Estimación puntual de parámetros.

Parámetro (θ): Característica de la población.

En estadística la forma funcional de $f(x;\theta)$ es conocida pero se desconoce θ total o parcialmente. La estimación del parámetro $(\hat{\theta})$ debe ser función de los datos de la muestra $x_1, x_2, ..., x_n$, es decir $\hat{\theta} = f(x_1, x_2, ..., x_n)$, pero los datos x_i son cantidades aleatorias por lo tanto $\hat{\theta}$ es aleatorio, entonces debe tener sentido preguntarse por la distribución de $\hat{\theta}$ y esta distribución debe describir por completo las propiedades del estimador.

Las propiedades más deseables de los estimadores son:

- Nos gustaría que la distribución de un estimador esté centrada en el parámetro que se desea estimar. Si la media de la distribución de un estimador $\hat{\theta}$ es igual al parámetro estimado θ, se dice que el estimador está insesgado. Si no es así, se dice que el estimador está sesgado.
- Además nos gustaría que la distribución de un estimador tuviera varianza mínima; es decir, que la dispersión de la distribución fuera lo más pequeña posible, de modo que las estimaciones tiendan a ser cercanas a θ.

Definición:
Un estimador $\hat{\theta}$ de un parámetro θ es insesgado si $E(\hat{\theta}) = \theta$. Si $E(\hat{\theta}) \neq \theta$, se dice que el estimador está sesgado.
El sesgo B de un estimador $\hat{\theta}$ es igual a: $B = E(\hat{\theta}) - \theta$

Un estimador insesgado que tiene la varianza más pequeña de todos los estimadores insesgados se denomina: estimador insesgado de varianza mínima (MVUE).

Hay ocasiones en las que no podemos lograr la falta de sesgo y también la varianza mínima en el mismo estimador. En un caso así, preferimos el estimador que minimiza el error cuadrado medio (ECM):

$$ECM = E\left[\left(\hat{\theta} - \theta\right)^2\right] = V(\hat{\theta}) + B^2$$

Por lo tanto, si $\hat{\theta}$ no está sesgado, es decir, si $B = 0$, entonces $ECM = V(\hat{\theta})$.

Si $\hat{\theta}_1$ y $\hat{\theta}_2$ son estimadores de θ se dice que $\hat{\theta}_1$ es más eficiente que $\hat{\theta}_2$ si y sólo si $ECM(\hat{\theta}_1) \leq ECM(\hat{\theta}_2)$, si son insesgados: $V(\hat{\theta}_1) \leq V(\hat{\theta}_2)$

- Consistencia. Se dice que $\hat{\theta}$ (estimador de θ) es consistente si $\lim_{n\to\infty} P\left[\left|\hat{\theta}_n - \theta\right| < \varepsilon\right] \to 1$

La "distancia" entre el estimador y el parámetro debe ser pequeña; para n muy grande $\hat{\theta} \approx \theta$.

Ejemplo: Sea $x_1, x_2, \ldots, x_n$ una m.a. de n observaciones. Se desconoce la distribución de la población muestreada. Demuestre que la varianza de la muestra, s^2, es un estimador insesgado de la varianza de la población, σ^2.

Por la definición de la varianza de la muestra tenemos que:

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

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

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

- Conocemos que:

  $V(X) = E\left(X^2\right) - \left(E(X)\right)^2 \Rightarrow \sigma^2 = E\left(x^2\right) - \mu^2 \Rightarrow E\left(x^2\right) = \sigma^2 + \mu^2$

  Como cada x_i ($i = 1, 2, \ldots, n$) se escogió al azar de una población con media μ y varianza σ^2, por lo que: $E\left(x_i^2\right) = \sigma^2 + \mu^2$. Además $E\left(\bar{x}^2\right) = \sigma_{\bar{x}}^2 + \left(\mu_{\bar{x}}\right)^2$, pero por el Teorema Central del Límite, sabemos que:

  $$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

  Entonces: $E\left(\bar{x}^2\right) = \sigma_{\bar{x}}^2 + \left(\mu_{\bar{x}}\right)^2 = \dfrac{\sigma^2}{n} + \mu^2$

$$\Rightarrow E\left(s^2\right) = \frac{1}{n-1}\left\{\sum_{i=1}^{n} E\left(x_i^2\right) - nE\left(\bar{x}^2\right)\right\} = \frac{1}{n-1}\left\{\sum_{i=1}^{n}\left(\sigma^2 + \mu^2\right) - n\left(\frac{\sigma^2}{n} + \mu^2\right)\right\}$$

$$= \frac{1}{n-1}\left\{\left(n\sigma^2 + n\mu^2\right) - \sigma^2 - n\mu^2\right\} = \frac{1}{n-1}\left(n\sigma^2 - \sigma^2\right) = \frac{1}{n-1}(n-1)\sigma^2 = \sigma^2$$

$$\Rightarrow E\left(s^2\right) = \sigma^2$$

Por lo tanto s^2 es un estimador insesgado de σ^2.

Ejemplo: Sea $x_1, x_2, ..., x_n$ variables aleatorias, $x_i \sim N\left(\mu, \sigma^2\right)$. Queremos estimar μ.

Sean $\hat{\mu}_1 = \sum_{i=1}^{3} \frac{y_i}{3}, \;\; \hat{\mu}_2 = \frac{y_1}{2} + \frac{y_2}{3} + \frac{y_3}{6}, \;\; \hat{\mu}_3 = y_2$, cuál dará la mejor estimación.

- ¿son insesgados?

$$E\left(\hat{\mu}_1\right) = E\left(\sum_{i=1}^{3} \frac{y_i}{3}\right) = \frac{1}{3}\sum_{i=1}^{3} E\left(y_i\right) = \frac{1}{3}\sum_{i=1}^{3} \mu = \frac{1}{3}(3\mu) = \mu \Rightarrow E\left(\hat{\mu}_1\right) = \mu$$

$$E\left(\hat{\mu}_2\right) = E\left(\frac{y_1}{2} + \frac{y_2}{3} + \frac{y_3}{6}\right) = \frac{E\left(y_1\right)}{2} + \frac{E\left(y_2\right)}{3} + \frac{E\left(y_3\right)}{6} = \frac{\mu}{2} + \frac{\mu}{3} + \frac{\mu}{6} = \frac{(3+2+1)\mu}{6} = \mu \Rightarrow E\left(\hat{\mu}_2\right) = \mu$$

$$E\left(\hat{\mu}_3\right) = E\left(y_2\right) = \mu \Rightarrow E\left(\hat{\mu}_3\right) = \mu$$

Los tres estimadores son insesgados

- ¿Cuál tiene la menor varianza?

$$V\left(\hat{\mu}_1\right) = V\left(\sum_{i=1}^{3} \frac{y_i}{3}\right) = \frac{1}{9}\sum_{i=1}^{3} V\left(y_i\right) = \frac{1}{9}\sum_{i=1}^{3} \sigma^2 = \frac{1}{9}\left(3\sigma^2\right) = \frac{1}{3}\sigma^2$$

$$V\left(\hat{\mu}_2\right) = V\left(\frac{y_1}{2} + \frac{y_2}{3} + \frac{y_3}{6}\right) = \frac{\sigma^2}{4} + \frac{\sigma^2}{9} + \frac{\sigma^2}{36} = \frac{(9+4+1)\sigma^2}{36} = \frac{7}{18}\sigma^2$$

$$V\left(\hat{\mu}_3\right) = V\left(y_2\right) = \sigma^2$$

El de menor varianza es $\hat{\mu}_1 = \sum_{i=1}^{3} \frac{y_i}{3} \Rightarrow \hat{\mu}_1$ es el mejor estimador.

