# 3 ESTIMACION

## 3.1 Introducción

En un problema estadístico, si los datos fueron generados a partir de una distribución de probabilidad F(x) desconocida, **la Inferencia Estadística** permite decir algo respecto de esta distribución. Cuando se supone que tal distribución no es totalmente desconocida - por ejemplo pertenece a una determinada familia de distribuciones - entonces son desconocidos sólo uno o varios **parámetros** que definen cada distribución de esta familia. En este caso la teoría de estimación tiene por objetivo dar valores a estos parámetros a partir de los valores muestrales.

Por ejemplo, $F(x)$ pertenece a la familia de las distribuciones normales $\mathcal{N}(\mu, 1)$ de varianza igual a 1 y de esperanza desconocida. Aquí $\mu$ es el único parámetro desconocido de la distribución. Pero si se supone la varianza también desconocida, se tendrán dos parámetros, la media $\mu$ y la varianza $\sigma^2$.

Los parámetros son constantes que toman valores en un espacio llamado **espacio de parámetros** $\Theta$:

| | |
|---|---|
| $\mathcal{N}(\mu, 1)$ | $\Theta = I\!R$ |
| $\mathcal{N}(\mu, \sigma)$ | $\Theta = I\!R \times ]0, +\infty[$ |
| $Exp(\beta)$ | $\Theta = ]0, +\infty[$ |
| Binomial(10,p) | $\Theta = [0, 1]$ |

Sean $x_1, ..., x_n$ los valores muestrales obtenidos sobre una muestra aleatoria simple de una v.a. X de función de densidad $f(x/\theta)$, en que $\theta$ es desconocido. Hay varias maneras de decir algo sobre $\theta$. Lo más simple consiste en dar un valor único para $\theta$. Es **la estimación puntual**: se busca elegir un valor para $\theta$ a partir de los valores muestrales. Es decir se tiene que definir una función $\delta : I\!R^n \longrightarrow \Theta$, que es un estadístico llamado **estimador** de $\theta$. El valor tomado por esta función sobre una muestra particular de tamaño n es una **estimación**. Otra forma de estimar un parámetro consiste en buscar no un sólo valor para $\theta$, sino un conjunto de valores, un intervalo en general, en el cual se tiene alta probabilidad de encontrar $\theta$. Es la **estimación por intervalo**.

Procediendo así, tratamos de acercar parámetros, que son considerados como constantes, a partir de estadísticos que son aleatorios. Ahora bien, frecuentemente se sabe algo más sobre los parámetros; este conocimiento obviamente no es preciso, sino no se tendría el problema de estimar estos parámetros; pero se tienen ideas sobre sus posibles valores, que pueden ser traducidas a una **función de distribución a priori** sobre el espacio de parámetro $\Theta$. Los estimadores bayesianos toman en cuenta de la distribución a priori y de los valores muestrales.

El problema es encontrar métodos que permitan construir estos estimadores. A continuación daremos los metodos usuales de estimación puntual.

## 3.2 Método de los Momentos

Vimos en el capítulo anterior que $\bar{x}_n \xrightarrow{c.s.} E(X) = \mu$. Más generalmente si el momento $\mu_r = E(X^r)$ existe, entonces por la ley de los grandes números:

$$m_r = \frac{1}{n}\sum x_i^r \xrightarrow{c.s.} \mu_r \qquad (I\!P(lim_{n\to\infty} = \mu_r) = 1)$$

Luego se puede estimar $\mu_r$ como $\hat{\mu}_r = m_r$.

Ejemplo: este método produce como estimador de la media $\mu$, $\hat{\mu} = \bar{x}_n$ y como estimador de la varianza $\sigma^2 = m_2 - \bar{x}_n^2 = S_n^2$

## 3.3 Método de Máxima Verosimilitud (Fisher 1912)

Sean $X_1, X_2, ..., X_n$ una muestra aleatoria simple de una v.a. de densidad $f(x/\theta)$ en que $\theta \in \Theta$, el espacio de parámetros.

**Definición 3.1** *Se llama* **función de verosimilitud** *a la densidad conjunta del vector de los valores muestrales; para todo vector observado* $\underline{x} = (x_1, x_2, ..., x_n)$ *en la muestra, se denota* $f_n(\underline{x}/\theta)$.

