omega$. Entonces,

$$f_\theta(x) = q_\theta(T(x))C$$

Encontrando las funciones g_θ, h.

Supongamos que existen funciones g_θ, h al que para todo $x \in \Omega, \theta \in \Theta$, se tiene que,

$$f_\theta(x) = g_\theta(T(x))h(x)$$

Sea $q_\theta(t)$ la densidad de $T(X)$ **terminar**

Ejemplo: Sea $X_1, \dots, X_n$ una muestra aleatoria proviene de una distribución uniforme en el conjunto $\{1, \dots, \theta\}$, con $\theta \in \mathbb{N}$. Encontremos un estadístico suficiente. Sabemos que la función densidad de una uniforme discreta es,

$$f_{X_i}(x_i) = \frac{1}{\theta}\mathbb{1}_{\{1,\dots,\theta\}}(x_i)$$

donde $x_i \in \{1, \dots, \theta\}$. Luego la conjunta es,

$$f_\theta(x) = \theta^{-n}\prod_{i=1}^{n}\mathbb{1}_{\{1,\dots,\theta\}}(x_i)$$

Suponiendo que existe un estadístico, entonces por el teorema de factorización existe g_θ, h tales que $f_\theta(x) = g_\theta(T(x))h(x)$. Definamos la muestra aleatoria ordenada $X_{(1)}, \dots, X_{(n)}$ definido de la siguiente forma,

$$X_{(1)} = \min_{i=1,\dots,n} X_i \leq X_{(2)} \leq \dots X_{(n-1)} \leq X_{(n)} = \max_{i=1,\dots,n} X_i$$

Entonces podemos tomar el estadístico $T(X) = X_{(n)}$ definido $T(x) : \Omega \to \{1, \dots, n\}$. Luego podemos tomar $g_\theta(x) = \theta^{-n}\mathbb{1}_{x\leq\theta}$ y $h(x) = 1$. Por lo que,

$$\begin{aligned} g_\theta(T(x))h(x) = \theta^{-n}\mathbb{1}_{\max x_i \leq \theta} &= \begin{cases} \theta^{-n}, & \max x_i \leq \theta \\ 0, & \text{e.o.c} \end{cases} \\ &= f_\theta(x) \end{aligned}$$

Para todo $x \in \Omega$ y para todo parámetro θ natural.

Ejemplo: Sea $X_1, \dots, X_n$ una muestra aleatoria proveniente de una distribución $N(\mu, \sigma^2)$ con varianza desconocido. Encontremos un estadístico sufiente para $\theta = (\mu, \sigma^2)$. En este caso estamos trabajando con dos parámetros desconocidos (μ, σ^2).

Para ello observemos que,

$$f_\theta(x) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right)$$

Podemos rescribir la conjunta usando $\overline{x}$, de forma que,

$$f_\theta(x) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left(-\frac{(n-1)s^2 + n(\overline{x} - \mu)}{2\sigma^2}\right)$$

donde $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$. Entonces podemos ver dos estadísticos. En particular, consideramos $T(X) = (S^2, \overline{X})$ y tomamos $g_{(\mu,\sigma^2)}(x, y)$ de tal forma que,

$$g_{(\mu,\sigma^2)}(x, y) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left(-\frac{(n-1)x + n(y - \mu)}{2\sigma^2}\right)$$

Y tomando $h(x) = 1$, se concluye que $f_\theta(x) = g_\theta(T(x))h(x)$, encontrando un estadístico suficiente.

Observemos lo siguiente. Sea $T(X)$ un estadístico suficiente para θ. Sea $\gamma : \mathbb{R}/\mathbb{R}^n \to \mathbb{R}/\mathbb{R}^n$ una función inyectiva. Entonces $T'(X) := \gamma(T(X))$ tambíen es un estadístico suficiente. **¿?**. A partir de esto podemos definir un estadístico suficiente minimal.

Definición: *Un estadístico sufienciente $T(X)$ se dice **suficiente minimal** si y sólo para cualquier otro estadístico suficiente $T'(X)$, existe una función ν, tal que,*

$$T(X) = \nu(T'(X))$$

Teorema: *Sea f_θ la función densidad de una muestra aleatoria $X_1, \ldots, X_n$. Existe un estadístico $T(X)$ tal que la razón,*

$$\frac{f_\theta(x)}{f_\theta(y)}$$

Es constante como función de θ si y sólo si $T(x) = T(y)$ para todo $x, y \in \Omega$. Entonces $T(X)$ es un estadístico suficiente minimal para θ.

Corolario: *Toda función inyectiva de un estadístico suficiente minimal, es también suficiente minimal.*

Demostración: Esto se evidente, puesto que la composición de funciones inyectivas, es una función inyectiva. ■

Ejemplo: Sea $X_1, \ldots, X_n$ una muestra aleatoria proveniente de una distribución $N(\mu, \sigma^2)$ con ambos parámetros desconocidos. Probemos que $T(X) = (\overline{X}, S^2)$ corresonde a un estadístico suficiente minimal para $\theta = (\mu, \sigma^2)$. Observemos que,

$$\begin{aligned}
\frac{f_\theta(x)}{f_\theta(y)} &= \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}((x_i - \overline{x})^2 - (y_i - \overline{y})^2) + n((\overline{x} - \mu) - (\overline{y} - \mu)^2)\right)\right) \\
&= \exp\left(-\frac{1}{2\sigma^2}\left((n-1)(s_x^2 - s_y^2) + n((\overline{x} - \mu) - (\overline{y} - \mu)^2)\right)\right)
\end{aligned}$$

Si la razón es constante, entonces necesariamente $s_x^2 = s_y^2$ y $\overline{x} = \overline{y}$ ya que el valor no varia al cambiar σ^2, μ. Por tanto encesariamente $T(x) = T(y)$ para todo $x, y \in \Omega$. Ahora observemos que si $T(x) = T(y)$ para todo $x, y \in \Omega$, entonces claramente la razón anterior es constante.

Ejemplo: Sea $X_1, \ldots, X_n$ una muestra con distribución uniforme en $(\theta, \theta + 1)$ con θ finito. Encuentre un estadístico suficiente minimal para θ.

Observemos lo siguiente,

$$\begin{aligned}
f_\theta(x) &= \begin{cases} 1, & \theta < x_{(1)} \leq \cdots \leq x_{(n)} < \theta + 1 \\ 0, & \text{e.o.c} \end{cases} \\
&= \begin{cases} 1, & x_{(n)} - 1 < \theta < x_{(1)} \\ 0, & \text{e.o.c} \end{cases}
\end{aligned}$$

Por lo que por el teorema de factorización podemos tomar el estadístico $T(X) = (X_{(1)}, X_{(n)})$. Probemos que es minimal. Tenemos que la razón,

$$\frac{f_\theta(x)}{f_\theta(y)}$$

es constante si $x_n - 1 < \theta < x_{(1)}$ y $y_n - 1 < \theta < x_{(1)}$ y solamente si $x_{(1)} = y_{(1)}, x_{(n)} = y_{(n)}$ (en caso de que $x_{(n)} < y_{(n)}$ entonces podemos tomar un θ que hace que la razón cambie de valor, cosa que es contradictorio). Por tanto necesariamente $T(x) = T(y)$ para todo $x, y \in \Omega$. Por tanto $T(X)$ es minimal.

1.2. Ancilaridad y Completitud

Definición: *Un estadístico $S(X)$ se dice* ***ancilar para un parámetro*** *θ si y sólo si su distribución no depende de este parámetro.*

Cambio de Variables: *Sea $T(u, v) = (x, y)$ una transformación donde $x = g(u, v), y = (h(u, v))$ son transformaciones uno a uno C^1 con Jacobiano no nulo en la región S sobre uv. swdqadwa*

Ejemplo: Consideremos una muestra aleatoria $X_1, \ldots, X_n$ con distribución uniforme sobre $(\theta, \theta+1)$. Sea definimos el rango de la muestra por $R(X) := X_{(n)} - X_{(1)}$. Probemos que R es un estadístico ancilar para el parámetro θ. Observemos que la distribución conjunta de la muestra es,

$$f_\theta(x) = \begin{cases} 1, & x_i \in (\theta, \theta+1), \;\; i = 1, \ldots, n \\ 0, & \text{e.o.c} \end{cases}$$

Observemos que podemos considerar $x_{(1)} = \text{mín}_{i=1,\ldots,n} x_i$ y $x_{(n)} = \text{máx}_{i=1,\ldots,n} x_i$. De esta forma se tiene podemos reescribir la conjunta de la siguiente forma,

$$f_\theta(x) = \begin{cases} 1, & x_{(n)} - 1 < \theta < x_{(1)} \\ 0, & \text{e.o.c} \end{cases}$$

Consideremos la conjunta de el vector $(X_{(1)}, X_{(n)})$,

$$f_\theta(X_{(1)}, X_{(n)}) = n(n-1)(X_{(n)} - X_{(1)})^{n-2}$$

Sea $M = (X_{(n)} + X_{(1)})/2$ otro estadístico. Entonces se cumple que,

$$X_{(n)} = \frac{2M + R}{2}$$
$$X_{(1)} = \frac{2M - R}{2}$$

Ahora usando cambio de variable, obtenemos que la densidad se determina por,

$$f_\theta(r, m) = n(n-1)r^{n-2}|J|$$

donde $|J|$ es la Jacobiana que tiene valor 1. Pero hemos obtenido la conjunto y nos interesa determinar la función densidad/probabilidad con respecto a r, por lo que debemos determinar la marginal. Obserbvemos que $0 < r < 1$ y $\theta + r/2 < m < \theta + 1 - r/2$. Esto último implica que,

$$f(r|\theta) = \int_{\theta+r/2}^{\theta+1-r/2} n(n-1)r^{n-2} dm = n(n-1)r^{n-2}(1-r)$$

con $0 < r < 1$. Recordemos la distribución beta, decimos que una variable aleatoria tiene distribución beta con parámetros α, β en ese orden si,

$$g(x|\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$$

con $0 < x < 1$. En este caso se cumple que la distribución de R es Beta$(n-1, 2)$. Por lo tanto, no depende de θ, de forma que es ancilar.

