Inferencia Estadística

Marisol García Peña

Departamento de Matemáticas
Pontificia Universidad Javeriana

Bogotá, 2022

Estimación puntual (cont.)

Ejemplo

Sea $X_1, \ldots, X_n$ una m.a. de una distribución Cauchy con $f(x, \theta) = \frac{1}{\pi(1+(x-\theta)^2)}$, para $-\infty < x < \infty$, $-\infty < \theta < \infty$.

$$L(\theta) = \frac{1}{\pi^n \prod_{i=1}^n [1 + (X_i - \theta)^2]}$$

$L(\theta)$ se maximiza cuando $\prod_{i=1}^n [1 + (X_i - \theta)^2]$ es mínimo o equivalentemente cuando $\sum_{i=1}^n \ln(1 + (X_i - \theta)^2)$ es mínimo. Para determinar el valor que minimiza esta expresión $\Longrightarrow$ métodos númericos.

Por ejemplo, suponga que las observaciones son $X_1 = 1, X_2 = X_3 = 2, X_4 = 3$, para maximizar $L(\theta)$, es necesario minimizar

$$\ln(1 + (1 - \theta)^2) + \ln(1 + (2 - \theta)^2) + \ln(1 + (2 - \theta)^2) + \ln(1 + (3 - \theta)^2)$$

Estimación puntual (cont.)

```
x <- c(1, 2, 2, 3)
g <- function(theta) sum(log(1 + (x-theta)^2))
optimize(g, interval = c(0, 4))
$minimum
[1] 2
$objective
[1] 1.386294

logL <- function(theta) sum(log(dcauchy(x, theta)))
optimize(logL, interval = c(0, 4), maximum = TRUE)
$maximum
[1] 2
$objective
[1] -5.965214
# La solución 2 es el estimador de \theta
```

Estimación puntual (cont.)

Ejemplo

Sean $X_1 \sim Pois(\lambda_1), X_2 \sim Pois(\lambda_2), X_3 \sim Pois(\beta\lambda_1)$ y $X_4 \sim Pois(\beta\lambda_2)$ variables aleatorias independientes y $\lambda_1, \lambda_2, \beta > 0$ parámetros desconocidos. Encontrar los EMV.

$$f(\boldsymbol{X}; \lambda_1, \lambda_2, \beta) = \frac{\lambda_1^{X_1+X_3}\lambda_2^{X_2+X_4}\beta^{X_3+X_4}e^{-(\lambda_1+\lambda_2)(\beta+1)}}{X_1!X_2!X_3!X_4!}$$

Suponga que se observan $(X_1, X_2, X_3, X_4) = (x_1, 3, 5, 7)$, x_1 es un valor perdido. $y = (3, 5, 7)$ denota los datos incompletos.

La función log-verosimilitud es

$$\begin{aligned}\ell(\lambda_1, \lambda_2, \beta|\boldsymbol{x}) = &(x_1 + x_3)\ln(\lambda_1) + (x_2 + x_4)\ln(\lambda_2)\\ &+ (x_3 + x_4)\ln(\beta) - (\lambda_1 + \lambda_2)(\beta + 1) + C\end{aligned}$$

donde C es una constante independiente de $\lambda_1, \lambda_2, \beta$.

Primero se calcula el valor esperado de la log-verosimilitud dado $y = (3, 5, 7)$ y estimar $\lambda_1^{(k)}$:

$$\begin{aligned}Q(\lambda_1, \lambda_2, \beta|\lambda_1^{(k)}) &= E[\ell(\lambda_1, \lambda_2, \beta|\boldsymbol{x})|y, \lambda_1^{(k)}]\\ &= (E[X_1|y, \lambda_1^{(k)}] + 5)\ln(\lambda_1) + 10\ln(\lambda_2)\\ &+ 12\ln(\beta) - (\lambda_1 + \lambda_2)(\beta + 1)\\ &= (\lambda_1^{(k)} + 5)\ln(\lambda_1) + 10\ln(\lambda_2) + 12\ln(\beta)\\ &- (\lambda_1 + \lambda_2)(\beta + 1)\end{aligned}$$

donde $E[X_1|y,\lambda_1^{(k)}] = E[X_1|\lambda_1^{(k)}] = \lambda_1^{(k)}$, pues X_i's se asumen independeintes. Al tomar las derivadas de Q con respecto a β, λ_1, y λ_2 e igualarlas a 0 se obtiene:

- $\beta^{(k+1)} = \dfrac{12}{\lambda_1^{(k)} + 3}$
- $\lambda_1^{(k+1)} = \dfrac{\lambda_1^{(k)+5}}{\beta^{(k+1)} + 1}$
- $\lambda_2^{(k+1)} = \dfrac{10}{\beta^{(k+1)} + 1}$

Si se toma como valor inicial $\lambda_1^{(0)} = 1$, entonces $\beta = 2.333, \lambda_1 = 2.143, \lambda_2 = 3.000$

Estimación puntual (cont.)

```
lambda1 <- 1 #initial value for lambda1

for (i in 1:20)
{
  beta <- 12/(lambda1 + 3)
  lambda1 <- (lambda1 + 5)/(beta + 1)
  lambda2 <- 10/(beta + 1)
}

beta
[1] 2.333355
lambda1
[1] 2.142829
lambda2
[1] 2.999981
```

Ejemplo

Suponga que una muestra aleatoria de tamaño n es seleccionada de una distribución Bernoulli con $P[X=x]=\theta^x(1-\theta)^{1-x}I_{\{0,1\}}(x)$, $0\leq\theta\leq 1$. Encuentre el EMV de θ.

$$L(\theta)=\prod_{i=1}^{n}\theta^{x_i}(1-\theta)^{1-x_i}=\theta^{\sum_{i=1}^{n}x_i}(1-\theta)^{n-\sum_{i=1}^{n}x_i},\quad S=\sum x_i$$

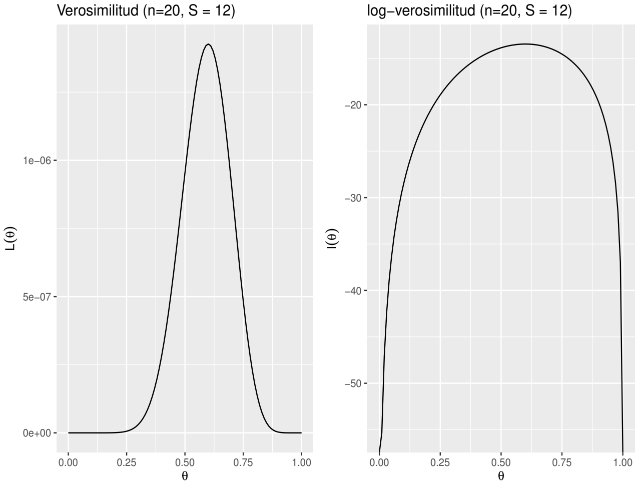
$$\ell(\theta)=S\ln(\theta)+(n-S)\ln(1-\theta)$$

Estimación puntual (cont.)

```
# Verosimilitud X_1,...,X_n ~ Bernoulli(theta)
L_bernoulli <- function(n, S){
    function(theta){
        theta ^ S * (1 - theta) ^ (n - S)
    }
}

# log-verosimilitud
l_bernoulli <- function(n, S){
    function(theta){
        S * log(theta) + (n - S) * log(1 - theta)
    }
}
```

Estimación puntual (cont.)



Estimación puntual (cont.)

Con $S = \sum x_i = 12$ y $n = 20$, $\widehat{\theta} = \overline{X} = \frac{12}{20} = 0.6$.

```
optimize(L_bernoulli(n=20, S=12), interval=c(0, 1), maximum=TRUE)
$ maximum
[1] 0.6

optimize(l_bernoulli(n=20, S=12), interval=c(0, 1), maximum=TRUE)
$ maximum
[1] 0.6
```

Estimación puntual (cont.)