Como los valores son independientes, se tiene:

$$f_n(\underline{x}/\theta) = f_n(x_1, x_2, ..., x_n/\theta) = \prod_{i=1}^{n} f(x_i/\theta)$$

Un estimador del parámetro $\theta$ basado en una muestra de tamaño n es una función $\delta$ de los valores muestrales $(X_1, X_2, ..., X_n)$ a valores en el espacio de parámetro $\Theta$.

El valor que toma el estimador $\delta$ sobre una muestra $(x_1, ..., x_n)$ se llama **estimación** o **valor estimado**.

**El estimador de Máxima Verosimilitud es el estimador que hace** $f_n(\underline{x}/\theta)$ **máxima.**

Tal estimador puede entonces no ser único, o bien no existir.

## 3.4 Ejemplos

Ejemplo 1: Una máquina produce diariamente un lote de piezas. Un criterio basado sobre normas de calidad vigente permite clasificar cada pieza fabricada como defectuosa o no defectuosa. El cliente aceptara el lote si la proporción de piezas $\theta$ defectuosas contenidas en el lote no sobrepasa el valor $\theta_o$. El fabricante tiene que controlar entonces la proporción $\theta$ de piezas defectuosas contenidas en cada lote que fabrica. Pero si la cantidad de piezas N de cada lote es muy grande, no podrá examinar cada una para determinar el valor de $\theta$. El fabricante efectúa entonces el control de calidad de una muestra aleatoria pequeña con n piezas. Se define la v.a. $X$ que toma el valor 1 si la pieza es defectuosa y 0 en el caso contrario. Sean $X_1, X_2, ..., X_n$ los valores abtenidos sobre la muestra.

$X_i \sim Bernoulli(p) \quad (0 \leq \theta \leq 1)$

$$f_n(\underline{x}/\theta) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i}$$

$$\max_\theta f_n(\underline{x}/\theta) \Longleftrightarrow \max_\theta Logf_n(\underline{x}/\theta)$$

$$Logf_n(\underline{x}/\theta) = \sum_{i=1}^{n}[x_i Log\theta + (1-x_i)Log(1-\theta)]$$

$$\frac{dLogf_n(\underline{x}/\theta)}{d\theta} = \frac{\sum x_i}{\theta} - \frac{n - \sum x_i}{1-\theta} = 0$$

Luego el estimador de máxima verosimilitud (E.M.V.) $\hat{\theta}$ de $\theta$ es la proporción de piezas defectuosas observada $\sum x_i/n$.

Ejemplo 2: El ministerio de la salud quiere conocer la talla promedia $\mu$ de las mujeres chilenas adultas. Si $X_1, X_2, ..., X_N$ son las tallas de todas las chilenas adultas, $\mu = \sum X_i/N$. Dado el tamaño grande de esta población, se obtiene la talla de muestra aleatoria de tamaño pequeño n. Sean $X_1, X_2, ..., X_n$.

Se supone que $X_i \sim \mathcal{N}(\mu, \sigma^2)$ con $\sigma^2$, $\mu$ y $\sigma^2$ desconocidos.

$$f_n(\underline{x}/\theta) = (1/2\pi\sigma^2)^{n/2} exp\{-\sum(x_i - \mu)^2/2\sigma^2\}$$

$Logf_n(\underline{x}/\theta)$ es máximo cuando $\mu = \bar{x}_n$ la media muestral y $\sigma^2 = S_n^2$ la varianza muestral.

Notas:
- Si se supone la varianza poblacional $\sigma^2$ conocida, el E.M.V. de $\mu$ queda igual a la media muestral $\bar{x}_n$.

- Se puede buscar el estimador de la varianza o bien de su raíz $\sigma$. El resultado no cambia.

Ejemplo 3: $X_i \sim Uniforme[0, \theta] \qquad \theta > 0$

$$f_n(\underline{x}/\theta) = 1/\theta^n \quad si \;\; 0 \leq x_i \leq \theta \;\; \forall i$$

Cuando $\theta \geq x_i$ para todo i, $f_n(\underline{x}/\theta)$ es no nulo y es decreciente en $\theta$; luego $f_n(\underline{x}/\theta)$ es máxima para el valor más pequeño de $\theta$ que hace $f_n(\underline{x}/\theta)$ no nulo: el E.M.V. de $\theta$ es entonces $\hat{\theta} = X_{(n)} = max\{X_1, ..., X_n\}$

El método de los momentos produce un estimador bien diferente. En efecto, como E(X) = $\theta/2$, el estimador de los momentos es $\tilde{\theta} = 2\bar{x}_n$.