Ejemplo: Vamos a generalizar el caso anterior. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una distribución en una familia de localización con función de distribución $F(x-\theta), \theta \in \mathbb{R}$. Probemos que $R(X) = X_{(n)} - X_{(1)}$ es ancilar.

Para ver esto tenemos que simplemente probar que la distribución del rango es independiente de θ. Pero claramente no es tan sencillo, debemos deteriminar la distribución de $X_{(n)} - X_{(1)}$. Pero para no hacerlo usaremos un pequeño truco, sea $Z_i := X_i + \theta$, entonces claramente se cumple que,

$$Z_{(n)} = X_{(n)} + \theta$$
$$Z_{(1)} = X_{(1)} + \theta$$

Y que,

$$\begin{aligned}\mathbb{P}(R \leq r) &= \mathbb{P}(\text{máx}\, X_i - \text{mín}\, X_i \leq r) \\ &= \mathbb{P}(\text{máx}\, Z_i - \text{mín}\, Z_i \leq r)\end{aligned}$$

Esto significa que el rango es independiente de θ, y por tanto es ancilar con respecto a θ.

Este truco está relacionado con un teorema que más adelante estudiaremos.

Dependencia de Estadísticos minimales y ancilares: Consideremos una muestra X_1, X_2 con distribución dada por,

$$\mathbb{P}_\theta(X = \theta) = \mathbb{P}_\theta(X = \theta + 1) = \mathbb{P}_\theta(X = \theta + 2) = \frac{1}{3}$$

con θ valor entero desconocido. Sea el estadístico,

$$T(X) = (R(X), M(X)) = \left(X_{(2)} - X_{(1)}, \frac{X_{(1)} + X_{(2)}}{2} \right)$$

Sabemos que $R(X)$ es ancilar para θ, entocnes se puede demostar que $T(X)$ es suficiente minimal, sin embargo, ellos no son independientes.

Definición: *Sea f_θ una familia de funciones de densidad (probabilidad) para un estadístico $T(X)$. La familia f_θ se dice una* ***familia completa*** *si, para toda función g, se tiene que,*

$$\mathbb{E}_\theta(g(T(X))) = 0$$

para todo $\theta \in \Theta$ si y sólo si.

$$\mathbb{P}_\theta(g(T(X)) = 0) = 1$$

para todo $\theta \in \Theta$. En este caso se dice que $T(X)$ es un ***estadístico completo****.*

Ejemplo 1.29: Consideremos un estadístico T con distribución Binomial(n, θ). Probemos que T es un estadístico completo.

Observemos que la función densidad de T es,

$$f(t|\theta) = \binom{n}{t}\theta^t(1-\theta)^{n-t}$$

Supongamos que $\mathbb{E}_\theta(g(T)) = 0$ para toda función g. Entonces,

$$\mathbb{E}_\theta(g(T)) = \sum_{t=0}^{n} g(t)\binom{n}{t}\theta^t(1-\theta)^{n-t} = 0$$

Lo que implica que necesariamente $g(t) = 0$ para todo $t = 0, 1 \ldots, n$. Lo que implica que $\mathbb{P}(g(t) = 0) = 1$, de forma qeu $T(X)$ es un estadístico completo.

Por otro lado, si $g(T) = 0$ entonces se tiene que $\mathbb{E}_\theta(g(T(X)) = 0$. Por lo tanto, T es un estadístico coompleto.

Teorema (Teorema de Basú): *Sea $T(X)$ un estadístico suficiente minimal completo. Entonces $T(X)$ es independiente de todo estadístico ancilar $S(X)$.*

Teorema: *Sea $X_1, \ldots, X_n$ una muestra aleatoria proveniente de una distribución en una familia exponencial, con función de densidad dada por,*

$$f_\theta(x) = h(x)c(\theta)\exp\left(\sum_{j=1}^{k}\omega_j(\theta)t_j(x)\right)$$

donde $\theta = (\theta_1, \ldots, \theta_k)$. Si el conjunto $\{(\omega_1(\theta), \ldots \omega_k(\theta)) : \theta \in \Theta\}$ contiene un conjunto abierto en $\mathbb{R}^k$, entonces, el estadístico,

$$T(X) = \left(\sum_{i=1}^{n} t_1(X_i), \ldots, \sum_{i=1}^{n} t_k(X_i)\right)$$

es completo.

Ejemplo: Sea $X_1, \ldots, X_n$ una muestra aleatoria proveniente de una distribución Normal(μ, σ^2) con σ conocido. Probemos que los estadísticos S^2 y $\overline{X}$ son independientes.

Observemos que $\overline{X}$ es un estadístico suficiente y minimal con respecto al parámetro μ (si σ fuera desconocidos, entonces decimos que $\overline{X}$ es suficiente minimal para $\theta = (\mu, \sigma^2)$). Observemos lo siguiente,

$$\begin{aligned} f_\theta(x_i) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x_i - \mu)^2\right) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x_i^2 - 2x_i\mu + \mu^2)\right) \end{aligned}$$

Tomando $h(x) = 1/\sqrt{2\pi\sigma^2}$, $c(\mu) = 1$, $k = 3$ y

$$\begin{aligned} \omega_1(\mu)t_1(x_i) &= -\frac{1}{2\sigma^2}x_i^2 \\ \omega_2(\mu)t_2(x_i) &= \frac{1}{\sigma^2}x_i\mu \\ \omega_3(\mu)t_3(x_i) &= -\frac{1}{2\sigma^2}\mu^2 \end{aligned}$$

Entonces el estadístico,

$$T(X) = \left(\sum_{i=1}^n X_i^2, \sum_{i=1}^n X_i, n\right)$$

Es completo. Esto implica que la segunda coordenada, que es un estadístico, es completo. Por lo tanto $\overline{X}$ es completo. Finalmente queda ver que S^2 sea ancilar, pero observemos que $\mathbb{E}(S^2) =$

1.3. Familia Exponencial

Definición: *Se dice que una familia de distribución paramétrica, con espacio paramétrico Θ y función de densidad f_θ, corresponde a una* ***familia exponencial*** *si y sólo si,*

$$f_\theta(x) = h(x)c(\theta) \exp\left(\sum_{j=1}^k \omega_j(\theta)t_j(x)\right)$$

con $\theta = (\theta_1, \ldots, \theta_k) \in \Theta$ para $k \in \mathbb{N}$.

Ejemplo: Consideremos la siguiente familia de distribuciones,

$$\mathcal{F} = \{\text{Normal}(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma > 0\}$$

Tomemos la normal con media μ y varianza σ^2, entonces la función densidad es de la forma,

$$\begin{aligned} f_\theta(x) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x^2 - 2x\mu + \mu^2)\right) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\mu^2}{2\sigma^2}\right) \exp\left(-\frac{1}{2\sigma^2}x^2 + \frac{\mu}{\sigma^2}x\right) \end{aligned}$$

Claramente consistuye a una familia exponencial.

Ejemplo: Consideremos la siguiente familia,

$$\mathcal{F} = \{\text{Uniforme}(0,\theta), \theta > 0\}$$

Veamos si es una exponencial. Sea la uniforme que va de 0 a θ, entonces la función densidad es,

$$\begin{aligned} f_\theta(x) &= \begin{cases} \frac{1}{\theta}, & x \in (0,\theta) \\ 0, & \text{e.o.c} \end{cases} \\ &= \frac{1}{\theta}\mathbb{1}_{(0,\theta)}(x) \end{aligned}$$

Se tiene que no pertenece a una familia exponencial pues el soporte de x no es constante.

Definición: *Consideremos una familia exponencia, el parámetro:*

$$\eta = (\omega_1(\theta), \ldots, \omega_k(\theta))$$

Se denomina ***parámetro natural*** *de la distribución.*

Definición: *Consideremos una familia exponencial, el conjunto,*

$$\tau = \left\{ \eta \in \mathbb{R}^k : \int_\Omega h(x) \exp\left(\sum_{j=1}^{k} \omega_j t_j(x) \right) dx < \infty \right\}$$

Se denomina ***espacio paramétrico natural****.*

Definición: *Considere una muestra aleatoria $X_1, \ldots, X_n$ de una distribución en una familia exponencial. Entonces, el estadístico:*

$$T(X) = \left(\sum_{i=1}^{n} t_i(X_i), \ldots, \sum_{i=1}^{n} t_k(X_i) \right)$$

Es un estadístico suficiente, y se denomina ***estadístico suficiente natural****.*

Definición: *Consideremos una familia exponencial. Si el espacio parámetrico natural τ, contiene un conjunto abierto en $\mathbb{R}^k$, el estadístico suficiente natural es completo.*

Ejemplo: Consideremos la familia,

$$\mathcal{F} = \{\text{Normal}(\mu,\sigma^2), \mu \in \mathbb{R}, \sigma > 0\}$$

- Determinemos el espacio paramétrico natural.
- Encuentre un estadístico suficiente natural y determine si es completo.