Ejemplo

Energía eólica

- La cantidad de energía producida por una turbina depende de la velocidad del viento.
- Los ingenieros usan información sobre la velocidad del viento para determinar las ubicaciones adecuadas para construir una turbina eólica o para optimizar el diseño de una turbina.
- La distribución de Weibull es la más utilizada para modelar la velocidad del viento.
- La f.d.p. de una Weibull está dada por $f(x) = \frac{kx^{k-1}}{\lambda^k} e^{-(x/\lambda)^k}, x \geq 0$, con $k > 0$ parámetro de forma y $\lambda > 0$ parámetro de escala.

Estimación puntual (cont.)

- Usar la distribución para modelar las velocidades promedio del viento (m/s).
- Sitio de la turbina eólica de Carleton durante 168 días del 14 de febrero al 1 de agosto de 2010 (no hubo datos para el 2 de julio).

$$
\begin{aligned}
L(k;\lambda) &= \prod_{i=1}^{n} \frac{kx_i^{k-1}}{\lambda^k} e^{(-x_i/\lambda)^k} \\
&= \frac{k^n}{\lambda^{kn}} \prod_{i=1}^{n} x_i^{k-1} e^{-\sum_{i=1}^{n}(x_i/\lambda)^k}
\end{aligned}
$$

$$
\ell(k;\lambda) = n\ln(k) - kn\ln(\lambda) + (k-1)\sum_{i=1}^{n}\ln(x_i) - \sum_{i=1}^{n}\left(\frac{x_i}{\lambda}\right)^k
$$

Estimación puntual (cont.)

Calculando las derivadas parciales de $\ell(k;\lambda)$ con respecto a k y λ e igualando a 0, obtenemos:

$$\frac{\partial \ell(k;\lambda)}{\partial k} = \frac{n}{k} - n\ln(\lambda) + \sum_{i=1}^{n} \ln(x_i) - \sum_{i=1}^{n} \left(\frac{x_i}{\lambda}\right)^k \ln\left(\frac{x_i}{\lambda}\right) = 0$$

$$\frac{\partial \ell(k;\lambda)}{\partial \lambda} = -\frac{kn}{\lambda} + \frac{k}{\lambda^{k+1}} \sum_{i=1}^{n} x_i^k = 0$$

Recuerde, la derivada de c^x con respecto a x siendo c una constante, $c^x \ln c$.
De la segunda ecuación obtenemos

$$\lambda^k = \frac{1}{n} \sum_{i=1}^{n} x_i^k$$

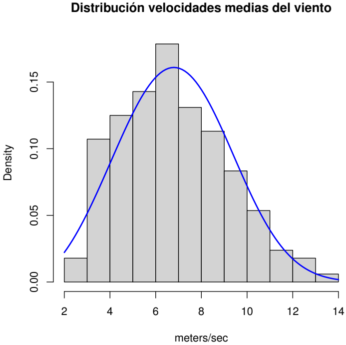
Sustituyendo esto en la primera ecuación

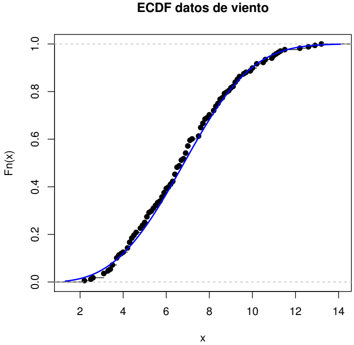
$$\frac{1}{k} + \frac{1}{n} \sum_{i=1}^{n} \ln(x_i) - \frac{1}{\alpha} \sum_{i=1}^{n} x_i^k \ln(x_i) = 0; \quad \alpha = \sum_{i=1}^{n} x_i^k$$

Estimación puntual (cont.)