En este ejemplo, una dificultad se presenta cuando se toma el intervalo $]0, \theta[$ abierto, dado que no se puede tomar como estimador el máximo $X_{(n)}$; en este caso no existe E.M.V. Puede ocurrir que no es único también: si se define el intervalo $[\theta, \theta + 1]$, la función de verosimilitud es:

$$f_n(\underline{x}/\theta) = 1 \quad si \quad \theta \leq x_i \leq \theta + 1 \quad \forall i$$

es decir:

$$f_n(\underline{x}/\theta) = 1 \quad si \quad max\{x_1, ..., x_n\} - 1 \leq \theta \leq min\{x_1, ..., x_n\}$$

Por lo cual todo elemento del intervalo $[max\{x_1, ..., x_n\} - 1, min\{x_1, ..., x_n\}]$ es E.M.V.

Aquí el estimador de los momentos, que es igual a $\bar{x}_n - 1/2$, es bien diferente también.

## 3.5 Propiedades

¿Cómo elegir un estimador? ¿Cómo decidir si un estimador es aceptable? Para eso tiene que cumplir ciertas propiedades razonables.

- Invarianza
  Observamos en las notas del ejemplo 2, que el E.M.V. de $\sigma$ se puede obtener directamente o como la raiz del E.M.V. de $\sigma^2$. Eso se debe de la propiedad de **invarianza** del E.M.V. por transformación funcional:

  **Proposición 3.1** *Si $\hat{\theta}$ es el E.M.V. del parámetro $\theta$, si $g : \Theta \longrightarrow \Theta$ es biyectiva, entonces $g(\hat{\theta})$ es el E.M.V. de $g(\theta)$*

Demostración: en efecto si $\tau = g(\theta)$, como g es biyectiva, $\theta = g^{-1}(\tau)$; si $f_n(\underline{x}/\theta) = f_n(\underline{x}/g^{-1}(\tau))$ es máxima para $\hat{\tau}$ tal que $g^{-1}(\hat{\tau}) = \hat{\theta}$. $\hat{\tau}$ es necesariamente el E.M.V. y como g es biyectiva, $\hat{\tau} = g(\hat{\theta})$.

- Consistencia
  Un estimador depende del tamaño de la muestra a traves de los valores muestrales; los estimadores $\hat{\theta}_n$ asociados a muestras de tamaño n ($n \in I\!N$) constituyen sucesiones de v.a. Un buen estimador deberia converger en algún sentido hacia $\theta$.

**Definición 3.2** *Se dice que un estimador $\hat{\theta}_n$ de un parámetro $\theta$ es* **consistente** *cuando converge en probabilidad hacia $\theta$:*

$$I\!P(|\hat{\theta}_n - \theta| < \epsilon) \overset{n\to\infty}{\longrightarrow} 1$$

Los momentos empíricos de una v.a. real son estimadores consistentes de los momentos teóricos correspondientes. Más aún la convergencia es casi-segura y la distribución asintótica de estos estimadores es normal.

- Estimador insesgado

**Definición 3.3** *Se dice que un estimador $\hat{\theta}$ de $\theta$ es* **insegado** *si $E(\hat{\theta}) = \theta$.*

Vimos que la media muestral $\bar{x}_n$ es un estimador insesgado de la media poblacional si la muestra es aleatoria simple, pero la varianza muestral $S_n^2 = 1/n \sum(x_i - x_n)^2$ no es un estimador insesgado para la varianza poblacional $\sigma^2$:

$$E(S_n^2) = \frac{n-1}{n}\sigma^2$$

Pero, la diferencia $|E(S_n^2) - \sigma^2| = \sigma^2/n$, que es el sesgo, tiende a cero.

**Definición 3.4** *Se dice que el estimador es* **asintoticamente insesgado** *cuando $E(S_n^2) \overset{n\to\infty}{\longrightarrow} \sigma^2$.*

Por otro lado se puede construir un estimador insesgado de $\sigma^2$ a partir de $S_n^2$: $\tilde{\sigma}^2 = \sum(x_i - \bar{x}_n)^2/(n-1)$. Pero observamos que $\tilde{\sigma}^2 = (\frac{n}{n-1})^2\sigma^2$, es decir que el estimador insesgado $\tilde{\sigma}^2$ tiene mayor varianza que $S_n^2$.