Ya sabemos que pertenece a una familia exponencial. Descomponiendo obtenemos que,

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

Tomando $\eta = (1/\sigma^2, \mu/\sigma^2)$ como parámetro natural, se obtiene el estadístico suficiente natural $T(X) = (-x^2/2, x)$. Entonces queremos los $\eta \in \mathbb{R}^2$ tales que la matriz,

$$\int_{\mathbb{R}} \exp\left(-\frac{1}{2\sigma^2}x^2 + \frac{\mu}{\sigma^2}x\right) < \infty$$

Para que la integral sea finita, el coeficiente asociado a x^2 debe ser negativo. Es decir,

$$\eta_1 => 0$$

Si η_0 entonces no hay restricción sobre η_2, finalmente el espacio paramétrico natural, está dado por,

$$\tau = \{(\eta_1, \eta_1) : \eta_1 > 0, -\infty < \eta_2 < \infty\}$$

Ahora los estadísticos son,

$$T(X) = \left(-\frac{X^2}{2}, X\right)$$

Que al provenir de una famialia exponecial, se tiene que es completo.

Teorema: *Si $T(X)$ correpsinde a un estadístico suficiente completo de dimensión finita, entonces es suficiente minimal.*

Teorema: *Si $T(X)$ corresponde a un estadístico suficiente completo de dimensión finita y $R(X)$ es un estadístico anciclar, entonces $T(X)$ y $R(X)$ son independientes.*

2. Estimación Puntual

2.1. Máxima Verosimilitud

Prinicipio de Equivarianza frente a cambio e Escala de medición: El principio de equivarianza bajo cambio de escala de medición indica que la inferencias basadas en observaciones que solo difieren en su escala de medida deben ser idénticas.

Prinicio de Suficiencia: Si $T(X)$ es un estadístico suficiente para θ, toda inferencia sobre θ debe depender de la muestra solo a través de $T(X)$.

Principio de Verosimilitud: Si L corresponde a la función de verosimilitud de θ y x e y corresponden a dos puntos muestrales tales que, para todo $\theta \in \Theta$ se cumple que:

$$L(\theta, x) = C(x, y)L(\theta, y)$$

las inferencias basadas en x e y deben ser las mismas.

Definición: *Un **estimador puntual** corresponde a cualquier estadístico $T(X)$.*

Definición: *Dado un estadístico $T(X)$ y una muestra $X = x$, una **estimación puntual** corresponde a $T(X)$.*

Lo que nos interesa ahora son métodos para encontrar estimadores. Para ellos estudiaremos tres estimadores:

- Método de Máxima Verosimilitud.
- Método de Momentos.
- Estimadores de Bayes.

Definición: *Sea $X_1, \dots, X_n$ una muestra aleatoria proveniente de una población con función de densidad/probabilidad f_θ, de modo que su función de verosimilitud corresponde a:*

$$L(\theta, x) = \prod_{i=1}^{n} f_\theta(x_i)$$

*para cada punto muestra x. Sea $\hat{\theta}(x)$ un valor del parámetro θ donde L alcanza un máximo como función de θ con x fijo. Un **estimador máximo verosímil (EMV)** de θ, basado en una meustra X, corresponde a,*

$$\hat{\theta}(X)$$

Supongamos que $L : \mathbb{R}^k \to \mathbb{R}$ es diferenciable como función de θ, candidatos a EMV de θ corresponden a los puntos de la forma $(\theta_1, \dots, \theta_k)$ que satisfacen,

$$\frac{\partial}{\partial \theta_j} L(\theta, x) = 0$$

para todo $j = 1, \ldots, k$; donde $k = \dim(\theta)$. También son puntos candidatos los puntos extremos del espacio paramétrico θ.

Observación: Los puntos críticos no necesariamente son máximales, por lo que en caso de que la función sea dos veces diferenciable, se podría verificar si es máximo.

Ejemplo: Sea $X_1, \ldots, X_n$ una muestra aleatoria de una distribución Normal$(\theta, 1)$. Determinemos el estimador máxima verósimil de θ. Sabemos que la función de verosimilitud conjunta es,

$$\begin{aligned} L(\theta, x) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x_i - \theta)^2\right) \\ &= \left(\frac{1}{2\pi}\right)^{n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i - \theta)^2\right) \end{aligned}$$

Ahora vamos a derivar con respecto a θ, por lo que,

$$\frac{\partial L(\theta, x)}{\partial \theta} = 0 \Longleftrightarrow \sum_{i=1}^{n}(x_i - \theta) \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i - \theta)^2\right) = 0$$

Como la exponencia no puede ser 0, necesariamente,

$$\sum_{i=1}^{n}(x_i - \theta) = 0 \Longleftrightarrow \theta = \overline{x}$$

Por lo tanto, como candidato tenemos a $\theta(x) = \overline{x}$. Estudiando la segunda derivada vemos que,

$$\frac{\partial^2 L(\theta, x)}{\partial \theta^2}(\overline{x}) < 0$$

Por tanto el θ anterior es un máximo y finalmente nuestro EMV es $\hat{\theta} = \overline{X}$.

Ejemplo: Sea $X_1, \ldots, X_n$ una muestra aleatoria de una distribución Normal$(\theta, 1)$ con $\theta > 0$. Determinemos el estimador máximo verosímil de θ. Aunque es similar al ejemplo anterior, en este caso pedimos que $\theta > 0$ y esto cambia radicalmente el ejericicio, puesto que $\overline{X}$ podría tomar valores negativos, lo que no es posible por la restricción inicial. Por lo que necesariamente restructuramos la función de verosimilitud para que todo funciones, de esta forma,

$$\hat{\theta} = \begin{cases} \overline{X}, & \text{si } \overline{X} > 0 \\ 0, & \text{si } \overline{X} < 0 \end{cases}$$

Ejemplo: Sea $X_1, \ldots, X_n$ una muestra aleatoria de una distribución uniforme $[0, \theta]$ con $\theta > 0$. Determinemos el EMV de θ.

Observemos que la función de verosimilitud conjunta es,

$$\begin{aligned} L(\theta, x) &= \prod_{i=1}^{n} \frac{1}{\theta} \mathbb{1}_{[0,\theta]}(x_i) \\ &= \frac{1}{\theta^n} \prod_{i=1}^{n} \mathbb{1}_{[0,\theta]}(x_i) \end{aligned}$$

Para x fijo tenemos que $L(\theta, x)$ es una función continua a partir de θ grande, en particular, tal que cuando $\theta \geq x_{(n)}$. Observamos que también es una función decreciente, por tanto, el máximo se alcanza en $\theta = x_{(n)}$, de esta forma $\hat{\theta}(x) = x_{(n)}$ y por tanto el EMV es $\hat{\theta}(X) = X_{(n)}$.

Ejemplo: Sea $X_1, \ldots, X_n$ una muestra aleatoria de distribución Binomial(k, p) con p conocido. Encontremos el EMV con respecto al parámetro k.

Por definición tenemos que estudiar la función,

$$L(k, x) = \prod_{i=1}^{n} \binom{k}{x_i} p^{x_i} (1-p)^{k-x_i}$$

donde $k \in \mathbb{N}$ y $x_i = 0, 1, \ldots, k$ para todo $i = 1, \ldots, n$. Ocurre que no es una función continua por lo que noes fácil estudiar o determinar el EMV, por lo que se necesita hacer una comparación de crecimiento con respecto a k. Primero que nada, si $k < x_{(n)}$, entonces $L(k, x) = 0$. Supongamso que $0 \leq x_i \leq k$ para todo $i = 1, \ldots, n$ con $x = (x_1, \ldots, x_n)$ fijo. Queremos un k máximal, por una parte se cumple que,

$$\frac{L(k, x)}{L(k-1, x)} \geq 1 \iff k^n (1-p)^n \geq \prod_{i=1}^{n} (k - x_i)$$

Y por otro lado,

$$\frac{L(k+1, x)}{L(k, x)} \leq 1 \iff (k+1)^n (1-p)^n \leq \prod_{i=1}^{n} (k - x_i)$$

Sea $z = 1/k$, entonces obtenemos el siguiente problema a resolver,

$$(1-p)^n \geq \prod_{i=1}^{n} (1 - x_i z)$$

donde $0 < z \leq 1/x_{(n)}$. Se tiene que la función,

$$g(z) := \prod_{i=1}^{n} (1 - x_i z)$$

es decreciente en z con valor 1 en $z = 0$ y 0 cuando $z = 1/x_{(n)}$. Por lo tanto existe un único valor de z que satisface la ecuación. El problema podría ser que no sea entero, pero el mayor entero menor o igual a este será el EMV.

Ejemplo: Sea $X_1, \ldots, X_n$ una muestra aleatoria con distribución Bernoulli(θ). Determinemos el EMV.

Sabemos que la función de verosimilitud es,

$$\begin{aligned} L(\theta, x) &= \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} \\ &= \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i} \end{aligned}$$

Aplicando logaritmo natural a la función verosimilitud, obtenemos,

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

Derivando con respecto a θ obtenemos que,

$$\frac{\partial \log L}{\partial \theta} = \frac{1}{\theta}\sum_{i=1}^{n} x_i + \frac{1}{1-\theta}\left(n - \sum_{i=1}^{n} x_i\right)$$

Igualando a 0 y despejando θ, se obtiene como posible candidato $\hat{\theta} = \overline{x}$. Si derivamos de nuevo y estudiamos $\theta = \overline{x}$, se comprueba que la segunda derivada con respecto a $\theta = \overline{x}$ es un valor negativo y por tanto el EMV es $\overline{X}$.

Observación: En el ejemplo anterior aplicamos una función log y de igual forma hemos obtenido la función EMV, y esto se debe a que log es una función creciente biyectiva.

Observación: A veces es más fácil máximizar la función $\log(L)$ en vez de L.

Invarianza del Estimador de Máxima Verosimilitud: Si $\hat{\theta}$ es el EMV de θ y $\tau(\theta)$ es función uno a uno, entonces el EMV de $\tau(\theta)$, es $\tau(\hat{\theta})$.