¿Cómo estimar θ?

- No existe una única manera de estimar un parámetro.
- No existe un único estimador de un parámetro $\Rightarrow \exists$ un mejor estimador.

- Las principales técnicas de estimación son: $\begin{cases} \textit{por momentos} \\ \textit{por máxima verosimilitud} \end{cases}$

Máxima verosimilitud.

Si seleccionamos al azar una muestra de n observaciones independientes e idénticamente distribuidas $x_1, x_2, ..., x_n$ de una v.a. x, y si la función de densidad $f(x;\theta)$ es función de un sólo parámetro θ entonces la función de densidad conjunta de los valores $x_1, x_2, ..., x_n$ es:

$$\underbrace{f(x_1, x_2, ..., x_n; \theta) = f(x_1) \cdot f(x_2) \cdots f(x_n) \equiv L}_{verosimilitud}$$

$$\Rightarrow L = \prod_{i=1}^{n} f(x_i; \theta)$$

Fisher sugirió que se debería escoger como estimación de θ el valor que maximiza L, es decir, debemos encontrar el valor de θ que maximice la observación de la muestra conjunta $x_1, x_2, ..., x_n$.

Si la verosimilitud L de la muestra es función de dos parámetros θ_1 y θ_2 entonces las estimaciones de máxima verosimilitud de θ_1 y θ_2 son los valores que maximizan L.

Suponiendo que L es función de un sólo parámetro θ, entonces el valor de θ que maximiza la verosimilitud es el valor para el cual $\frac{dL}{d\theta} = 0$, esta derivada en ocasiones puede ser difícil de obtener, ya que L es el producto de varias cantidades que dependen de θ; por lo tanto se usa el hecho de que la función logaritmo es una función monótonamente creciente, entonces: L será máxima con el mismo valor de θ que maximiza su logaritmo. Por lo que el valor de θ que maximiza la verosimilitud será el valor para el cual $\frac{d \log L}{d\theta} = 0$

Ojo.

L será función sólo de θ, los valores de x_i se fijan y lo que importa es el recorrido del parámetro θ; por lo que no importa que si x_i son v.a. discretas, podemos derivar libremente.

Ejemplo: Sea $x_1, x_2, ..., x_n$ una m.a. de n observaciones de la variable aleatoria x con función de densidad exponencial:

$$f(x;\beta) = \begin{cases} \dfrac{e^{-x/\beta}}{\beta} & si\ 0 \le x < \infty \\ 0 & en\ los\ demás\ puntos \end{cases}$$

Determine el estimador de máxima verosimilitud (MLE) del parámetro β.

$$L=\prod_{i=1}^{n} f(x_i;\theta)=f(x_1)\cdot f(x_2)\cdots f(x_n)=\left(\frac{e^{-x_1/\beta}}{\beta}\right)\cdot\left(\frac{e^{-x_2/\beta}}{\beta}\right)\ \cdot\ \cdot\ \cdot\ \left(\frac{e^{-x_n/\beta}}{\beta}\right)=\frac{e^{-\sum x_i/\beta}}{\beta^n}$$

$$\Rightarrow \log L=\log\left(e^{-\sum x_i/\beta}\right)-n\log\beta=-\frac{\sum x_i}{\beta}-n\log\beta$$

$$\frac{d\log L}{d\beta}=\frac{\sum x_i}{\beta^2}-\frac{n}{\beta}=0\Rightarrow n\hat{\beta}=\sum x_i\Rightarrow \hat{\beta}=\frac{\sum_{i=1}^{n} x_i}{n}=\overline{x}$$

¿es máximo?

$$\left.\frac{d^2\log L}{d\beta^2}\right|_{\beta\to\hat{\beta}}=-2\frac{\sum x_i}{\hat{\beta}^3}+\frac{n}{\hat{\beta}^2}=-2\frac{\sum x_i}{\left(\frac{\sum x_i}{n}\right)^3}+\frac{n}{\left(\frac{\sum x_i}{n}\right)^2}=-2\frac{n^3}{\left(\sum x_i\right)^2}+\frac{n^3}{\left(\sum x_i\right)^2}=-\frac{n^3}{\left(\sum x_i\right)^2}<0\ (\textit{máximo})$$

$\therefore\ \hat{\beta}=\overline{x}$ es el MLE de β

Ejemplo: Sea $x_1,x_2,...,x_n$ una m.a. de n observaciones de la variable aleatoria x con función de densidad normal con media μ y varianza σ^2, obtenga los MLEs de μ y σ^2:

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

$$L=\prod_{i=1}^{n} f(x_i;\theta)=f(x_1)\cdot f(x_2)\cdots f(x_n)=$$

$$=\left(\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(x_1-\mu)^2}{2\sigma^2}\right\}\right)\cdot\left(\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(x_2-\mu)^2}{2\sigma^2}\right\}\right)\ \cdot\ \cdot\ \cdot\ \left(\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(x_n-\mu)^2}{2\sigma^2}\right\}\right)$$

$$=\frac{\exp\left\{-\sum\frac{(x_i-\mu)^2}{2\sigma^2}\right\}}{\sigma^n(2\pi)^{n/2}}=\left(\sigma^{-n/2(2)}(2\pi)^{-n/2}\right)\exp\left\{-\sum\frac{(x_i-\mu)^2}{2\sigma^2}\right\}$$

$$\Rightarrow \log L=-\frac{n}{2}\log\left(\sigma^2\right)-\frac{n}{2}\log 2\pi-\frac{\sum(x_i-\mu)^2}{2\sigma^2}$$

- $$\frac{d\log L}{d\mu}=\frac{\sum 2(x_i-\mu)}{2\sigma^2}=0\Rightarrow\sum(x_i-\hat\mu)=0\Rightarrow\sum x_i-n\hat\mu=0\Rightarrow n\hat\mu=\sum x_i\Rightarrow\hat\mu=\frac{\sum x_i}{n}=\bar{x}$$

- $$\frac{d\log L}{d\sigma^2}=\frac{\sum(x_i-\mu)^2}{2(\sigma^2)^2}-\frac{n}{2}\left(\frac{1}{\sigma^2}\right)=0\Rightarrow\frac{\sum(x_i-\hat\mu)^2}{2(\hat\sigma^2)^2}=\frac{n}{2}\left(\frac{1}{\hat\sigma^2}\right)\Rightarrow\frac{\sum(x_i-\bar{x})^2}{\hat\sigma^2}=n$$

$$\Rightarrow\hat\sigma^2=\frac{\sum(x_i-\bar{x})^2}{n}$$

sesgado

$\therefore\ \hat\mu=\bar{x}\ ,\ \hat\sigma^2=\dfrac{\sum(x_i-\bar{x})^2}{n}$ son los MLEs de μ y σ^2