Por otro lado observamos que si $\hat{\theta}_n^2$ es un estimador sesgado de $\theta$, se tiene:

$$E(\hat{\theta}_n - \theta)^2 = Var(\hat{\theta}_n) + (sesgo)^2$$

En efecto,

$$E(\hat{\theta}_n-\theta)^2 = E[(\hat{\theta}_n - E(\hat{\theta}_n) + E(\hat{\theta}_n) - \theta)^2]$$
$$E(\hat{\theta}_n-\theta)^2 = E[(\hat{\theta}_n - E(\hat{\theta}_n))^2] + [E(\hat{\theta}_n) - \theta)]^2$$

Si $[E(\hat{\theta}_n) - \theta)]^2 \longrightarrow 0$ entonces $\hat{\theta}_n$ converge en media cuadrática hacia 0. ($\hat{\theta}_n \stackrel{m.c.}{\longrightarrow}$ 0)

**Proposición 3.2**

$$E(\hat{\theta}_n-\theta)^2 \longrightarrow 0 \Longleftrightarrow Var(\hat{\theta}_n) \to 0 \quad y \quad E(\hat{\theta}_n) \to \theta$$

Como la convergencia en media cuadrática implica la convergencia en probabilidad se tiene:

**Proposición 3.3** *Si $\hat{\theta}_n$ es un estimador consistente de $\theta$ y $E(\hat{\theta}_n)$ es finito entonces $\hat{\theta}_n$ es asintoticamente insesgado.*

**Proposición 3.4** *Si $Var(\hat{\theta}_n) \to 0 \quad y \quad E(\hat{\theta}_n) \to \theta$, entonces $\hat{\theta}_n$ es un estimador consistente de $\theta$.*

Nota: Es una condición suficiente pero no necesaria.

- Suficiencia
  En el ejemplo 1, se busca deducir de las observaciones de una muestra aleatoria de n piezas una información sobre la proporción $\theta$ de piezas defectuosas en el lote total. Es más simple considerar el número de piezas defectuosas encontradas en la muestra en vez de la sucesión de resultados $X_1, X_2, ..., X_n$. El conocimiento de los valores individuales no procura ninguna información aditiva para la proporción $\theta$ que $\sum_{i=1}^{n} X_i$. Se redujo los n datos a un sólo valor, que es función de estos datos, sin perder información para determinar $\theta$.

  En el ejemplo 2, la media muestral $\bar{x}$ permite simplificar la información dada por los n valores muestrales. Pero nos preguntamos si se pierde información usando la media muestral para estimar la media $\mu$ de la población.

  Observamos que si suponemos la varianza conocida, la función de verosimilitud puede escribirse como función unicamente de la media muestral y del tamaño n de la muestra:

  $$f_n(\underline{x}/\theta) = (1/\sqrt{2\pi})^n exp\{-n(\bar{x}_n - \theta)^2/2\}$$

  Es decir que la única información relevente para estimar $\theta$ es dada por la media muestral. En este caso se dice que la media muestral es un estadístico suficiente. Un estadístico suficiente que se toma como estimador del parámetro $\theta$, debería contener toda la información que llevan los valores muestrales sobre $\theta$.

**Definición 3.5** *Un estadístico $T(x_1, ..., x_n)$, función de los valores muestrales y a valor en $\Theta$ se dice* **suficiente** *para $\theta$ si la distribución conjunta de los valores muestrales condicionalmente a $T(x_1, ..., x_n)$ no depende de $\theta$.*

**Definición 3.6** *Se dice que un estadístico $T$ es suficiente minimal si no se puede encontrar otro estadístico suficiente que hace una mejor reducción de los datos que $T$.*

No es siempre fácil detectar si un estadístico es suficiente. Los dos siguientes teoremas permiten enunciar condiciones para que un estadístico sea suficiente.