Ejemplo: Considere una variable aleatoria con distribución Binomial(n, θ) con $0 < \theta < 1$. Determinemos el EMV de las posibilidades de éxitos dadas por:

$$\eta = \log\frac{\theta}{1-\theta}$$

Tenemos un cambio de parámetro. Por la invarianza del EMV se tiene que si $\hat{\theta}$ es el EMV con respecto al parámetro θ, entonces el EMV con respecto a η es,

$$\hat{\eta} = \log\left(\frac{\hat{\theta}}{1-\hat{\theta}}\right)$$

Determinemos el EMV de la binomial. Por definición,

$$\begin{aligned} L(\theta, x) &= \prod_{i=1}^{n} \binom{n}{x_i} \theta^{x_i}(1-\theta)^{n-x_i} \\ &= C(x)\theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n^2-\sum_{i=1}^{n} x_i} \end{aligned}$$

donde C es una constante que solo depende de x y no de θ. Aplicando logartimo e ignorando la constate C, se obtiene que,

$$\log(L) = \sum_{i=1}^{n} x_i \log\theta + \left(n^2 - \sum_{i=1}^{n} x_i\right)\log(1-\theta)$$

Derivando e igualando a 0 se obtiene que,

$$\hat{\theta}(x) = \frac{\sum_{i=1}^{n} x_i}{n^2}$$

Se puede comprobar que es máximal y por tanto, el EMV es,

$$\hat{\theta}(X) = \frac{\overline{X}}{n}$$

Finalmente aplicamos el cambio de parámetro y por tanto,

$$\hat{\eta}(X) = \log\left(\frac{\overline{X}}{n-\overline{X}}\right)$$

2.2. Métodos de Momentos

Definición: *Sea X una variable aleatoria. Se denomina k-ésimo momento poblacional de su distribución al parámetro:*

$$\mu_k = \mathbb{E}(X^k)$$

para todo $k = \mathbb{N}$.

Definición: *Sea $X_1, \dots, X_n$ una muestra aleatoria de una variable X. Se denomina k-ésimo momento muestral al estadístico:*

$$m_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

para todo $k = \mathbb{N}$.

Sea $X_1, \dots, X_n$ una muestra aleatoria de una variable X proveniente de una población con función densidad/probabildad f_θ donde $\theta = (\theta_1, \dots, \theta_n)$. El estimador de momentos $\widetilde{\theta}$ de θ corresponde a la solución del sistema de ecuaciones:

$$\begin{aligned} m_1 &= \mu_1(\theta) \\ &\vdots \\ m_k &= \mu_k(\theta) \end{aligned}$$

Ley débil de los Grandes Números: *Sea $X_1, X_2, \dots$ una secuencia de variables aleatorias independientes e idénticamente distribuidas, cada una con media $\mu < \infty$. Entonces:*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mu$$

Ley fuerte de los Grandes Números:*Sea $X_1, X_2, \dots$ una secuencia de variables aleatorias independientes e idénticamente distribuidas, cada una con media $\mu < \infty$. Entonces:*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{c.s} \mu$$

Ejemplo: Sea $X_1, \dots, X_n$ una muestra aleatoria con distribución Normal(μ, σ^2) con μ y σ^2 desconocidos. Determinemos los estimadores de momentos de μ y σ^2.

Sabemos lo siguiente:

$$m_1 = \frac{1}{n}\sum_{i=1}^{n} X_i$$
$$m_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

con media μ_1, μ_2. Claramente $\mu_1 = \mu$, por lo que $\mu = \overline{X}$. Determinemos μ_2. Notemos que,

$$\begin{aligned} \text{Var}(X_i) &= \mathbb{E}((X_i - \mu)^2) \\ &= \mathbb{E}(X_i^2 - 2\mu X_i + \mu^2) \\ &= \mathbb{E}(X_i^2) - 2\mu\mathbb{E}(X_i) + \mu^2 \\ &= \mathbb{E}(X_i^2) - \mu^2 = \sigma^2 \end{aligned}$$

Entonces $\mathbb{E}(X_i^2) = \sigma^2 + \mu^2$. De esta forma

$$\begin{aligned} \sigma^2 &= \left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right) - \overline{X}^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \end{aligned}$$

Determinando los estimadores de momentos.

Ejemplo: Sea $X_1, \dots, X_n$ una muestra aleatoria de una distribución Binomial(k, p) con k y p desconocidos. Encuentre los estimadores de momentos de k y p.

Observemos que,

$$\overline{X} = kp$$
$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 = kp(1-p) + p^2k^2$$

Entonces al resolver, obtenemos que,

$$k = \frac{\overline{X}}{\overline{X} - \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

2.3. Estimadores de Bayes

Definición: *Sea X una variable aleatoria con función densidad (probabilidad) f_θ y sea $\pi(\theta)$ una función de creencia sobre θ, que cumple las propiedades de una función densidad (probabilidad). La función $\pi(\theta)$ se denomina* ***función de densidad/probabilidad a priori para*** *θ. La función:*

$$\pi(\theta|x) = \frac{\pi(\theta)f_\theta(x)}{m(x)}, \quad m(x) = \int_{\theta\in\Theta} \pi(\theta)f_\theta(x)d\theta$$

se denomina ***función de densidad/probabilidad a posteriori para θ.***

Ejemplo: Sea $X_1, \ldots, X_n$ una muestra aleatoria de una variable con distribución Bernoulli(θ), donde la creencia sobre θ corresponde a una distribución Beta(α, β). Determinemos la distribución a posteriori de θ.

Tenemos que la distribución de existo de las n variables sigue una distribución Binomial(n, θ), por lo que,

$$f_\theta(x) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$$

Por otro lado, la creencia sobre θ, sigue una distribución Beta(α, β), por lo que,

$$\pi(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

donde $\theta \in (0, 1)$. Ahora nos falta determinar $m(x)$. Por definición tenemos que estudiar la intgral,

$$\int_0^1 \binom{n}{x}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{(\alpha+x)-1}(1-\theta)^{(n+\beta-x)-1}d\theta$$

Observemos que la binomial de n sobre x con las funciones gammas son cosntante, por lo que podemos sacar la integral, obteniendo,

$$\binom{n}{x}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\int_0^1 \theta^{(\alpha+x)-1}(1-\theta)^{(n+\beta-x)-1}d\theta$$

Ahora, observemos que estamso integrando una parte de una función densidad g_θ de la distribución Gamma($\alpha + x, n + \beta - x$), la cual además cumple que,

$$\int_0^1 \frac{\Gamma(\alpha+x+n+\beta-x)}{\Gamma(\alpha+x)\Gamma(n+\beta-x)}\theta^{(\alpha+x)-1}(1-\theta)^{(n+\beta-x)-1}d\theta = 1$$

Por tanto,

$$\pi(\theta|x) = \frac{\Gamma(\alpha+x+n+\beta-x)}{\Gamma(\alpha+x)\Gamma(n+\beta-x)}\theta^{(\alpha+x)-1}(1-\theta)^{(n+\beta-x)-1}$$

Es decir, $\pi(\theta|x)$ tiene distribución Gamma$(\alpha+x, n+\beta-x)$.

Un estimador natural para θ es la media a posteriori o media de la distribución a posteriori,

$$\begin{aligned}\hat{\theta}_B &= \frac{x+\alpha}{x+\alpha+n-x+\beta}\\ &= \frac{x+\alpha}{\alpha+\beta+n}\\ &= \frac{\sum_{i=1}^n x_i+\alpha}{\alpha+\beta+n}\end{aligned}$$

Definición: *Dado un parámetro θ con función de densidad/probabilidad a posteriori $\pi(\theta|x)$, su estimador de Bayes corresponde a:*

$$\hat{\theta}_B := \mathbb{E}(\theta|x) = \int_{\theta\in\Theta}\theta\pi(\theta|x)d\theta$$

Ejemplo: Consideremos el ejemplo anterior. Supongamos que la distribución a priori para θ corresponde a una distribución Beta$(2, 1)$. Determinemos el estimador de Bayes de θ si se observar 8 éxitos en un total de 10 repeticiones del experimento.

Definición: *Sea $\mathcal{F}_\theta$ una clase de distribución para una variable aleatoria X. Una clase de distribuciones a prioir para θ, Π se denominda una* ***familia de distribuciones conjugadas*** *para $\mathcal{F}_\theta$ so y sólo si la distribución a posteriori para θ está en la clase Π, para todas las distribuciones $f \in \mathcal{F}_\theta, \pi \in \Pi, x \in \Omega$.*

Ejemplo: Sea X una variable aleatoria X con función densidad Normal(θ, σ^2) con σ^2 conocido y una distribución a priori Normal(μ, τ^2) para θ. Encontremos el estimador de Bayes de θ.

Para ello necesitamos determinar la función densidad de $\pi(\theta|x)$. Observemos que,

$$\begin{aligned}\pi(\theta) &= \frac{1}{\sqrt{2\pi\tau^2}}\exp\left(-\frac{1}{2\tau^2}(\theta-\mu)^2\right)\\ f_\theta(x) &= \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2\sigma^2}(x-\theta)^2\right)\end{aligned}$$

Entonces,

$$\pi(\theta)f_\theta(x) = \frac{1}{\sqrt{2\pi\tau^2}}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2}\left(\frac{(\theta-\mu)^2}{\tau^2}+\frac{(x-\theta)^2}{\sigma^2}\right)\right)$$

Queremos poder expresar todo esto con respecto a una densidad de una Normal(a, b), para ello basta estudiar la exponencial, en particular,

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

Donde C es una constante que surge al reescribir la expresión. Lo importante es que se obtiene que tenemos presencia de una distribución Normal$\left(\frac{x\tau^2+\mu\sigma^2}{\tau^2+\sigma^2}, \sigma^2\tau^2/(\sigma^2+\tau^2)\right)$. De hecho haciendo las cuentas con $m(x)$ y la misma técnica usada en los ejemplos anteriores, se obtiene que,

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

Finalmente podemos determinar $\hat{\theta}_B = \mathbb{E}(\pi(\theta|x))$, es decir, $\hat{\theta}_B = x\frac{\tau^2}{\tau^2+\sigma^2} + \mu\frac{\sigma^2}{\tau^2+\sigma^2}$.