Ejemplo: Sea $x_1,x_2,\ldots,x_n$ una m.a. de n observaciones de la variable aleatoria x con función de probabilidad Bernoulli:

$$f(x;p)=p^x(1-p)^{1-x}\qquad x=0,1$$

Determine el estimador de máxima verosimilitud (MLE) del parámetro p.

$$L=\prod_{i=1}^{n}f(x_i;\theta)=\prod_{i=1}^{n}p^{x_i}(1-p)^{1-x_i}=p^{\sum x_i}(1-p)^{\sum 1-x_i}=p^{\sum x_i}(1-p)^{\left(n-\sum x_i\right)}$$
$$\Rightarrow\log L=\sum x_i\log p+\left(n-\sum x_i\right)\log(1-p)$$

- $$\frac{d\log L}{dp}=\frac{\sum x_i}{p}-\frac{\left(n-\sum x_i\right)}{1-p}=0\Rightarrow\frac{\sum x_i}{\hat p}=\frac{\left(n-\sum x_i\right)}{1-\hat p}\Rightarrow\frac{1-\hat p}{\hat p}=\frac{\left(n-\sum x_i\right)}{\sum x_i}$$

$$\Rightarrow\frac{1}{\hat p}-1=\frac{n}{\sum x_i}-1\Rightarrow\hat p=\frac{\sum x_i}{n}\ \textit{proporción}$$

$\therefore$ la proporción $\hat p=\dfrac{\sum x_i}{n}$ es el MLE de p.

Problema: ¿cómo garantizar que dentro de los estimadores $\{\hat\theta_i\}$ de una característica de la población se tiene "el mejor" en algún sentido?

Cota inferior de Cramér- Rao.

Sea $x_1,x_2,\ldots,x_n$ m.a. con $f(\bullet,\theta)$. Sea $\hat\tau(\theta)=T=t(x_1,x_2,\ldots,x_n)$ un estimador insesgado de $\tau(\theta)$. Entonces si se cumplen los siguientes supuestos, llamados condiciones de regularidad

i) $\frac{\partial}{\partial\theta}\log f(x;\theta)\exists\forall x$ y $\forall\theta$

ii) $\frac{\partial}{\partial\theta}\int\int\cdots\int\prod_{i=1}^{n} f(x_i;\theta)dx_1dx_2\cdots dx_n=\int\int\cdots\int\frac{\partial}{\partial\theta}\prod_{i=1}^{n} f(x_i;\theta)dx_1dx_2\cdots dx_n$

iii)
$$\frac{\partial}{\partial\theta}\int\int\cdots\int t(x_1,x_2,...,x_n)\prod_{i=1}^{n} f(x_i;\theta)dx_1dx_2\cdots dx_n=\int\int\cdots\int t(x_1,x_2,...,x_n)\frac{\partial}{\partial\theta}\prod_{i=1}^{n} f(x_i;\theta)dx_1dx_2\cdots dx_n$$

iv) $0<E\left\{\left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right)^2\right\}<\infty\forall\theta$

tenemos que: $$V\left[T=t(x_1,x_2,...,x_n)\right]\geq\frac{\left[\tau'(\theta)\right]^2}{nE\left[\left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right)^2\right]}$$

cota inferior para la varianza de un estimador insesgado de una característica.

- Proporciona el valor mínimo que puede tomar la varianza de un estimador insesgado.
- Si tenemos varios estimadores insesgados, entonces el que tenga la varianza más cercana a esta cota será el mejor de ellos.
- Si tenemos un estimador cuya varianza alcanza la cota: este es el mejor estimador de todos.

Demostración:

$$L=\prod_{i=1}^{n} f(x_i;\theta)\Rightarrow\log L=\sum_{i=1}^{n}\log f(x_i;\theta)$$

$$\Rightarrow\frac{\partial}{\partial\theta}\log L=\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f(x_i;\theta)\equiv U\quad \textit{función score (de puntajes)}$$

- $E\left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right)=\int_{-\infty}^{\infty}\frac{\partial}{\partial\theta}\log f(x;\theta)\cdot f(x;\theta)dx=\int_{-\infty}^{\infty}\frac{\frac{\partial}{\partial\theta}f(x;\theta)}{f(x;\theta)}f(x;\theta)dx=\int_{-\infty}^{\infty}\frac{\partial}{\partial\theta}f(x;\theta)dx=$

$$\underset{condición(ii)}{=}\frac{\partial}{\partial\theta}\int_{-\infty}^{\infty}f(x;\theta)dx=\frac{\partial}{\partial\theta}(1)=0$$

$$\therefore E\left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right)=0\Rightarrow E(U)=E\left(\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f(x_i;\theta)\right)=\sum_{i=1}^{n}E\left(\frac{\partial}{\partial\theta}\log f(x_i;\theta)\right)=0$$

$$\Rightarrow E(U)=0$$

- $$E\left(U^2\right)=E\left\{\left(\sum_{i=1}^{n}\left(\frac{\partial}{\partial\theta}\log f\left(x_i;\theta\right)\right)^2\right)\right\}$$ como las x_i son v.a.i.i.d.

  $$\Rightarrow E\left(U^2\right)=E\left\{\left(\sum_{i=1}^{n}\left(\frac{\partial}{\partial\theta}\log f\left(x_i;\theta\right)\right)^2\right)\right\}=nE\left\{\left(\frac{\partial}{\partial\theta}\log f\left(x_i;\theta\right)\right)^2\right\}$$

Pero $V(X)=E\left(X^2\right)-\left(E(X)\right)^2 \Rightarrow V(U)=E\left(U^2\right)-\left(\cancel{E(U)}\right)^2=E\left(U^2\right)$

$$\therefore \underbrace{V(U)=E\left(U^2\right)=nE\left\{\left(\frac{\partial}{\partial\theta}\log f\left(x_i;\theta\right)\right)^2\right\}}_{\text{información de } Fisher}\ldots(*)$$

Ahora por insesgamiento de $\hat{\tau}(\theta)=T=t\left(x_1,x_2,\ldots,x_n\right)$:

$$\tau(\theta)=E\left(\hat{\tau}(\theta)\right)=\int\int\cdots\int\hat{\tau}(\theta)\cdot f\left(x_1;\theta\right)\cdot f\left(x_2;\theta\right)\cdots f\left(x_n;\theta\right)dx_1dx_2\cdots dx_n=$$

$$=\int\int\cdots\int\hat{\tau}(\theta)\prod_{i=1}^{n}f\left(x_i;\theta\right)dx_1dx_2\cdots dx_n=\int\int\cdots\int\hat{\tau}(\theta)L\left(\theta,\underline{x}\right)dx_1dx_2\cdots dx_n$$

$$\Rightarrow \tau'(\theta)=\frac{\partial}{\partial\theta}\int\int\cdots\int\hat{\tau}(\theta)L\left(\theta,\underline{x}\right)dx_1dx_2\cdots dx_n \underset{condición(iii)}{=}\int\int\cdots\int\hat{\tau}(\theta)\frac{\partial}{\partial\theta}L\left(\theta,\underline{x}\right)dx_1dx_2\cdots dx_n$$

$$=\int\int\cdots\int\hat{\tau}(\theta)\underbrace{\frac{\partial}{\partial\theta}\left(\log L\left(\theta,\underline{x}\right)\right)}_{U}\cdot L\left(\theta,\underline{x}\right)dx_1dx_2\cdots dx_n$$

$$=\int\int\cdots\int\hat{\tau}(\theta)\cdot U\prod_{i=1}^{n}f\left(x_i;\theta\right)dx_1dx_2\cdots dx_n=E\left(\hat{\tau}(\theta)\cdot U\right)$$

$$\Rightarrow \left(\tau'(\theta)\right)^2=E^2\left(\hat{\tau}(\theta)\cdot U\right)$$