**Teorema 3.1** *Teorema de factorización*
*Si $T(\underline{x})$ es suficiente para $\theta$ y $g(T(\underline{x})/\theta)$ es la densidad de $T(\underline{x})$, entonces*

$$f_n(\underline{x}/\theta) = g(T(\underline{x})/\theta)h(\underline{x}/T(\underline{x}))$$

**Teorema 3.2** *Theorema de Darmois-Koopman*
*Si $X$ es una variable real cuyo dominio de variación no depende del parámetro $\theta$, una condición necesaria y suficiente para que existe un estadístico suficiente es que la función de densidad de $X$ sea de la forma:*

$$f(x, \theta) = b(x)c(\theta)exp\{a(x)q(\theta)\}$$

$T_n(X) = \sum_{i=1}^{n} a(X_i)$ *es un estadístico suficiente minimal.*

Si $X \sim \mathcal{N}(\theta, 1)$ y una muestra aleatoria es $x_1, ..., x_n$ de X,

$$f_n(x_1, ..., x_n/\theta) = \frac{1}{2\pi)^n/2} exp(-\frac{1}{2}\sum x_i^2) exp(-\frac{n\theta^2}{2} + n\theta\bar{x})$$

El término $exp(-\frac{1}{2}\sum x_i^2)$ no depende de $\theta$ y el término $exp(-\frac{n\theta^2}{2} + n\theta\bar{x})$ depende de $\theta$ y $\bar{x}$.

$n\bar{x} = \sum x_i$ es un estadístico suficiente; también toda función biyectiva de $\bar{x}$ lo, en particular $\bar{x}$.

## 3.6 Estimadores Bayesianos

### 3.6.1 Distribuciones a priori

En el problema de estimación de un parámetro de una distribución de función de densidad $f(x/\theta)$, es frecuente tener algunas ideas sobre los valores que puede tomar $\theta$; en este caso conviene tomar en cuenta este conocimiento o **creencia** que se puede traducir en una distribución de probabilidad sobre el espacio de parámetros $\Theta$, sea $\pi(\theta)$. Es decir' que ahora $\theta$ ya no es un parámetro constante, sino una variable aleatoria. Esta distribución no depende de los valores muestrales. Está definida previo al muestreo.

Por ejemplo, en un proceso de fabricación se tiene la proporción $\theta$ desconocida de piezas defectuosas. Si no se sabe nada respecto a $\theta$, se puede suponer que todos los valores son equiprobables: $\theta \sim \mathcal{U}(0,1)$. Pero uno puede sopechar que los valores alrededor de 0.10 son más probables; en este caso se podrá tomar una distribución más concentrada en 0.10.

**Definición 3.7** *Se llama* **distribución a priori** *a la distribución atribuida a un parámetro poblacional, antes de tomar alguna muestra.*

### 3.6.2 Distribuciones a posteriori

Ahora hay que relacionar los valores muestrales con la distribución a priori $\pi(\theta)$.

La función de verosimilitud $f_n(\underline{x}/\theta)$ es ahora una densidad condicional y $h(\underline{x},\theta) = f_n(\underline{x}/\theta)\pi(\theta)$ es la densidad conjunta de $(\underline{x},\theta)$. De la cual se puede deducir la distribución condicional de $\theta$ dado los valores muestrales $\underline{x}$:

**Definición 3.8** *La distribución condicional de $\theta$ dada la muestra $(x_1,...,x_n)$ se llama* **distribución a posteriori** *y su densidad es igual a $\xi(\theta/\underline{x}) = \frac{f_n(\underline{x}/\theta)\pi(\theta)}{g_n(\underline{x})}$, en que $g_n(\underline{x}) = \int_\Theta h(\underline{x},\theta)d\theta$ es la densidad marginal de $\underline{x}$.*

La distribución a posteriori representa la actualización de la información a priori $\pi(\theta)$ en vista de la información contenida en los valores muestrales, $f_n(\underline{x}/\theta)$. Podemos entonces estudiar esta distribución a posteriori de $\theta$ dando la moda, la media, la mediana, la varianza, etc. Un estimador natural en este caso es tomar la moda de $\xi(\theta/\underline{x})$, que aparece como el máximo de la verosimilitud corregida.

Ejemplo 4: Sean $X \sim Bernoulli(p)$ y $p \sim \beta eta(\alpha,\beta)$, con $\alpha$ y $\beta$ dados.

$$f_n(\underline{x}/p) = p^{n\bar{x}_n}(1-p)^{n-n\bar{x}_n}$$

$$\pi(p) = p^{\alpha-1}(1-p)^{\beta-1}/B(\alpha,\beta) \quad 0 \le p \le 1$$

en que $B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$.
La densidad a posteriori de p es entonces:

$$\xi(p/\underline{x}) = p^{\alpha+n\bar{x}_n-1}(1-p)^{\beta+n-n\bar{x}_n-1}/B(\alpha+n\bar{x}_n, \beta+n-n\bar{x}_n)$$

que es la distribución $\beta eta(\alpha+n\bar{x}_n, \beta+n-n\bar{x}_n)$. La moda de esta distribución, cuando está definida, es $(\alpha-1)/\alpha+\beta$.