- Usar métodos numéricos $\Longrightarrow$ valor aproximado para k, en la última ecuación.
- Usando los datos de Turbinas de Carleton.
- $\widehat{k} = 3.169$.
- $\lambda^{3.169} = \dfrac{1}{168}\displaystyle\sum_{i=1}^{168} x_i^{3.169}$.
- $\widehat{\lambda} = 7.661$.

Estimación puntual (cont.)





Estimación puntual (cont.)

Prueba chi-cuadrado

Interval	[0.0,3.77]	(3.77,4.77]	(4.77,5.53]	(5.53,6.20]	(6.20,6.80]
Observado	17	18	19	13	20
Esperado	16.8	16.8	16.8	16.8	16.8
Interval	(6.80,7.45]	(7.45,8.12]	(8.12,8.90]	(8.90,9.97]	$(9.97,\infty)$
Observado	14	17	17	14	19
Esperado	16.8	16.8	16.8	16.8	16.8

Se tienen 168 datos/puntos, luego se esperan 16.8 puntos en cada intervalo.

Estadística de prueba chi-cuadrado, $c = 3.071$ con p-valor$= 0.878$.

La distribución de los datos de velocidad del viento es consistente con una distribución Weibull con $\widehat{k} = 3.169$ y $\widehat{\lambda} = 7.661$.

Estimación puntual (cont.)

Método de mínimos cuadrados

Suponiendo que las observaciones Y_i pueden se escritas de la siguiente forma $Y_i = g_i(\theta_1, \ldots, \theta_k) + e_i$; $i = 1, 2, \ldots, n$ donde $g_i(\theta_1, \ldots, \theta_k)$. Son funciones conocidas y $\theta_1, \ldots, \theta_k$ son parámetros desconocidos de interés y $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$. Suponga también que e_i son v.a's que satisfacen los siguientes supuestos.

- $E[e_i] = 0$.
- $V[e_i] = \sigma^2, \forall i = 1, 2, \ldots, n$.
- $Cov[e_i, e_j] = 0, \forall i \neq j$

Estimación puntual (cont.)

El estimador de mínimos cuadrados del vetor de parámetros $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$ es el valor de $\boldsymbol{\theta}$ que minimiza la función

$$G(\boldsymbol{\theta}) = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} [y_i - g_i(\theta)]^2$$

Estimación puntual (cont.)

Observaciones

Si las funciones $g_i(\theta_1, \ldots, \theta_k)$ son diferenciables $\forall i = 1, 2, \ldots, n$, entonces el estimador de mínimos cuadrados de los parámetros $\boldsymbol{\theta}$ es obtenido resolviendo el sistema de ecuaciones

$$\begin{bmatrix} \frac{\partial G(\boldsymbol{\theta})}{\partial \theta_1} = 0 \\ \vdots \\ \frac{\partial G(\boldsymbol{\theta})}{\partial \theta_k} = 0 \end{bmatrix}$$

Estas ecuaciones se conocen como ecuaciones normales.

Para estimar $\sigma^2 \Longrightarrow E[CMRes]$

Estimación puntual (cont.)

Ejemplo

Considerando el modelo de regresión lineal simple $Y_i = \beta_0 + \beta_1 X_i + e_i; i = 1, 2, \ldots, n$, donde X_i es fijo y los errores son v.a's independientes tal que $E[e_i] = 0$, $V[e_i] = \sigma^2 \; \forall i = 1, \ldots, n$, $cov[e_i, e_j] = 0, \forall i \neq j$. Encontrar los estimadores de mínimos cuadrados.