Ahora bien $Cov(X,Y)=E(X\cdot Y)-E(X)E(Y)$

Si $X=\hat{\tau}(\theta), Y=U$, tenemos:

$$Cov\left(\hat{\tau}(\theta),U\right)=E\left(\hat{\tau}(\theta)\cdot U\right)-E\left(\hat{\tau}(\theta)\right)\cancel{E(U)}\Rightarrow Cov^2\left(\hat{\tau}(\theta),U\right)=E^2\left(\hat{\tau}(\theta)\cdot U\right)=\left(\tau'(\theta)\right)^2\ldots(**)$$

Por otra parte, sabemos que para v.a.: $Cov^2(X,Y)\le V(X)\cdot V(Y)$

$$\Rightarrow Cov^2\left(\hat{\tau}(\theta),U\right)\le V\left(\hat{\tau}(\theta)\right)\cdot V(U)\Rightarrow V\left(\hat{\tau}(\theta)\right)\ge\frac{Cov^2\left(\hat{\tau}(\theta),U\right)}{V(U)}$$

entonces, sustituyendo (*) y (**), obtenemos la **Cota inferior de Cramér- Rao (CICR)**:

$$V\left(\hat{\tau}(\theta)\right) \geq \frac{Cov^2\left(\hat{\tau}(\theta), U\right)}{V(U)} \Rightarrow V\left[T = t\left(x_1, x_2, \ldots, x_n\right)\right] \geq \frac{\left[\tau'(\theta)\right]^2}{nE\left[\left(\frac{\partial}{\partial\theta}\log f\left(x;\theta\right)\right)^2\right]} \quad Q.E.D$$

Resultado útil: $E\left[\left(\frac{\partial}{\partial\theta}\log f\left(x;\theta\right)\right)^2\right] = -E\left\{\frac{\partial^2}{\partial\theta^2}\log f\left(x;\theta\right)\right\}$

Ejemplo: ¿Existe un estimador $\hat{p}$ de p (distribución Bernoulli), que alcance la cota inferior de Cramér-Rao?

$$\tau(p) = p$$

$$\tau'(p) = 1$$

$$f\left(x;p\right) = p^x\left(1-p\right)^{1-x}$$

$$\log f\left(x;p\right) = x\log p + (1-x)\log(1-p)$$

$$\frac{\partial}{\partial p}\log f\left(x;p\right) = \frac{x}{p} - \frac{1-x}{1-p}$$

$$\frac{\partial^2}{\partial p^2}\log f\left(x;p\right) = -\frac{x}{p^2} - \frac{1-x}{(1-p)^2}$$

$$-E\left\{\frac{\partial^2}{\partial p^2}\log f\left(x;p\right)\right\} = E\left\{\frac{x}{p^2} + \frac{1-x}{(1-p)^2}\right\} = \frac{E(x)}{p^2} + \frac{E(1-x)}{(1-p)^2} = \frac{p}{p^2} + \frac{1-p}{(1-p)^2} = \frac{1}{p} + \frac{1}{(1-p)} = \frac{1}{p(1-p)}$$

$$-nE\left\{\frac{\partial^2}{\partial p^2}\log f\left(x;p\right)\right\} = \frac{n}{p(1-p)}$$

$$\Rightarrow V[\hat{p}] \geq \frac{\left[\tau'(p)\right]^2}{nE\left[\left(\frac{\partial}{\partial\theta}\log f\left(x;p\right)\right)^2\right]} = \frac{\left[\tau'(p)\right]^2}{-nE\left\{\frac{\partial^2}{\partial\theta^2}\log f\left(x;\theta\right)\right\}} = \frac{1}{\frac{n}{p(1-p)}} = \frac{p(1-p)}{n}$$

$$\Rightarrow V[\hat{p}] \geq \frac{p(1-p)}{n} \rightarrow (CICR)$$

Ya sabemos que la proporción $\hat{p} = \frac{\sum x_i}{n}$ es el MLE de p, ¿alcanza la CICR?

Primero veamos si es insesgado:

$$E(\hat{p}) = E\left(\frac{\sum x_i}{n}\right) = \frac{1}{n}\sum_{i=1}^{n} E(x_i) = \frac{1}{n}\sum_{i=1}^{n} p = \frac{np}{n} = p \Rightarrow \hat{p} \text{ es insesgado}$$

Ahora

$$V(\hat{p}) = V\left(\frac{\sum x_i}{n}\right) = \frac{1}{n^2}\sum_{i=1}^{n} V(x_i) = \frac{1}{n^2}\sum_{i=1}^{n} p(1-p) = \frac{1}{n^2} np(1-p) = \frac{p(1-p)}{n} \to es\ la \text{ cota}$$

$\Rightarrow \hat{p}\ alcanza\ la\ CICR \Rightarrow \hat{p}$ es el estimador de menor varianza y es el mejor estimador de todos.

Ejemplo: Normal(μ, σ^2), con μ desconocida y σ^2 conocida.

$$\tau(\mu) = \mu \Rightarrow \tau'(\mu) = 1$$

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

$$\log f(x;\mu,\sigma^2) = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{(x-\mu)^2}{2\sigma^2}$$

$$\frac{\partial}{\partial\mu}\log f(x;\mu,\sigma^2) = \frac{(x-\mu)}{\sigma^2}$$

$$\frac{\partial^2}{\partial\mu^2}\log f(x;\mu,\sigma^2) = -\frac{1}{\sigma^2}$$

$$-E\left\{\frac{\partial^2}{\partial\mu^2}\log f(x;\mu,\sigma^2)\right\} = -E\left\{-\frac{1}{\sigma^2}\right\} = \frac{1}{\sigma^2}$$

$$-nE\left\{\frac{\partial^2}{\partial\mu^2}\log f(x;\mu,\sigma^2)\right\} = \frac{n}{\sigma^2}$$

$$\Rightarrow V[\hat{\mu}] \geq \frac{[\tau'(\mu)]^2}{nE\left[\left(\frac{\partial}{\partial\theta}\log f(x;\mu,\sigma^2)\right)^2\right]} = \frac{[\tau'(\mu)]^2}{-nE\left\{\frac{\partial^2}{\partial\theta^2}\log f(x;\mu,\sigma^2)\right\}} = \frac{1}{\frac{n}{\sigma^2}} = \frac{\sigma^2}{n}$$

$$\Rightarrow V[\hat{\mu}] \geq \frac{\sigma^2}{n} \to (CICR)$$

¿conocemos un estimador insesgado de μ cuya varianza sea $\frac{\sigma^2}{n}$? Si, ya que por el Teorema Central del Límite, sabemos que:

$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

es decir $V\left(\bar{x}\right) = \frac{\sigma^2}{n} \Rightarrow \hat{\mu} = \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ *alcanza la CICR* $\Rightarrow \bar{x}$ es MVUE.