Ejemplo 5: Sean $X \sim \mathcal{N}(\theta, 1)$ y $\theta \sim \mathcal{N}(0, 10)$.
$\xi(\theta/\underline{x}) \propto f_n(\underline{x}/\theta)\pi(\theta)$ ($\propto$ se refiere a la proporcionalidad con respecto a $\theta$).
$\xi(\theta/\underline{x}) \propto exp(-\frac{\sum(x_i-\theta)^2}{2} - \frac{\theta^2}{20})$
$\xi(\theta/\underline{x}) \propto exp(-\frac{-11\theta^2}{20} + n\theta\bar{x}_n)$
$\xi(\theta/\underline{x}) \propto exp(-\frac{-11}{20}(\theta - (10n\bar{x}_n/11))^2)$
La distribución a posteriori de $\theta$ es entonces $\mathcal{N}(\frac{10}{11}n\bar{x}_n, \frac{10}{11})$. La moda de la distribución es la media $\frac{10}{11}n\bar{x}_n$.

### 3.6.3 Funciones de pérdida

Los metodos de estimación propuestos hasta ahora no toman en cuenta un aspecto importante del problema, que son las consecuencias de tales estimaciones.

Dado que los estimadores son la base de una decisión final, es importante poder comparar los procedimientos que conducen a estas decisiones mediente algún criterio de evaluación, que mide las consecuencias de cada estimación en función de los valores del parámetro $\theta$.

**Definición 3.9** *Se llama* **función de pérdida** *o* **función de costo** *a la función L: $\Theta \times \Theta \longrightarrow [0, +\infty[$, en que $L(\theta, \delta)$ es creciente con el error entre el parámetro $\theta$ y su estimador $\delta$.*

No es siempre fácil definir esta función de pérdida, que es específica de cada problema y puede tener algún aspecto subjectivo (noción de utilidad). Sin embargo, se puede elegir entre diversas funciones de pérdida clásicas, cuando no se puede construir una propia:

- Función de pérdida cuadrática
  Es la función de pérdida más utilizada y más criticada:

$$L(\theta, \delta) = (\theta - \delta)^2$$

que penaliza demasiado los errores grandes.

- Función de pérdida absoluta
  Una solución alternativa a la función cuadrádica es usar el valor absoluto:
  $$L(\theta, \delta) = |\theta - \delta|$$
  o bien una función afín por parte:
  $$L(\theta, \delta) = \begin{cases} k_1(\theta - \delta) & \text{si } \theta > \delta \\ k_2(\delta - \theta) & \text{sino} \end{cases}$$

- Función de perdida "0-1"
  Sea $I_\varepsilon(\delta)$ el intervalo de centro $\delta$ y largo $2\varepsilon$.
  $$L(\theta, \delta) = \begin{cases} 0 & \text{si } \theta \in I_\varepsilon(\delta) \\ 1 & \text{sino} \end{cases}$$

### 3.6.4 Estimadores de Bayes

La función de pérdida $L(\theta, \delta)$ es una función de $\theta$ considerado como aleatoria con la distribución a posteriori $\xi(\theta/\underline{x})$. Luego es natural de buscar un estimador $\delta(\underline{x})$ de $\theta$ tal que la pérdida promedia sea mínima.

**Definición 3.10** *El estimador de Bayes es solución de*
$$\min_\delta E(L(\theta, \delta)/\underline{x})$$

- Función de pérdida cuadrática
  Para la función de pérdida cuadrática $L(\theta, \delta) = (\theta - \delta)^2$, el estimador de Bayes es simple de encontrar. $E((\theta - \delta)^2/\underline{x})$ es mńimo para $\delta(\underline{x}) = E(\theta/\underline{x})$.