3. Calidad de un Estimador

3.1. Error Cuadrádito Medio

Definición: *El* ***error cuadrático medio*** *de un estimador $\hat{\theta}$ de θ corresponde a:*

$$\begin{aligned} ECM_\theta(\hat{\theta}) &= \mathbb{E}_\theta\left((\hat{\theta}-\theta)^2\right) \\ &= Var(\hat{\theta}) + (Sesgo_\theta(\hat{\theta}))^2 \end{aligned}$$

donde $Sesgo_\theta(\hat{\theta}) = \mathbb{E}(\hat{\theta}) - \theta$

Observación: Sea la variable aleatoria $X = \hat{\theta} - \theta$. Si,

$$\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$$

Entonces,

$$\begin{aligned} \text{ECM}_\theta(\hat{\theta}) &= \text{Var}(\hat{\theta}-\theta) + \left(\mathbb{E}(\hat{\theta}-\theta)\right)^2 \\ &= \text{Var}(\hat{\theta}) + \text{Var}(\theta) + \left(\mathbb{E}(\hat{\theta}-\theta)\right)^2 \\ &= \text{Var}(\hat{\theta}) + \left(\mathbb{E}(\hat{\theta}-\theta)\right)^2 \end{aligned}$$

Ejemplo: Sea $X_1, \dots, X_n$ una muestra aleatoria con distribución Normal(θ, σ^2). Consideremos los estimadores para σ^2:

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

Determinemos el mejor estimador en términso del error cuadrático medio.

- **Primer estimador:** Necesitamos conocer la esperanza y la varianza de S^2. Observemos que,

$$\frac{(n-1)}{\sigma^2}S^2 \sim \mathcal{X}^2_{(n-1)}$$

Entonces,

$$\mathbb{E}\left(\frac{(n-1)}{\sigma^2}S^2\right) = n-1 \Longleftrightarrow \mathbb{E}(S^2) = \sigma^2$$

Es decir, S^2 es un estimador insesgado. Por lo que,

$$\text{Sesgo}_{\sigma^2}(S^2) = \sigma^2 - \sigma^2$$

Por tanto,

$$\text{ECM}_{\sigma^2}(S^2) = \text{Var}(S^2)$$

Sabemos que,

$$\text{Var}\left(\frac{(n-1)}{\sigma^2}S^2\right) = 2(n-1) \Longleftrightarrow \text{Var}(S^2) = \frac{2\sigma^4}{(n-1)}$$

Por tanto,

$$\text{ECM}_{\sigma^2}(S^2) = \frac{2\sigma^4}{(n-1)}$$

- **Segundo estimador:**

Ejemplo: Sea $X_1, \ldots, X_n$ una muestro aleatoroa con distribución Bernoulli(θ) y considere los estimadores de máxima verosimilitud y de Bayes bajo priori Beta para θ. Comparemos estos estimadores en términos de error cuadrático medio.

Determinemos la máxima verosimilitud. Por definición,

$$\begin{aligned} L(\theta, x) &= \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} \\ &= \theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i} \end{aligned}$$

Derivando con respecto a θ, obtenemos que,

$$\theta^{\sum_{i=1}^n x_i - 1}(1-\theta)^{(n-1)-\sum_{i=1}^n x_i}\left(\sum_{i=1}^n x_i + \left(n - 2\sum_{i=1}^n x_i\right)\theta\right)$$

Igualando a 0 obtenemos que,

$$\hat{\theta} = \frac{\overline{X}}{2\overline{X}-1}$$

Faltan argumentos.

- **Primer estimador:** Observemos que $\overline{X}$ es un estadístico suficiente minimal completo y que $\hat{\theta}$ es ancilar, de forma que,

$$\mathbb{E}(\overline{X}) = \mathbb{E}\left(\frac{\overline{X}}{2\overline{X}-1}\right)\mathbb{E}(2\overline{X}-1) \Longleftarrow \mathbb{E}\left(\frac{\overline{X}}{2\overline{X}-1}\right) = \frac{\theta}{2\theta-1}$$

terminar

3.2. Mejores Estimadores Insesgados

Consideremos un estimador $\theta, T^*(X)$ tal que,

$$\mathbb{E}_\theta(T^*(X)) = \tau(\theta)$$

Considere el conjunto:

$$C_\tau = \{T(X) : \mathbb{E}_\theta(T^*(X)) = \tau(\theta)\}$$

Entonces para todo $T_1(X), T_2(X) \in C_\tau$ se cumple que:

$$\mathrm{ECM}_\theta(T_1(X)) - \mathrm{ECM}_\theta(T_2(X)) = \mathrm{Var}_\theta(T_1(X)) - \mathrm{Var}_\theta(T_2(X))$$

Definición: *Un estiamdor $W(X)$ es mejor estimador insesgado si y sólo si,*

$$\mathbb{E}_\theta(W(X)) = \tau(\theta)$$

y, para todo $T(X)$ tal que $\mathbb{E}_\theta(T(X)) = \tau(\theta)$ se cumple que,

$$Var_\theta(W(X)) \leq Var_\theta(TX)$$

para todo $\theta \in \Theta$. El estimador $W(X)$ se dice también ***insesgado y de varianza uniformemente mínima*** *o UMVUE de $\tau(\theta)$.*

Teorema (Cramér-Rao): *Sea $X_1, \ldots, X_n$ una muestra aleatoria con función de densidad conjunta $f_\theta(x)$ y sea $T(X)$ un estimador tal que $\mathbb{E}_\theta(T(X))$ es diferenciable como función θ. Supongamso que la densidad de X satisface:*

$$\frac{d}{d\theta} \int \cdots \int h(x) f_\theta(x) dx = \int \cdots \int h(x) \frac{d}{d\theta} f_\theta(x) dx$$

para toda función h con $\mathbb{E}_\theta |h(X)| < \infty$. Entonces,

$$Var_\theta(T(X)) \geq \frac{\left(\frac{d}{d\theta}\mathbb{E}_\theta(T(X))\right)^2}{\mathbb{E}_\theta\left(\frac{d}{d\theta}\log(f_\theta(X))\right)^2}$$

Corolario: *Sea $X_1, \ldots, X_n$ una muestra aleatoria proveniente de una distribución con función de densidad $f_\theta(x)$, y sea $T(X)$ un estimador donde $\mathbb{E}_\theta(T(X))$ es diferenciable como función de θ. Si la densidad conjunta $\prod_{i=1}^n f_\theta(x_i)$ cumple con las condiciones del Teorema de Cramér-Rao, entonces:*

$$Var_\theta(T(X)) \geq \frac{\left(\frac{d}{d\theta}\mathbb{E}_\theta(T(X))\right)^2}{\mathbb{E}_\theta\left(\frac{d}{d\theta}\log(f_\theta(X))\right)^2}$$

Ejemplo: Consideremos $X_1, \ldots, X_n$ una muestra aleatoria de distribución Poisson(θ). Probemos que $\bar{X}$ es el mejor estimador insesgado de θ.

Definición: *Sea $X_1, \ldots, X_n$ una muestra aleatoria de una distribución f_θ y sea $\hat{\theta}$ el estimador máximo de verosímil de θ. Entonces,*

$$\lim_{n\to\infty} \mathbb{E}\left((\hat{\theta}-\theta)^2\right) = 0$$

(En este caso, se dice que el estimador máximo verosímil es consistente para estimar θ.)

Definición: *Sea $X_1, \ldots, X_n$ una muestra aleatoria de una distribución con función densidad (probabilidad) f_θ y sea $\hat{\theta}$ el estimador máximo verosímil de θ. Entonces, la distribución de:*

$$\sqrt{n}(\hat{\theta}-\theta)$$

converge a una distribución:

$$Normal\left(0, \frac{1}{\mathbb{E}_\theta\left(\frac{d}{d\theta}\log(f_\theta(X))\right)^2}\right)$$

3.3. Estimación Intervalar

Definición: *Una **estimación intervalar** para un parámetro real valorado θ corresponde a cualquier par de funciones $L(x_1, \ldots, x_n)$ y $U(x_1, \ldots, x_n)$ de una muestra $x = (x_1, \ldots, x_n)$, que satisface:*

$$L(x) \leq U(x)$$

*para todo $x \in \Omega$. El intervalo aleatorio $[L(X), U(X)]$ se denomina un **estimador intervalar** para θ.*

Definición: *Para un estimador intervalar $[L(X), U(X)]$ de un parámetro θ, se define su **probabilidad de cobertura** como la probabilidad de contener al verdadero valor parámetro θ, es decir:*

$$\mathbb{P}_\theta(\theta \in [L(X), U(X)])$$

Ejemplo: Sea $X_1, \ldots, X_4$ una muestra aleatoria con distribución Normal$(\mu, 1)$. El intervalo de confianza para μ es $[\overline{X}-1, \overline{X}+1]$ donde $\overline{X} \sim$ Normal$(\mu, 1/4)$. Se tiene que IC (intervalo de confianza) es menos preciso que el estimador puntual, sin embargo se gana seguridad en la estimación.