¿Qué pasa para σ^2 desconocida?

$$\tau\left(\sigma^2\right) = \sigma^2 \Rightarrow \tau'\left(\sigma^2\right) = 1$$

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

Sea $\sigma^2 = \theta$

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

$$\log f\left(x;\mu,\theta\right) = -\frac{1}{2}\log\left(2\pi\theta\right) - \frac{\left(x-\mu\right)^2}{2\theta}$$

$$\frac{\partial}{\partial\theta}\log f\left(x;\mu,\theta\right) = -\frac{2\pi}{2\left(2\pi\theta\right)} + \frac{\left(x-\mu\right)^2}{2\theta^2} = -\frac{1}{2\theta} + \frac{\left(x-\mu\right)^2}{2\theta^2}$$

$$\frac{\partial^2}{\partial\theta^2}\log f\left(x;\mu,\theta\right) = \frac{1}{2\theta^2} + \frac{\left(x-\mu\right)^2}{\theta^3}$$

$$-E\left\{\frac{\partial^2}{\partial\theta^2}\log f\left(x;\mu,\theta\right)\right\} = -E\left\{\frac{1}{2\theta^2} + \frac{\left(x-\mu\right)^2}{\theta^3}\right\} = -\frac{1}{2\theta^2} + \frac{1}{\theta^3}E\left\{\left(x-\mu\right)^2\right\}$$

$$E\left\{\left(x-\mu\right)^2\right\} = E\left\{x^2 - 2x\mu + \mu^2\right\} = E\left(x^2\right) - 2\mu E\left(x\right) + \mu^2 = E\left(x^2\right) - 2\mu^2 + \mu^2 = E\left(x^2\right) - \mu^2$$

Pero $V\left(X\right) = E\left(X^2\right) - \left(E\left(X\right)\right)^2 \Rightarrow E\left(x^2\right) = V\left(x\right) + \left(E\left(x\right)\right)^2 = \sigma^2 + \mu^2$

$$\Rightarrow E\left\{\left(x-\mu\right)^2\right\} = E\left(x^2\right) - \mu^2 = \left(\sigma^2 + \mu^2\right) - \mu^2 = \sigma^2 = \theta$$

$$\therefore -E\left\{\frac{\partial^2}{\partial\theta^2}\log f\left(x;\mu,\theta\right)\right\} = -\frac{1}{2\theta^2}+\frac{1}{\theta^3}\theta = -\frac{1}{2\theta^2}+\frac{1}{\theta^2} = \frac{1}{2\theta^2}$$

$$\Rightarrow -nE\left\{\frac{\partial^2}{\partial\theta^2}\log f\left(x;\mu,\theta\right)\right\} = \frac{n}{2\theta^2}$$

$$\Rightarrow V\left[\hat{\theta}\right] \geq \frac{\left[\tau'(\theta)\right]^2}{nE\left[\left(\frac{\partial}{\partial\theta}\log f\left(x;\mu,\theta\right)\right)^2\right]} = \frac{\left[\tau'(\theta)\right]^2}{-nE\left\{\frac{\partial^2}{\partial\theta^2}\log f\left(x;\mu,\theta\right)\right\}} = \frac{1}{\frac{n}{2\theta^2}} = \frac{2\theta^2}{n} = \frac{2\sigma^4}{n}$$

$$\Rightarrow V\left[\hat{\sigma}^2\right] \geq \frac{2\sigma^4}{n} \rightarrow (CICR)$$

- Si μ es conocida $\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}\left(x_i-\mu\right)^2}{n} \Rightarrow V\left(\hat{\sigma}^2\right) = \frac{2\sigma^4}{n} \Rightarrow$ *alcanza la CICR*
- Si μ es desconocida $\hat{\sigma}^2 = s^2 = \frac{\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2}{n-1} \Rightarrow V\left(s^2\right) = \frac{2\sigma^4}{n-1} \Rightarrow$ Sólo si $n\to\infty$ se alcanza la CICR

Ejemplo:

Uniforme $(0,\theta)$ $\Rightarrow$ $f\left(x;\theta\right) = \frac{1}{\theta}$ $\quad x \in (0,\theta)$

Bajo las condiciones de regularidad ya demostramos que se debe cumplir

$$E\left(\frac{\partial}{\partial\theta}\log f\left(x;\theta\right)\right) = 0$$

Pero para la función Uniforme$(0,\theta)$ tenemos que:

$$\log f\left(x;\theta\right) = -\log\theta \Rightarrow \frac{\partial}{\partial\theta}\log f\left(x;\theta\right) = -\frac{1}{\theta}$$

$$\Rightarrow E\left(\frac{\partial}{\partial\theta}\log f\left(x;\theta\right)\right) = \int_0^\theta -\frac{1}{\theta}f\left(x;\theta\right)dx = -\int_0^\theta \frac{1}{\theta^2}dx = -\frac{1}{\theta^2}x\Big|_0^\theta = -\frac{1}{\theta^2}\theta = -\frac{1}{\theta} \neq 0$$

$\therefore$ La función Uniforme$(0,\theta)$ no cumple las condiciones de regularidad $\Rightarrow \nexists CICR$ (No existe la cota inferior de Cramér-Rao).

Estimemos el parámetro θ, para la Uniforme$(0,\theta)$, usando la técnica de momentos y la de máxima verosimilitud:

Momentos	**Máxima verosimilitud**
$E\left(x^m\right)=\frac{1}{n}\sum x_i^m$ $m=1$ $E\left(x\right)=\frac{1}{n}\sum x_i \Rightarrow \frac{\theta}{2}=\overline{x}$ $\Rightarrow \hat{\theta}_M = 2\overline{x}$	$L=\prod_{i=1}^{n}\frac{1}{\theta}=\frac{1}{\theta^n}\Rightarrow \log L=-n\log\theta$ $\frac{\partial}{\partial\theta}\log L=-\frac{n}{\theta}=0!!!$ no podemos maximizar $\log L$, pero si L, obsérvese que $\frac{1}{\theta^n}$ es una función decreciente de θ. Pero debe ocurrir que $\theta\geq x_i, \forall x_i$ en particular $\theta\geq x_{(n)}$ (máximo de la muestra). $\Rightarrow \frac{1}{\theta}\leq\frac{1}{x_{(n)}}\Rightarrow\frac{1}{\theta^n}\leq\frac{1}{x_{(n)}^n}\Rightarrow\hat{\theta}_{MV}=x_{(n)}$ (MLE)

cuál de los dos estimadores es "el mejor"

- ¿son insesgados?