$\boldsymbol{\theta} = (\beta_0, \beta_1)$, $g_i(\beta_0, \beta_1) = \beta_0 + \beta_1 X_i$, luego

$$G(\boldsymbol{\theta}) = \sum_{i=1}^{n} [y_i - (\beta_0 + \beta_1 x_i)]^2 \Longrightarrow \text{ Función a minimizar}$$

$$\frac{\partial G(\boldsymbol{\theta})}{\partial \beta_0} = \sum_{i=1}^{n} 2[y_i - (\beta_0 + \beta_1 x_i)](-1) = 0$$

$$\sum_{i=1}^{n} -2[y_i - (\beta_0 + \beta_1 x_i)] = 0 \quad (i)$$

$$\frac{\partial G(\boldsymbol{\theta})}{\partial \beta_1} = \sum_{i=1}^{n} 2[y_i - (\beta_0 + \beta_1 x_i)](-x_i) = 0$$

$$\sum_{i=1}^{n} -2x_i[y_i - (\beta_0 + \beta_1 x_i)] = 0 \quad (ii)$$

Estimación puntual (cont.)

$$\sum_{i=1}^{n} y_i - n\beta_0 - \beta_1 \sum_{i=1}^{n} x_i = 0 \quad (i)$$

$$\sum_{i=1}^{n} x_i y_i - \beta_0 \sum_{i=1}^{n} x_i - \beta_1 \sum_{i=1}^{n} x_i^2 = 0 \quad (ii)$$

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{XY}}{\sum_{i=1}^{n} x_i^2 - n\overline{X}^2}$$

$$\widehat{\beta}_0 = \frac{\sum_{i=1}^{n} y_i - \widehat{\beta}_1 \sum_{i=1}^{n} x_i}{n} = \overline{Y} - \widehat{\beta}_1 \overline{X}$$

Estimación puntual (cont.)

Método de chi-cuadrado mínimo

Sea $X_1, \ldots, X_n$ una m.a. a partir de una densidad $f(x;\theta)$ y sea $\mathcal{L}_1, \ldots, \mathcal{L}_k$ una partición del rango de X. La probabilidad de que una observación pertenezca a la celda $\mathcal{L}_j$, $j = 1, \ldots, k$ denotada por $p_j(\theta)$ puede calcularse. Recuerde que $\displaystyle\sum_{j=1}^{k} p_j(\theta) = 1$.

Sea N_j la v.a. que denota el número de X_i's en la muestra que pertenecen a la celda $\mathcal{L}_j$, $j = 1, \ldots, k$; entonces $\displaystyle\sum_{j=1}^{k} N_j = n$, n es el tamaño de la muestra.

Estimación puntual (cont.)

$$\chi^2 = \sum_{j=1}^{k} \frac{[n_j - np_j(\theta)]^2}{np_j(\theta)}$$

n_j es un valor de N_j.

- Númerador $\Longrightarrow$ cuadrado de la diferencia entre valores observados y esperados de observaciones que pertenecen a la celda $\mathcal{L}_j$.
- La estimación de chi-cuadrado mínimo de θ es $\widehat{\theta}$ tal que minimiza χ^2.
- El estimador de chi-cuadrado mínimo depende de la partición $\mathcal{L}_1, \ldots, \mathcal{L}_k$ seleccionada.

Estimación puntual (cont.)

Ejemplo

Sea $X_1, \ldots, X_n$ una m.a. de una distribución Bernoulli, $P[X = x; \theta] = \theta^x (1-\theta)^{1-x}$ para $x = 0, 1$. N_j es el número de observaciones iguales a j para $j = 0, 1$. El rango de observación de X es dividido en dos conjuntos 0 y 1.

$$\begin{aligned}
\chi^2 &= \sum_{j=0}^{1} \frac{[n_j - np_j(\theta)]^2}{np_j(\theta)} = \frac{[n_0 - n(1-\theta)]^2}{n(1-\theta)} + \frac{(n_1 - n\theta)^2}{n\theta} \\
&= \frac{[(n-n_1) - n(1-\theta)]^2}{n(1-\theta)} + \frac{(n_1 - n\theta)^2}{n\theta} = \frac{(n_1 - n\theta)^2}{n} \frac{1}{\theta(1-\theta)}
\end{aligned}$$

Estimación puntual (cont.)