Sabemos que $\mathbb{P}(\overline{X} = \mu) = 0$ (probabilidad puntual en una distribución continua). Pero en un IC se tiene que,

$$\begin{aligned}
\mathbb{P}(\mu \in [\overline{X}-1, \overline{X}+1]) &= \mathbb{P}(\overline{X}-1 \leq \mu \leq \overline{X}+1) \\
&= \mathbb{P}(-1 \leq \mu - \overline{X} \leq 1) \\
&= \mathbb{P}\left(\frac{-1}{\sqrt{1/4}} \leq \frac{\overline{X}-\mu}{\sqrt{1/4}} \leq \frac{1}{\sqrt{1/4}}\right) \\
&= \mathbb{P}\left(-2 \leq \frac{\overline{X}-\mu}{\sqrt{1/4}} \leq 2\right) \\
&= \mathbb{P}(-2 \leq Z \leq 2)
\end{aligned}$$

donde $Z = (\overline{X} - \mu)/\sqrt{1/4} \sim \text{Normal}(0,1)$. Entonces,

$$\begin{aligned}\mathbb{P}(\mu \in [\overline{X} - 1, X + 1]) &= \mathbb{P}(-2 \leq Z \leq 2)\\ &= \Phi(2) - \Phi(-2)\\ &= 0{,}9544\end{aligned}$$

Ejemplo: Sea $X_1, \ldots, X_n$ una muestra aleatoria proveniente de una distribución Uniforme$(0, \theta)$. Suponga que interesa un intervalo de confianza para θ, y se proponen los intervalos:

$$\begin{aligned}&[aX_{(n)}, bX_{(n)}], && 1 \leq a \leq b,\\ &[X_{(n)} + c, X_{(n)} + d], && 0 \leq c < d\end{aligned}$$

Determinemos la probabilidad de cobertura de cada uno de los intervalos propuestos.

- **Primer intervalo:** Observemos que,

$$\begin{aligned}\mathbb{P}(\theta \in [aX_{(n)}, bX_{(n)}]) &= \mathbb{P}(aX_{(n)} \leq \theta \leq bX_{(n)})\\ &= \mathbb{P}\left(\frac{\theta}{b} \leq X_{(n)} \leq \frac{\theta}{a}\right)\end{aligned}$$

Observemos que $X_{(n)}$ tiene función densidad $nF^{n-1}(t)f(t)$ donde $F(t) = t/\theta, f(t) = 1/\theta$ donde $t \in (0, \theta)$. Por otro lado notemos que,

$$0 < \frac{\theta}{b} \leq \frac{\theta}{a} < \theta$$

por como hemos escogido a, b. Entonces,

$$\begin{aligned}\mathbb{P}\left(\frac{\theta}{b} \leq X_{(n)} \leq \frac{\theta}{a}\right) &= \int_{\theta/b}^{\theta/a} \frac{n}{\theta^n} t^{n-1} dt\\ &= \left(\frac{1}{a}\right)^n - \left(\frac{1}{b}\right)^n\end{aligned}$$

- **Segundo intervalo:** Observemos que,

$$\begin{aligned}\mathbb{P}(\theta \in [X_{(n)} + c, X_{(n)} + d]) &= \mathbb{P}(X_{(n)} + c \leq \theta \leq X_{(n)} + d)\\ &= \mathbb{P}(\theta - d \leq X_{(n)} \leq \theta - c)\\ &= \int_{\theta-d}^{\theta-c} \frac{n}{\theta^n} t^{n-1} dt\\ &= \left(1 - \frac{c}{\theta}\right)^n - \left(1 - \frac{d}{\theta}\right)^n\end{aligned}$$

Observemos que la probabilidad de cobertura depende del parámetro θ.

Definición: *Para un estimador intervalar $[L(X), U(X)]$ de un parámetro θ, se define su* ***coeficiente de confianza*** *como el ínfimo de las probabilidades de cobertura, es decir:*

$$\inf_{\theta} \mathbb{P}_{\theta} \in [L(X), U(X)]$$

Ejemplo: En el problema anterior, encuentre el coeficiente de confianza de cada estimador propuesto.

- **Primer intervalo:** Del ejemplo anterior, sabemos que la probabilidad de cobertura no depende de θ, por lo tanto, el coeficiente de confianza es,

$$\left(\frac{1}{a}\right)^n - \left(\frac{1}{b}\right)^n$$

- **Segundo intervalo:** Observemos que la probabilidad de cobertura depende de θ. Además el ínfimo se alcanza cuando $\theta \to \infty$, entonces,

$$\lim_{\theta \to \infty} \left(1 - \frac{c}{\theta}\right)^n - \left(1 - \frac{d}{\theta}\right)^n = 0$$

Definición: *Para un parámetro cualquiera θ, se define un* ***conjunto de confianza*** $1-\alpha$ *como un conjunto $C(X)$ tal que:*

$$\inf_{\theta} P_{\theta}(\theta \in C(X)) = 1 - \alpha$$

Definición: *Una variable aleatoria $\mathcal{Q}(X, \theta)$ corresponde a un pivote para θ si y sólo si su distribución no depende de ningún parámetro desconocido. Es decir, la distribución de $\mathcal{Q}(X, \theta)$ es la misma para todo θ.*

Ejemplo: Determinemos los pivotes de las siguientes situaciones

- Familia de distribuciones de localización, con parámetro de localización θ.
- Familia de distribuciones de escala, con parámetro de escala σ.
- Familia de distribución de localización y escala, con parámetro θ y σ.

Pivotes:

- Sea Z una variable aleatoria con función densidad $f(z)$ sin parámetros. Digamos que Z tiene distribución Normal$(0, 1)$. Una familia de localización tiene la siguiente forma $f(x|\theta) = f(x-\theta)$ donde $X = Z+\theta$. Entonces $X \sim$ Normal$(\theta, 1)$, en particular, $X-\theta = Z$ de forma que es un pivote. Veamos como se distribuye $\overline{X} - \theta$ con respecto a $X_i = Z_i + \theta$ donde Z_i tiene la misma distribución de Z. Entonces,

$$\begin{aligned}
\overline{X} - \theta &= \frac{1}{n}\sum_{i=1}^{n}(Z_i + \theta) - \theta \\
&= \frac{1}{n}\sum_{i=1}^{n} Z_i \\
&= \overline{Z}
\end{aligned}$$

Entonces $\overline{X} = \overline{Z} + \theta$. Entonces $\overline{X} - \theta$ es una función de los datos y el parámetro cuya distribución no depende de ningún parámetro desconocido.

- **Distribución de escala:** Sea $X_1, \ldots, X_n$ una muestra aleatoria con densidad en la familia de escala $Z \sim f(z)$, dada por $X = \sigma Z$, entonces,

$$f_Z(x/\sigma) = \frac{1}{\sigma} f(z)$$

Un pivote $\mathcal{Q}(x, \sigma)$ es una función cuya distribución no dependa de σ. Haciendo el mismo truco anterior obtenemos que,

$$\frac{1}{\sigma}\overline{X} = \overline{Z}$$

Por lo tanto, la distribución $\frac{1}{\sigma}\overline{X}$ es libre y por tanto es un pivote para σ.

- **Familia de localización y escala:** Sea $X_1, \ldots, X_n$ una muestra aleatoria, sea $Z \sim f(z)$ tal que $X = \sigma Z + \theta$. Entonces,

$$\frac{X - \theta}{\sigma} = Z$$

Observamos que,

$$\begin{aligned} \frac{\overline{X} - \theta}{\sigma} &= \frac{1}{\sigma}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \theta\right) \\ &= \overline{Z} \end{aligned}$$

Ejemplo Encuentre un pivote para θ en la distribución Exponencial de media θ.

Sea $X_1, \ldots, X_n$ una muestra con distribución Exp$(1/\theta)$, es decir,

$$f(x_i) = \frac{1}{\theta} e^{-x_i/\theta}$$

para todo $i = 1, \ldots, n$. Sea $T = X_1 + \cdots + X_n \sim \text{Gamma}(n, 1/\theta)$. Entonces,

$$\begin{aligned} f_T(t) &= \frac{(1/\theta)^n}{\Gamma(n)} t^{n-1} e^{-t/\theta} \\ &= \frac{1}{\theta^n (n-1)!} t^{n-1} e^{-t/\theta} \end{aligned}$$

para $t > 0$. Sea $Z = T/\theta$, luego tenemos que $g(t) = w = t/\theta$ donde $g^{-1}(t) = w\theta = t$, que es diferenciable, luego,

$$f_W(w) = \frac{1}{(n-1)!} w^{n-1} e^{-w} \sim \text{Gamma}(n, 1)$$

awdawd terminar

Caso General: Considere un estadístico $T(X)$ con función de densidad f_θ. En general, se busca factorizar su densidad en la forma:

$$f_\theta(t) = g(\mathcal{Q}(t,\theta)) \left| \frac{\partial \mathcal{Q}(t,\theta)}{\partial t} \right|$$

con $\mathcal{Q}$ función monótona en t, para cada valor de θ.

Para encontrar un intervalo de $(1-\alpha)100\,\%$ de confianza, es posible plantear la igualdad:

$$P_\theta(a \leq \mathcal{Q}(t,\theta) \leq) \geq 1 - b$$

Ejemplo: Aplique la metodología anterior al caso Exponencial de media θ.