$$E\left(\hat{\theta}_M\right)=E\left(2\overline{x}\right)=2E\left(\frac{\sum x_i}{n}\right)=2\frac{\sum E\left(x_i\right)}{n}=2\frac{\sum\frac{\theta}{2}}{n}=2\frac{n\frac{\theta}{2}}{n}=\theta\rightarrow\hat{\theta}_M \text{ es insesgado.}$$

Ahora para determinar la esperanza del estimador máximo verosímil, primero necesito conocer la función de densidad de $x_{(n)}$:

$$f_{(n)}\left(x\right)=n\left(\frac{x}{\theta}\right)^{n-1}\left(\frac{1}{\theta}\right), 0\leq x\leq\theta$$

Entonces,

$$E\left(\hat{\theta}_{MV}\right)=E\left(x_{(n)}\right)=\int_0^\theta xf_{(n)}\left(x\right)dx=\int_{-\infty}^{\infty}xn\left(\frac{x}{\theta}\right)^{n-1}\left(\frac{1}{\theta}\right)dx=\int_0^\theta n\left(\frac{x}{\theta}\right)^{n-1}\left(\frac{x}{\theta}\right)dx=\int_0^\theta n\left(\frac{x}{\theta}\right)^{n}dx$$

$$=\left(\frac{n}{\theta^n}\right)\int_0^\theta x^n dx=\left(\frac{n}{\theta^n}\right)\frac{x^{n+1}}{n+1}\Bigg|_0^\theta=\left(\frac{n}{\theta^n}\right)\frac{\theta^{n+1}}{n+1}=\frac{n\theta}{n+1}\rightarrow\hat{\theta}_{MV}\text{ es sesgado}$$

Aunque el estimador $\hat{\theta}_{MV}$ es sesgado observamos que fácilmente podemos obtener el estimador insesgado: $\hat{\theta}_{MV} = \frac{n+1}{n} x_{(n)} \rightarrow$ es insesgado

Por lo que ya tenemos dos estimadores insesgados $\hat{\theta}_M = 2\overline{x}$ y $\hat{\theta}_{MV} = \frac{n+1}{n} x_{(n)}$

- ¿Cuál tiene la menor varianza?

$$V\left(\hat{\theta}_M\right) = V\left(2\overline{x}\right) = 4V\left(\frac{\sum x_i}{n}\right) = \frac{4}{n^2}\sum V\left(x_i\right) = \frac{4}{n^2}\sum \frac{\theta^2}{12} = \frac{4n\theta^2}{12n^2} = \frac{\theta^2}{3n}$$

$$V\left(\hat{\theta}_{MV}\right) = V\left(\frac{n+1}{n} x_{(n)}\right) = \left(\frac{n+1}{n}\right)^2 V\left(x_{(n)}\right) = \left(\frac{n+1}{n}\right)^2 \left[E\left(x_{(n)}{}^2\right) - \left(E\left(x_{(n)}\right)\right)^2\right]$$

Pero ya determinamos $E\left(x_{(n)}\right) = \frac{n\theta}{n+1}$

Y además:

$$E\left(x_{(n)}{}^2\right) = \int_0^\theta x^2 f_{(n)}\left(x\right) dx = \int_{-\infty}^{\infty} x^2 n\left(\frac{x}{\theta}\right)^{n-1}\left(\frac{1}{\theta}\right) dx = \int_0^\theta nx\left(\frac{x}{\theta}\right)^n dx$$

$$= \left(\frac{n}{\theta^n}\right)\int_0^\theta x^{n+1} dx = \left(\frac{n}{\theta^n}\right)\frac{x^{n+2}}{n+2}\Bigg|_0^\theta = \left(\frac{n}{\theta^n}\right)\frac{\theta^{n+2}}{n+2} = \frac{n\theta^2}{n+2}$$

$$\therefore V\left(\hat{\theta}_{MV}\right) = \left(\frac{n+1}{n}\right)^2 \left[E\left(x_{(n)}{}^2\right) - \left(E\left(x_{(n)}\right)\right)^2\right] = \left(\frac{n+1}{n}\right)^2 \left[\frac{n\theta^2}{n+2} - \left(\frac{n\theta}{n+1}\right)^2\right] =$$

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

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

Entonces tenemos que:

$V\left(\hat{\theta}_M\right) = \frac{\theta^2}{3n}$ y $V\left(\hat{\theta}_{MV}\right) = \frac{\theta^2}{n(n+2)}$ por lo que el de menor varianza es el estimador máximo verosímil $\hat{\theta}_{MV}$

$\Rightarrow \hat{\theta}_{MV} = \frac{n+1}{n} x_{(n)}$ es mejor estimador para la función Uniforme$(0,\theta)$, aunque no existe la cota inferior de Cramér-Rao.

Estadísticas suficientes.

Habíamos definido una estadística $T = t(x_1, x_2, ..., x_n)$, como una medida descriptiva numérica calculada a partir de datos de la muestra. Una estadística condensa o reduce las *n* v.a. en una v.a., por lo que debemos preguntarnos si ¿lo mismo que "decían" *n* v.a. sobre θ, lo dice ahora $T = t(x_1, x_2, ..., x_n)$? ¿habremos perdido información en esta "reducción"?

Estadística suficiente. Sean $x_1, x_2, ..., x_n$ una m.a. de una densidad $f(\bullet;\theta)$. Una estadística $T = t(x_1, x_2, ..., x_n)$ es suficiente para θ si y sólo si, la distribución condicional de $x_1, x_2, ..., x_n$ dado $T = t_i$ no depende de θ para ningún valor t_i de T.

- Basta conocer los valores de la estadística suficiente *T* para conocer θ.
- "Lo demás que no sea *T*" no aporta información sobre θ.
- *T* (una estadística suficiente) "condensa o reduce" el rango de valores de una manera que no haya perdida de información sobre θ.

Ejemplo:

Sean x_1, x_2, x_3 tres ensayos Bernoulli $f(x;p) = p^x(1-p)^{1-x} \quad x = 0,1$ y $T = t(x_1, x_2, x_3) = (x_1 \cdot x_2) + x_3$ una estadística.

Sea ω el rango de valores de la tripleta (x_1, x_2, x_3): $\omega = \left\{ \begin{matrix} (0,0,0),(0,0,1),(0,1,0),(1,0,0) \\ (1,1,1),(0,1,1),(1,0,1),(1,1,0) \end{matrix} \right\}$

$$T = t(x_1, x_2, x_3) = (x_1 \cdot x_2) + x_3 = \{0,1,2\} = \begin{cases} 0 & (x_1, x_2, x_3) = (0,0,0),(1,0,0),(0,1,0) \\ 1 & (1,1,0),(0,0,1),(0,1,1),(1,0,1) \\ 2 & (1,1,1) \end{cases}$$

La distribución condicional de x_1, x_2, x_3 dado $T = t$ será:

$$P(X_1 = x_1, X_2 = x_2, X_3 = x_3 | T = t) = \frac{P(X_1 = x_1, X_2 = x_2, X_3 = x_3)}{P(T = t)}$$

Para $T = 0$: tomando la tripleta $(0,0,0)$

$$P\left(X_1=0, X_2=0, X_3=0 \middle| T=0\right)=\frac{P\left(X_1=0, X_2=0, X_3=0\right)}{P\left(T=0\right)}=\frac{P\left(X_1=0\right)\cdot P\left(X_2=0\right)\cdot P\left(X_3=0\right)}{P\left(T=0\right)}=$$

$$=\frac{\left(1-p\right)^3}{\left(1-p\right)^3+p\left(1-p\right)^2+p\left(1-p\right)^2}=\frac{\left(1-p\right)^3}{\left(1-p\right)^3+2p\left(1-p\right)^2}=\frac{1}{1+\dfrac{2p}{\left(1-p\right)}}$$

depende de p, por lo tanto $T=t\left(x_1,x_2,x_3\right)=\left(x_1\cdot x_2\right)+x_3$ no es una estadística suficiente para p, lo que indica que se ha perdido información.