- Función de pérdida absoluta
  Para la función de pérdida absoluta $L(\theta, \delta) = |\theta - \delta|$, el estimador de Bayes es la mediana de la distribución a posteriori. Mostramos un resultado más general:

  **Proposición 3.5** *El estimador de Bayes asociado a la distribución a posteriori $\xi$ y a la función de perdida*
  $$L(\theta, \delta) = \begin{cases} k_1(\theta - \delta) & \textit{si } \theta > \delta \\ k_2(\delta - \theta) & \textit{sino} \end{cases}$$
  *es la fractila $\frac{k_1}{k_1+k_2}$ de $\xi$.*

Demostración: Se tiene

$$E[L(\theta,\delta)/\underline{x}] = k_2\int_{-\infty}^{\delta}(\delta-\theta)\xi(\theta/\underline{x})d\theta + k_1\int_{\delta}^{+\infty}(\theta-\delta)\xi(\theta/\underline{x})d\theta$$

Derivando con respecto a $\delta$, se obtiene:

$$k_2\mathbb{P}(\theta < \delta/\underline{x}) - k_1\mathbb{P}(\theta > \delta/\underline{x}) = 0$$

Es decir:

$$\mathbb{P}(\theta < \delta/\underline{x}) = \frac{k_1}{k_1+k_2}$$

En particular si $k_1 = k_2$, se obtiene la mediana de la distribución a posteriori de $\theta$.

- Función de pérdida "0-1"
  $E[L(\theta,\delta)]$ es mínimo cuando $\int_{I_\varepsilon(\delta)}\xi(\theta/\underline{x})d\theta$ es máximo. Si $\varepsilon \longrightarrow 0$, entonces $E[L(\theta,\delta)]$ es mínimo cuando $\xi(\theta/\underline{x})$ es máximo. El estimador de Bayes es la moda de $\xi(\theta/\underline{x})$.

**Teorema 3.3** *Theorema de Rao-Blackwell*
*Si T(X) es un estadístico suficiente para $\theta$ y si b(X) es un estimador insesgado de $\theta$, entonces*

$$\delta(T) = E(b(X)/T)$$

*es un estimador insesgado de $\theta$ basado sobre T mejor que b(X).*

Este teorema permite entonces construir estimadores insesgados mejores.

### 3.6.5 Estimadores de Bayes para muestras grandes

Se muestra aquí, a través de un ejemplo, los efectos de la distribución a priori y de la función de pérdida sobre el estimador de Bayes, para muestras grandes. Sea $\theta$ la proporción de defectuosos. Tomamos dos distribuciones a priori y dos funciones de pérdida:
$\pi(\theta) = 1$ para $\theta \in [0,1]$ y $\pi'(\theta) = 2(1-\theta)$ para $\theta \in [0,1]$
$L(\theta,\delta) = (\theta-\delta)^2)$ y $L'(\theta,\delta) = |\theta-\delta|$. Las distribuciones a posteriori son respectivamente;

$$\xi(\theta/\underline{x}) \propto \theta^{n\bar{x}_n}(1-\theta)^{n-n\bar{x}_n}$$

que es una $\beta eta(1+n\bar{x}_n, n+1-n\bar{x}_n)$ y

$$\xi'(\theta/\underline{x}) \propto \theta^{n\bar{x}_n}(1-\theta)^{n+1-n\bar{x}_n}$$

que es una $\beta eta(1 + n\bar{x}_n, n + 2 - n\bar{x}_n)$.
Los estimadores de Bayes para la pérdida cuadrática son las respectivas esperanzas de la distribución $\beta eta$:
$\delta = (1 + n\bar{x}_n)/(n + 2)$ para $\xi$ y $\delta' = (1 + n\bar{x}_n)/(n + 3)$ para $\xi'$.
Los estimadores de Bayes para la pérdida absoluta son las respectivas medianas de la distribución $\beta eta$, que se obtienen resolviendo la ecuación:

$$K \int_0^{\delta} \theta^{\alpha-1}(1-\theta)^{\beta-1} d\theta = 1/2$$

en que $\alpha = 1 + n\bar{x}_n$ y $\beta = n + 1 - n\bar{x}_n$ para $\xi$ y $\beta = n + 2 - n\bar{x}_n$ para $\xi'$.
Si n=100 y $n\bar{x}_n = 10$ entonces $\delta = 11/102 = 0.108$ y $\delta' = 11/103 = 0.107$ para la pérdida cuadrática. Se observara cómo la muestra corrige la distribución a priori, con las medias a priori $E(\theta) = 1/2$ con $\xi$ y $E(\theta) = 1/3$ con $\xi'$.
Encontramos ambos estimadores de Bayes a posteriori muy cercanos con n=100 y cercanos de la media muestral $\bar{x}_n = 10/100 = 0.100$.

En este ejemplo observamos que el estimador de Bayes cuadrático es consistente. No se puede siempre asegurar que el estimador de Bayes es consistente, pero bajo condiciones bastante generales es cierto.