Teorema: *Sea $T(X)$ una **variable aleatoria continua** con función de distribución F_θ. Sea $0 < \alpha < 1$ un valor constante. Para todo valor $t \in \mathcal{T}$, se define $\theta_L(t)$ y $\theta_U(t)$ tales que:*

1. *Si F_θ es decreciente en θ para cada t:*

$$F_{\theta_{U(t)}}(t) = \frac{\alpha}{2} \qquad F_{\theta_{L(t)}}(t) = 1 - \frac{\alpha}{2}$$

2. *Si F_θ es creciente en θ para cada t:*

$$F_{\theta_{U(t)}}(t) = 1 - \frac{\alpha}{2} \qquad F_{\theta_{L(t)}}(t) = \frac{\alpha}{2}$$

Entonces $[\theta_L(t), \theta_U(t)]$ corresponde a un intervalo de confianza $(1-\alpha)100\,\%$ para θ.

Ejemplo: Considere una distrubición Exponencial(1) con parámetro de localización θ. Utilice la metodología anterior para construir un intervalo de confianza para θ.

Teorema: *Sea $T(X)$ una **variable aleatoria discreta** con función de distribución F_θ. Sea $0 < \alpha < 1$ un valor constante. Para todo valor $t \in \mathcal{T}$, se definen $\theta_L(t)$ y $\theta_U(t)$ tales que:*

- *Si F_θ es decreciente en θ para cada t:*

$$P_{\theta_{U(t)}}(T \leq t) \leq \frac{\alpha}{2} \qquad P_{\theta_{L(t)}}(T \geq t) = \frac{\alpha}{2}$$

- *Si F_θ es creciente en θ para cada t:*

$$P_{\theta_{U(t)}}(T \geq t) \leq \frac{\alpha}{2} \qquad P_{\theta_{L(t)}}(T \leq t) = \frac{\alpha}{2}$$

Entonces $[\theta_L(t), \theta_U(t)]$ corresponde a un intervalo de confianza $(1-\alpha)\,\%$ para θ.

Ejemplo: Sea $X_1, \ldots, X_n$ una muestra aleatoria de una distrubición Poisson(θ). Utilice que, para todo $t \in \mathbb{R}$:

$$P_\theta\left(\sum_{i=1}^{n} X_i \leq t\right) = P(\mathcal{X}_{2(}t+1)^2 \geq 2n\theta)$$

Para obtener un intervalo de $(1-\alpha)100\,\%$ para θ.

4. Test de Hipótesis

Definición: *Una **hipótesis** es una afirmación sobre un parámetro poblacional.*

Definición: *Dos hipótesis complementarias en un problema de testeo de hipótesis se denominan **hipótesis nula** e **hipótesis alternativa**, H_0 y H_1, respectivamente.En general, escribimos,*

$$H_0 : \theta \in \Theta_0, \quad H_1 : \theta \in \Theta_0^c$$

Definición: *Un procedimiento de testeo de hipótesis es una **regla de decisión** que especifica para qué valores de la muestra la decisión es aceptar H_0 como verdadera y para qué valores de la muestra la decisión es aceptar H_1 como verdadera.*

Definición: *El subconjunto del espacio muestral para el cual H_0 es rechazada se denomina **región de rechazo o crítica**. Su complemento se denomina **región de aceptación**.*

Ejemplo: Sea X_i la variable aleatoria que cuenta la cantidad de automóviles que pasan por una esquina en una hora. Consideremos una muestra aleatoria $X_1, \dots, X_n$ con distribución Poisson(θ). Queremos estudiar $\theta > 800$, formulando las hipótesis,

$$\begin{aligned} H_0 :&\theta \leq 800 \\ H_1 :&\theta > 800 \end{aligned}$$

¿Para qué valores de $\overline{X}$ rechazamos H_0 y nos quedamos con H_1? La afirmación es que todas las muestras tales que $\overline{X} > 850$, hacen rechazar H_0 y se concluye $\theta > 800$.

4.1. Test de Razón de Verosimilitud

Definición *El **estadístico del Test de Razón de verosimilitud** para testear,*

$$H_0 : \theta \in \Theta_0 \; vs \; H_1 : \theta \in \Theta_0^c$$

basado en una muestra $X = x$, corresponde a:

$$\lambda(x) = \frac{\sup_{\theta \in \Theta_0} L(\theta, x)}{\sup_{\theta \in \Theta} L(\theta, x)}$$

Definición: *Un **test de razón de verosimilitud** (TRV) es cualquier test cuya región de rechazo es de la forma:*

$$\{x \in \Omega : \lambda(x) \leq c\}, \quad 0 \leq c \leq 1$$

Ejemplo: Sea $X_1, \dots, X_n$ una muestra aleatoria con distribución Bernoulli(θ) y la hipótesis:

$$H_0 : \theta = \theta_0, \quad H_1 : \theta \neq \theta_0$$

Encuentre la forma de la región de rechazo de un TRV.

Sabemos que,

$$\begin{aligned} L(\theta, x) &= \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} \\ &= \theta^S(1-\theta)^{n-S} \end{aligned}$$

donde $S = x_1 + \cdots + x_n$. Dado la hipótesis $H_0 : \theta = \theta_0$, se tiene que,

$$\sup_{\theta \in \Theta_0} L(\theta, x) = L(\theta_0, x) = \theta_0^S (1-\theta)^{n-S}$$

Ahora notemos que $L(\theta, x)$ con x fijo, está definido para $\theta \in [0, 1]$, entonces el supremo se alcanza y es el máximo, en particular, el máximo se alcanza en $\theta = S/n \in [0, 1]$. Por tanto,

$$\lambda(x) = \frac{\theta_0^S (1-\theta)^{n-S}}{(S/n)^S (1-S/n)^{n-S}}$$

Por tanto, una región de rechazo de un TRV son los $x \in \Omega$ tales que,

$$\frac{\theta_0^S (1-\theta)^{n-S}}{(S/n)^S (1-S/n)^{n-S}} \leq c$$

Ejemplo: Sea $X_1, \ldots, X_n$ una muestra aleatoria con distribución Normal$(\theta, 1)$, y sean las hipótesis:

$$H_0 : \theta = \theta_0, \quad H_1 : \theta \neq \theta_0$$

Determinemos la forma de la región de rechazo de un TRV.

Se tiene que,

$$\begin{aligned} L(\theta, x) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x_i - \theta)^2\right) \\ &= \left(\frac{1}{2\pi}\right)^{n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i - \theta)^2\right) \end{aligned}$$

Observemos que $L(\theta, x)$ alcanza su máximo en $\theta = S/n = (x_1 + \cdots + x_n)/n$, entonces,

$$\lambda(x) = \exp\left(-\frac{n}{2}\left(\frac{S}{n} - \theta_0\right)^2\right)$$

Luego para $c \neq 0$ se tiene que,

$$\exp\left(-\frac{n}{2}\left(\frac{S}{n} - \theta_0\right)^2\right) \leq c \iff \overline{x} \in \left[\frac{2}{n}\log c + \theta_0, -\frac{2}{n}\log c + \theta_0\right]$$

Luego la región de rechazo son los $x = (x_1, \ldots, x_n)$ tales que están en el intervalo anterior.

Ejemplo: Sea $X_1, \ldots, X_n$ una muestra aleatoria con función densidad:

$$f_\theta(x) = \begin{cases} e^{-(x-\theta)}, & x \geq \theta \\ 0, & \text{e.o.c} \end{cases}$$

con $\theta \in \mathbb{R}$. Y sean las hipótesis:

$$H_0 : \theta \leq \theta_0, \quad H_1 : \theta > \theta_0$$

Encuentre la forma de la región de rechazo de un TRV.

Sea tiene que,

$$\begin{aligned} L(\theta, x) &= \prod_{i=1}^{n} e^{-(x_i-\theta)} \mathbb{1}_{\{x_i \geq \theta\}} \\ &= e^{-(S-n\theta)} \mathbb{1}_{\{x_{(1)} \geq \theta\}} \end{aligned}$$

con $S = x_1 + \cdots + x_n$. Para $x = (x_1, \ldots, x_n)$ fijos, se tiene que $L(\theta, x)$ alcanza máximo en $\theta = x_{(1)}$, meintras que para $\theta > x_{(1)}$, la función de verosimilitud se anula. De forma que,

$$\sup_{\theta \in \mathbb{R}} L(\theta, x) = L(x_{(1)}, x) = e^{-(S-nx_{(1)})}$$

Ahora veamos que pasa si $\theta \leq \theta_0$, si $\theta_0 \leq x_{(1)}$ entonces,

$$\sup_{\theta \leq \theta_0} L(\theta, x) = L(\theta_0, x)$$

Si $\theta_0 > x_{(1)}$, entonces,

$$\sup_{\theta \leq \theta_0} L(\theta, x) = L(x_{(1)}, x)$$

Por lo tanto se tiene lo siguiente,

$$\lambda(x) = \begin{cases} 1, & \theta_0 > x_{(1)} \\ e^{n(\theta_0 - x_{(1)})}, & \theta_0 \leq x_{(1)} \end{cases}$$

Luego la región de rechazo depende de la hipótesis que se tome.

$$\text{Región de rechazo} = \begin{cases} \{x \in \Omega : 1 \leq c\}, & \theta_0 > x_{(1)} \\ \{x \in \Omega : & \end{cases}$$

terminar

Ejemplo: Sea $X_1, \ldots, X_n$ una muestra aleatoria con distribución Normal(μ, σ^2) y considere las hipótesis:

$$H_0 : \mu \leq \mu_0, \quad H_1 : \mu > \mu_0$$

con σ^2 desconocido. Determinemos una forma de región de rechazo de un TRV.