Ejemplo: Para $x_1,x_2,...,x_n$ una m.a. de n observaciones de la variable aleatoria x con función de probabilidad Bernoulli:

$$f\left(x;p\right)=p^x\left(1-p\right)^{1-x} \qquad x=0,1$$

Sabemos que el estimador de máxima verosimilitud del parámetro p, es $\hat{p}=\frac{\sum x_i}{n}$, ¿será $T=t=\sum x_i$ una estadística suficiente para p?

Tenemos que: $x_i \sim Bernoulli(p) \Rightarrow t=\sum x_i \sim Binomial(n,p)$, por lo tanto, la distribución condicional de x_1,x_2,x_3 dado $T=t=\sum x_i$, será

$$P\left(X_1=x_1, X_2=x_2,...,X_n=x_n \middle| T=\sum x_i=t\right)=\frac{P\left(X_1=x_1,X_2=x_2,...,X_n=x_n\right)}{P\left(T=\sum x_i=t\right)}=\frac{\prod_{i=1}^{n} p^{x_i}\left(1-p\right)^{1-x_i}}{\binom{n}{t}p^t\left(1-p\right)^{n-t}}=$$

$$=\frac{p^{\sum x_i}\left(1-p\right)^{n-\sum x_i}}{\binom{n}{t}p^t\left(1-p\right)^{n-t}}=\frac{p^t\left(1-p\right)^{n-t}}{\binom{n}{t}p^t\left(1-p\right)^{n-t}}=\frac{1}{\binom{n}{t}} \rightarrow \textit{no depende de } p \Rightarrow T=\sum x_i \textit{ es suficiente para } p$$

$\therefore\ \hat{p}=\frac{\sum x_i}{n}$ es una función 1 a 1 de la estadística suficiente, no hay perdida de información sobre p.

Estadísticas conjuntamente suficientes. Sea $x_1,x_2,...,x_n$ una m.a. de densidad $f\left(\bullet;\theta\right)$. Las estadísticas $S_1,S_2,...,S_k$ son conjuntamente suficientes, si la distribución condicional de $x_1,x_2,...,x_n$ dado $S_1=s_1,S_2=s_2,...,S_k=s_k$ no dependen de θ.

¿Habrá una forma más económica de decidir si una estadística es suficiente para un parámetro θ? $\rightarrow$Criterio de factorización.

Criterio de factorización. Sea $x_1, x_2, ..., x_n$ una m.a. de densidad $f(\bullet;\theta)$. Una estadística $S = s(x_1, x_2, ..., x_n)$ es suficiente si y sólo si la densidad conjunta de $x_1, x_2, ..., x_n$ se puede factorizar como:

$$f_{\underline{x}}(x_1, x_2, ..., x_n) = g(s(x_1, x_2, ..., x_n);\theta)h(x_1, x_2, ..., x_n) = g(S;\theta)h(x_1, x_2, ..., x_n)$$

Equivalentemente, para estadísticas conjuntamente suficientes:

$$\begin{aligned} f_{\underline{x}}(x_1, x_2, ..., x_n) &= g(s_1(x_1, x_2, ..., x_n), s_2(x_1, x_2, ..., x_n), ..., s_k(x_1, x_2, ..., x_n);\theta)h(x_1, x_2, ..., x_n) \\ &= g(S_1, S_2, ..., S_k;\theta)h(x_1, x_2, ..., x_n) \end{aligned}$$

h es una función no negativa que no depende de θ

g es no negativa y depende de $x_1, x_2, ..., x_n$ sólo a través de $s_1(x_1, x_2, ..., x_n), ..., s_k(x_1, x_2, ..., x_n)$

Ejemplo: Sea $x_1, x_2, ..., x_n$ una m.a. de n observaciones de la variable aleatoria x con función de probabilidad Bernoulli:

$$f(x;p) = p^x(1-p)^{1-x} \qquad x = 0,1$$

$$\Rightarrow f_{\underline{x}}(x_1, x_2, ..., x_n) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} \cdot 1_{\{0,1\}}^{(x_i)} = p^{\sum x_i}(1-p)^{n-\sum x_i} \prod 1_{\{0,1\}}^{(x_i)} = g(S;p)h(x_1, x_2, ..., x_n)$$

$$\Rightarrow \begin{Bmatrix} h(x_1, x_2, ..., x_n) = \prod 1_{\{0,1\}}^{(x_i)} \\ g(S;p) = p^{\sum x_i}(1-p)^{n-\sum x_i} \end{Bmatrix} \Rightarrow S = \sum x_i \textit{ es la estadistica suficiente para p.}$$

Ejemplo: Sea $x_1, x_2, ..., x_n$ una m.a. Normal(μ, σ^2)

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

$$\begin{aligned} \Rightarrow f_{\underline{x}}(x_1, x_2, ..., x_n) &= \prod_{i=1}^{n} f(x;\mu,\sigma^2) = \left(\frac{1}{2\pi}\right)^{n/2} \sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2}\sum(x_i-\mu)^2\right\} = \\ &= \left(\frac{1}{2\pi}\right)^{n/2} \sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2}\left(\sum x_i^2 - 2\mu\sum x_i + n\mu^2\right)\right\} = \\ &= \left(\frac{1}{2\pi}\right)^{n/2} 1_{\{0,1\}}^{(x_i)} \cdot \underbrace{\sigma^{-n}\left[\exp\left\{-\frac{1}{2\sigma^2}\sum x_i^2\right\}\exp\left\{-\frac{1}{2\sigma^2}\left(-2\mu\sum x_i + n\mu^2\right)\right\}\right]}_{g\left(S_1, S_2;\mu,\sigma^2\right)} \end{aligned}$$

$$\Rightarrow \left\{ \begin{array}{c} h\left(x_1, x_2, ..., x_n\right) = \left(\frac{1}{2\pi}\right)^{n/2} 1_{\{0,1\}}^{(x_i)} \\ g\left(S_1, S_2; \mu, \sigma^2\right) = \sigma^{-n}\left[\exp\left\{-\frac{1}{2\sigma^2}\sum x_i^2\right\}\exp\left\{-\frac{1}{2\sigma^2}\left(-2\mu\sum x_i + n\mu^2\right)\right\}\right] \end{array} \right\} \Rightarrow S_1 = \sum x_i;\ S_2 = \sum x_i^2$$

$\Rightarrow \left(\sum x_i, \sum x_i^2\right)$ son conjuntamente suficientes para $\left(\mu, \sigma^2\right)$ y por lo tanto:

$\hat{\mu} = \overline{x} = \dfrac{\sum x_i}{n}$ y $\hat{\sigma}^2 = s^2 = \dfrac{\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}}{n-1}$ son funciones 1 a 1 de la estadística suficiente.