- El mínimo de la función χ^2 se encuentra cuando $\theta = \frac{n_1}{n}$, es decir, $\widehat{\theta}$.
- En este caso, solo hay una selección posible de la partición.
- El estimador obtenido es el mismo que el del método de los momentos y máxima verosimilitud.
- Puede cambiarse $np_j(\theta)$ por $n_j \Longrightarrow$ Método de chi-cuadrado mínimo modificado.

Estimación puntual (cont.)

Método de la distancia mínima

Sea $X_1, \ldots, X_n$ una m.a. de una distribución con $F(x;\theta)$ y $d(F,G)$ es la distancia que mide que tan separadas están dos funciones de distribución acumulada F y G. Por ejemplo, $d(F,G) = \sup_x |F(x) - G(x)|$.

- El estimador de θ, $\widehat{\theta}$ es aquel que minimiza $d(F(x;\theta), F_n(x))$, donde $F_n(x)$ es la función de distribución acumulada muestral.

- $\widehat{\theta}$ es seleccionada tal que $F(x;\widehat{\theta})$ este más cerca a $F_n(x)$.

Estimación puntual (cont.)

Ejemplo

Sea $X_1, \ldots, X_n$ una m.a. de una distribución Bernoulli, entonces $F(x,\theta) = (1-\theta)I_{[0,1)}(x) + I_{[1,\infty)}(x)$, donde $0 \leq \theta \leq 1$.

Sea n_j el número de observaciones iguales a $j, j = 0, 1$, entonces $F_n(x) = \frac{n_0}{n} I_{[0,1)}(x) + I_{[1,\infty)}(x)$.

Usando la función distancia $d(F,G) = \underset{\text{sup}}{x}|F(x) - G(x)|$, $d(F(x;\theta), F_n(x))$ se minimiza si $1-\theta = \frac{n_0}{n}$ o $\theta = \frac{n_1}{n} = \sum \frac{x_i}{n}$, es decir, $\widehat{\theta} = \overline{x}$.

Estimación puntual (cont.)

Propiedades de un estimador puntual

Más concentrado y el más concentrado

Sean $T(X_1, \ldots, X_n)$ y $T'(X_1, \ldots, X_n)$ dos estimadores de $\tau(\theta)$. T' es un estimador más concentrado de $\tau(\theta)$ que T si y sólo si

$$P_\theta[\tau(\theta) - \lambda < T' \leq \tau(\theta) + \lambda] \geq P_\theta[\tau(\theta) - \lambda < T \leq \tau(\theta) + \lambda], \forall \lambda > 0$$

Un estimador $T^*(X_1, \ldots, X_n)$ es *el más concentrado* si es más concentrado que cualquier otro estimador.

$P_\theta[\bullet]$ indica que la probabilidad depende de θ.

Estimación puntual (cont.)

- La concentración es una propiedad deseable en un estimador $\Longrightarrow$ raramente existe.
- Restringir los posibles estimadores $\Longrightarrow$ Estimador con otras propiedades deseables $\Longrightarrow$ Estimador más concentrado en la restricción.
- Existen otros criterios para comparar estimadores.

Estimador más cercano de Pitman y el más cercano de Pitman

Sean $T(X_1, \ldots, X_n)$ y $T'(X_1, \ldots, X_n)$ dos estimadores de $\tau(\theta)$. T' es el estimador de $\tau(\theta)$ más cercano de Pitman si y sólo si

$$P_\theta[|T' - \tau(\theta)| < |T - \tau(\theta)|] \geq \frac{1}{2}$$

Un estimador $T^*(X_1, \ldots, X_n)$ es *el más cercano de Pitman* si es más cercano de Pitman que cualquier otro estimador.

Al igual que la propiedad de concentración, un estimador de Pitman $\Longrightarrow$ raramente existe.