Se tiene que la función de similitud es,

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

Sabemos que el máximo se alcanza en $\mu = \overline{x}$, el problema es estudiar el supremo con respecto a la hipótesis $\mu \leq \mu_0$. Observemos lo siguientes,

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

Entonces,

$$\frac{\partial L(\mu, x)}{\partial \mu} > 0, \quad \mu < \overline{x}$$
$$\frac{\partial L(\mu, x)}{\partial \mu} < 0, \quad \mu > \overline{x}$$

Es decir, $L(\mu, x)$ es creciente para todo $\mu < \overline{x}$ y decreciente para todo $\mu > \overline{x}$, de esta forma,

$$\lambda(x) = \begin{cases} \exp\left(-\frac{n}{2\sigma^2}(\overline{x} - \mu_0)\right), & \mu < \overline{x} \\ 1, & \mu \geq \overline{x} \end{cases}$$

terminar

TRV y Suficiencia: Sea $X_1, \ldots, X_n$ una muestra aleatoria con distribución función densidad f_0. Si $T(X)$ es un estadístico suficiente para θ y $\lambda(x)$ y $\lambda^*(t)$ corresponden a los estadísticos de los test de Razón de verosimilitud basados en X y $T(X)$, respectivamente, entonces:

$$\lambda^*(T(x)) = \lambda(x)$$

para todo $x \in \Omega$.

4.2. Métodos para Evaluar Test de Hipótesis

Definición: *Supongamos que queremos testear las hipótesis:*

$$H_0 : \theta \in \Theta_0, \quad vs \quad H_1 : \theta \in \Theta_0^C$$

Cuando se rechaza H_0, siendo que esta hipótesis es verdadera, se dice que está en presencia de ***Error de Tipo I****. La probabilidad de cometer este error corresponde a:*

$$P_\theta(X \in R)$$

para $\theta \in \Theta_0$, donde R corresponde a la región crítica del test. Por otro lado, cuando no se rechaza H_0, siendo que esta hipótesis es falsa, se dice que está en presencia de ***Error de Tipo II****. La probabilidad de cometer este error corresponde a:*

$$1 - P_\theta(X \in R)$$

con $\theta \in \Theta_0^c$

Definición: *La* ***función de potencia*** *de un test de hipótesis con región crítica R corresponde a:*

$$\beta(\theta) = P_\theta(X \in R)$$

con $\theta \in \Theta$.

Ejemplo: Consideremos una variable aleatoria X con distribución Binomial$(5, \theta)$. Interesa testear las hipótesis:

$$H_0 : \theta \leq 1/2, \quad \text{vs} \quad H_1 : \theta > 1/2$$

Compare las funciones de potencia de las reglas de rechazo con regiones críticas:

$$R_1 = \{5\}$$

y

$$R_2 = \{3, 4, 5\}$$

- $R_1 = \{5\}$**:** Por definición,

  $$\beta(\theta) = \mathbb{P}_\theta(X = 5) = \theta^5$$

  donde $\theta \in (0, 1)$.

- $R_2 = \{3, 4, 5\}$**:** Por definición,

  $$\begin{aligned} \beta(\theta) &= \mathbb{P}_\theta(X \in \{3, 4, 5\}) \\ &= \mathbb{P}_\theta\left(X = \{3\} \cup \{4\} \cup \{5\}\right) \\ &= \sum_{i=3}^{5} \mathbb{P}_\theta(X = i) = \sum_{i=3}^{5} \binom{5}{i} \theta^i (1-\theta)^{5-i} \end{aligned}$$

Ejemplo: Considere una muestra aleatoria $X_1, \ldots, X_n$ proveniente de una distribución Normal(θ, σ^2), con σ^2 conocido. El test de razón de verosimilitud para las hipótesis:

$$H_0 : \theta \leq \theta_0, \quad \text{vs} \quad H_1 : \theta > \theta_0$$

posee región crítica de la forma:

$$R(X) = \left\{ X; \frac{\overline{X} - \theta_0}{\sigma/\sqrt{n}} \geq k \right\}$$

con k un número positivo. Obtenga su función de potencia cuando $k = 1,28$.

Observemos que,

$$\beta(\theta) = \mathbb{P}_\theta(X \in R(X)) = \mathbb{P}_\theta\left(\frac{\overline{X} - \theta_0}{\sigma/\sqrt{n}} \geq k \right)$$

Sabemos que $\overline{X}$ tiene distribución Normal$(\theta, \sigma^2/n)$, entonces,

$$\begin{aligned} \mathbb{P}_\theta\left(\frac{\overline{X} - \theta_0}{\sigma/\sqrt{n}} \geq k \right) &= \mathbb{P}_\theta\left(\frac{\overline{X} - \theta}{\sigma/\sqrt{n}} \geq k + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}} \right) \\ &= \int_{k + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}}}^{\infty} \frac{1}{2\pi} \exp\left(-\frac{1}{2}x^2 \right) dx \end{aligned}$$

Ejemplo: Con respecto al ejemplo anterior supongamos que,

$$\max_{\theta \in \Theta_0} P_\theta(\text{Error tipo I}) = 0,1$$
$$\max_{\theta > \Theta_0 + \sigma} P_\theta(\text{Error tipo II}) = 0,2$$

Encuentre el tamaño de muestra necesario para cumplir estas condiciones.

Definición: *Se dice que un test con función de potencia $\beta(\theta)$ es de* ***tamaño*** *α con $0 \leq \alpha \leq 1$ si y sólo,*

$$\sup_{\theta \in \Theta_0} \beta(\theta) = \alpha$$

Definición: *Se dice que un test con función de potencia $\beta(\theta)$ es de* ***nivel*** *α con $0 \leq \alpha \leq 1$ si y sólo si,*

$$\sup_{\theta \in \Theta_0} \beta(\theta) \leq \alpha$$

Ejemplo: Considere una muestra aleatoria de tamaño n proveniente de una población Normal$(\theta, 1)$ junto a las hipótesis:

$$H_0 : \theta = \theta_0, \quad \text{vs} \quad H_1 : \theta \neq \theta_0$$

El test de razón de verosimilitud para estas hipótesis posee región de rechazo de la forma:

$$R(X) = \{X; |\overline{X} - \theta_0| \geq k\}$$

Determinemos el valor de k de modo que el test sea de tamaño α.

Observemos que,

$$R(X) = \{X; \overline{X} - \theta_0 \geq k\} \cup \{X; \overline{X} - \theta_0 \leq k\}$$

Y que $\overline{X}$ tiene distribución Normal$(\theta, 1/n)$. Por lo que,

$$\beta(\theta) = \mathbb{P}_\theta\left(\frac{\overline{X} - \theta}{1/\sqrt{n}} \geq \frac{k + \theta_0 - \theta}{1/\sqrt{n}}\right) + \mathbb{P}_\theta\left(\frac{\overline{X} - \theta}{1/\sqrt{n}} \leq \frac{-k + \theta_0 - \theta}{1/\sqrt{n}}\right)$$

Si consideramos $\theta = \theta_0$, entonces,

$$\begin{aligned}
\sup_{\theta=\theta_0} \beta(\theta) &= \mathbb{P}_\theta\left(\frac{\overline{X} - \theta}{1/\sqrt{n}} \geq \frac{k}{1/\sqrt{n}}\right) + \mathbb{P}_\theta\left(\frac{\overline{X} - \theta}{1/\sqrt{n}} \leq \frac{-k}{1/\sqrt{n}}\right) \\
&= 1 - \int_{-k\sqrt{n}}^{k\sqrt{n}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) dx \\
&= 1 - 2\Phi(-k\sqrt{n})
\end{aligned}$$

Finalmente,

$$k = -\frac{1}{\sqrt{n}}\Phi^{-1}\left(\frac{1-\alpha}{2}\right)$$

Ejemplo: Considere una muestra aleatoria $X_1, \ldots, X_n$ proveniente de una distribución con función de densidad,

$$f_\theta(x) = \begin{cases} \exp(-(x-\theta)), & x \geq \theta \\ 0, & \text{e.o.c.} \end{cases}$$

El test de razón de verosimilitud para las hipótesis:

$$H_0 : \theta \leq \theta_0, \text{ vs } \theta > \theta_0$$

tiene región crítica de la forma:

$$R(X) = \{X; X_{(1)} \geq k\}$$

Determinemos el valor de k de modo que el test sea de tamaño α.

Sabemos que $f_\theta(x) = \exp(-(x-\theta)\mathbb{1}_{\{x\geq\theta\}}(x)$. Observemos que,

$$\beta(\theta) = \mathbb{P}_\theta(X_{(1)} \geq k) = \begin{cases} 1 - e^{-n(k-\theta)}, & k \geq \theta \\ 0, & \text{e.o.c.} \end{cases}$$

Si pensamos en $k \geq \theta$, obtemeos que,

$$\sup_{\theta\leq\theta_0} \beta(\theta) = 1 - e^{-n(k-\theta_0)}$$

Luego al tomar un α obtenemos que,

$$1 - e^{-n(k-\theta_0)} = \alpha \Longleftrightarrow k = \theta_0 - \frac{1}{n}\log(1-\alpha)$$

Vemos que $k \geq \theta$.

Definición: *El* ***valor-p*** *del valor muestra x es el menor valor de α para el cual se rechaza la hipótesis nula.*

Ejemplo: Considere una muestra aleatoria de tamaño n proveniente de una distribución Normal$(\theta, 1)$, junto a las hipótesis:

$$H_0 : \theta = \theta_0, \quad \text{vs} \quad H_1 : \theta \neq \theta_0$$

El test de razón de verosimilud para estas hipótesis rechaza H_0 si y sólo si,

$$|\overline{X} - \theta_0| \geq \frac{1}{\sqrt{n}}\mathcal{Z}_{1-\alpha/2}$$

(El z está raro). Determine el valor-p del test con $n = 1$ y $\theta_0 = 10$, cuando se observa $\overline{X} = 11{,}75$.