Un método para deducir un estadístico suficiente y minimal es el de Lehmann y Scheffé, que emplea la razón de verosimilitudes evaluadas en dos puntos $\left(x_1, x_2, ..., x_n\right)$ y $\left(y_1, y_2, ..., y_n\right)$:

$$\frac{L\left(x_1, x_2, ..., x_n; \theta\right)}{L\left(y_1, y_2, ..., y_n; \theta\right)}$$

Muchas veces es posible encontrar una función $\mathrm{g}\left(x_1, x_2, ..., x_n\right)$ tal que esta razón no depende de θ si y sólo si $\mathrm{g}\left(x_1, x_2, ..., x_n\right) = \mathrm{g}\left(y_1, y_2, ..., y_n\right)$. si es posible encontrar dicha función entonces $g\left(X_1, X_2, ..., X_n\right)$ es un estadístico suficiente minimal para θ.

Ejemplo: Sea $x_1, x_2, ..., x_n$ una m.a. de una función de probabilidad Bernoulli:

$$f\left(x; p\right) = p^x\left(1-p\right)^{1-x} \qquad x = 0,1$$

$$\frac{L\left(x_1, x_2, ..., x_n; \theta\right)}{L\left(y_1, y_2, ..., y_n; \theta\right)} = \frac{\prod_{i=1}^{n} p^{x_i}\left(1-p\right)^{1-x_i}}{\prod_{i=1}^{n} p^{y_i}\left(1-p\right)^{1-y_i}} = \frac{p^{\sum x_i}\left(1-p\right)^{n-\sum x_i}}{p^{\sum y_i}\left(1-p\right)^{n-\sum y_i}} = \left(\frac{p}{1-p}\right)^{\sum x_i - \sum y_i}$$

Para que esta razón no dependa de $\theta = p$, la única posibilidad es que

$$\sum x_i - \sum y_i = 0 \Rightarrow \sum x_i = \sum y_i \Rightarrow \text{ Pero } \mathrm{g}\left(x_1, x_2, ..., x_n\right) = \mathrm{g}\left(y_1, y_2, ..., y_n\right)$$

$$\Rightarrow \mathrm{g}\left(x_1, x_2, ..., x_n\right) = \sum x_i$$

$\therefore \sum x_i$ es una estadística suficiente y minimal para p.

Familia exponencial.

Definición. Familia exponencial de k parámetros.

Una familia de densidades $f(\bullet;\theta_1,\theta_2,\ldots,\theta_k)$ que puede expresarse como:

$$f(x;\theta_1,\theta_2,\ldots,\theta_k)=a(\theta_1,\theta_2,\ldots,\theta_k)b(x)\exp\left\{\sum_{j=1}^{k}c_j(\theta_1,\theta_2,\ldots,\theta_k)d_j(x)\right\}$$

se dice que pertenece a la familia exponencial.

Ejemplo: $\text{Bernoulli}(\theta)\qquad \theta=0,1$

$$f(x;\theta)=\theta^x(1-\theta)^{1-x}=e^{x\log\theta}e^{(1-x)\log(1-\theta)}=e^{x\log\theta}\cdot e^{\log(1-\theta)}e^{-x\log(1-\theta)}=e^{x(\log\theta-\log(1-\theta))}e^{\log(1-\theta)}$$
$$=e^{x\log\left(\frac{\theta}{1-\theta}\right)}e^{\log(1-\theta)}$$

Entonces, definimos:

$$\left.\begin{array}{l}a(\theta)=e^{\log(1-\theta)}\\ b(x)=1_{\{0,1\}}^{(x_i)}\\ c_1(\theta)=\log\left(\dfrac{\theta}{1-\theta}\right)\\ d_1(x)=x\end{array}\right\}\Rightarrow \textit{la Bernoulli es de la familia}\ \text{exponencial}$$

Ejemplo: Poisson (θ)

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

Entonces, definimos:

$$\left.\begin{array}{l}a(\theta)=e^{-\theta}\\ b(x)=\dfrac{1}{x!}\\ c_1(\theta)=\log\theta\\ d_1(x)=x\end{array}\right\}\Rightarrow \textit{la Poisson es de la familia}\ \text{exponencial}$$

Ejemplo: Normal$\left(\mu,\sigma^2\right)$

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

Entonces, definimos:

$$\left.\begin{array}{l} a\left(\mu,\sigma^2\right)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{\mu^2}{2\sigma^2}\right\} \\ b\left(x\right)=1_{\{-\infty,\infty\}}^{(x_i)} \\ c_1\left(\mu,\sigma^2\right)=\frac{\mu}{\sigma^2} \\ d_1(x)=x \\ c_2\left(\mu,\sigma^2\right)=-\frac{1}{2\sigma^2} \\ d_2\left(x\right)=x^2 \end{array}\right\}\Rightarrow \textit{la}\ \text{Normal}\left(\mu,\sigma^2\right)\textit{es de la familia}\ \text{exponencial}$$

Familia exponencial: Binomial, Exponencial, Beta, Gamma, Normal, χ^2 .

- **Resultado importante:** Si $f\left(\bullet;\theta\right)$ es de la familia exponencial entonces $\sum_{i=1}^{n}d\left(x_i\right)$ es una estadística suficiente para θ.

Algunas "cosas" sobre el estimador máximo verosímil:

- **Principio de invarianza.** Sea $\underline{\hat{\theta}}=\left(\hat{\theta}_1,\hat{\theta}_2,...,\hat{\theta}_k\right)$ el estimador máximo verosímil de $\underline{\theta}$. Si $\tau\left(\underline{\theta}\right)=\left(\tau_1\left(\theta_1\right),\tau_2\left(\theta_2\right),...,\tau_k\left(\theta_k\right)\right)$ es una transformación del espacio parametral, entonces el estimador máximo verosímil de $\tau\left(\underline{\theta}\right)$ es:
$$\tau\left(\underline{\hat{\theta}}\right)=\left(\tau_1\left(\hat{\theta}_1\right),\tau_2\left(\hat{\theta}_2\right),...,\tau_k\left(\hat{\theta}_k\right)\right)$$

Ejemplo. Poisson:

$$\hat{\theta} = \overline{x} \rightarrow estimador\ máximo\ verosímil$$
$$transformación: \tau(\theta) = P(x = 0) = e^{-\theta}$$
$$\tau(\hat{\theta}) = e^{-\hat{\theta}} = e^{-\overline{x}} \rightarrow estimador\ máximo\ verosímil\ de\ la\ transformación$$

- **Bajo condiciones de regularidad.** Un estimador máximo verosímil, que depende de n, se distribuye asintóticamente como:

$$\hat{\theta}_{MV(n)} \overset{a}{\sim} N\left(\theta, \frac{1}{nE\left\{\left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right)^2\right\}}\right)$$