Estimación puntual (cont.)

Error cuadrático medio - $ECM/MSE(T(\theta))$

Sea $T(X_1,\ldots,X_n)$ un estimador de $\tau(\theta)$. $E_\theta[[T-\tau(\theta)]^2]$ es el error cuadrático medio del estimador T.

$$
\begin{aligned}
ECM(T(\theta)) &= ECM(\widehat{\theta}) = E[[T-\tau(\theta)]^2] \\
&= E[(T-E[T]+E[T]-\tau(\theta))^2] \\
&= E[((T-E[T])+(E[T]-\tau(\theta)))^2] \\
&= E[(T-E[T])^2]+2E[T-E[T]][E[T]-\tau(\theta)]+(E[T]-\tau(\theta))^2 \\
&= V[T]+[E[T]-\tau(\theta)]^2 = V[T]+(b_T[\tau(\theta)])^2
\end{aligned}
$$

$b_T[\tau(\theta)]$ se conoce como sesgo del estimador T y puede ser positivo, negativo o cero.

Estimación puntual (cont.)

- ECM mide la distancia cuadrática promedio entre el estimador y el parámetro. Combina sesgo y varianza.
- Medida de dirpersión de T alrededor de $\tau(\theta) \Longrightarrow$ varianza de una v.a. con respecto a su media.
- Cuanto menor el valor del ECM $\Longrightarrow$ mejor el estimador.
- El *ECM* es la suma de dos cantidades no negativas.
- Muestra como esta relacionado el *ECM* con la varianza y el sesgo de un estimador.

Estimación puntual (cont.)

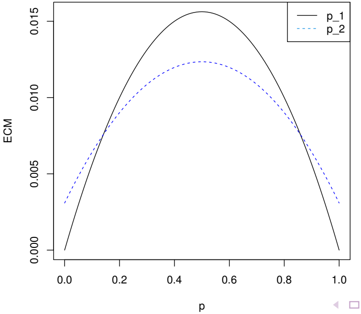
Ejemplo

Sea $X \sim Bin(n, p)$, n conocido y p desconocido. La proporción muestral $\widehat{p}_1 = \frac{X}{n}$ es un estimador de p, con $E[\widehat{p}_1] = p$, $V[\widehat{p}_1] = \frac{p(1-p)}{n}$ y $ECM[\widehat{p}_1] = \frac{p(1-p)}{n}$.

Un estimador alternativo de p es $\widehat{p}_2 = \frac{X+1}{n+2}$. Encuentre $E[\widehat{p}_2]$, $V[\widehat{p}_2]$, $b_{\widehat{p}_2}[p]$ y $ECM[\widehat{p}_2]$ y determine cuál estimador es mejor respecto al ECM.

Estimación puntual (cont.)

```
n <- 16
curve(x*(1-x)/n, from=0, to=1, xlab= "p", ylab="CME")
curve(n*(1-x)*x/(n+2)^2 + (1-2*x)^2/(n+2)^2, add=TRUE,
col="blue", lty=2)
legend("topright", c("p_1","p_2"), col=c(1,4),lty = c(1,2))
```



Estimación puntual (cont.)

Ejemplo

Sea $X_1, \ldots, X_n$ una m.a. de una población $N(\mu, \sigma^2)$. Considere los EMV de μ y σ^2 dados por

$$\widehat{\mu} = \frac{\sum_{i=1}^n X_i}{n} = T_1(\boldsymbol{X}) \qquad \widehat{\sigma}^2 = \frac{\sum_{i=1}^n (X_i - \overline{X})^2}{n} = T_2(\boldsymbol{X})$$

Se pide:

- Encontrar $ECM(\mu, \widehat{\mu})$ y el sesgo de $\widehat{\mu}$.
- Encontrar $ECM(\sigma^2, \widehat{\sigma}^2)$ y el sesgo de $\widehat{\sigma}^2$.