$4^{\underline{a}}$ P.A.

Inferencia Estadística

2 de noviembre de 2023

ENUNCIADO

Sea X una variable aleatoria con función de densidad $f(x) = (a + b{\cdot}x) \cdot I(-1 \leq x \leq 1)$

1. Comprueba que $a = 0{,}5$ y $|b| \leq 0{,}5$.
2. Calcula su media y su varianza.
3. Comprueba que la media no es un estadístico suficiente. Busca un estadístico suficiente minimal.
4. Genera al azar un valor de b y genera una muestra aleatoria de tamaño $n = 500$.
5. Calcula las estimaciones por el método de los momentos (EMM) y el de máxima verosimilitud (EMV).
6. Compara el EMV con el EMM.
7. Calcula la distribución asintótica del EMV.
8. Calcula una expresión que aproxime el EMV.

RESOLUCIÓN

1. Comprueba que $a = 0{,}5$ y $|b| \leq 0{,}5$.

Una densidad f debe cumplir que $f \geq 0$ y $\int f = 1$. Entonces

$$\begin{aligned} 1 &= \int_{-1}^{1} (a + bx)\, dx = \left[ax + \frac{b}{2}x^2\right]_{-1}^{1} = \left[a + \frac{b}{2}\right] - \left[-a + \frac{b}{2}\right] = 2a \\ &\Rightarrow \ a = 1/2 = 0{,}5 \end{aligned}$$

$$\forall x,\ f(x) \geq 0 \quad \Longrightarrow \quad \left\{ \begin{array}{lcr} f(1) = \dfrac{1}{2} + b \geq 0 & \Longrightarrow & b \geq -0{,}5 \\ f(-1) = \dfrac{1}{2} - b \geq 0 & \Longrightarrow & b \leq 0{,}5 \end{array} \right\} \quad \Longrightarrow \quad |b| \leq 0{,}5$$

2. Calcula su media y su varianza.

$$E(X) = \int x\, f(x)\, dx = \int_{-1}^{1} x\left(\frac{1}{2} + bx\right) dx = \left[\frac{x^2}{4} + \frac{b}{3}x^3\right]_{-1}^{1} = \frac{2}{3}\, b$$

$$\begin{aligned} \mathrm{Var}(X) &= E\left(X^2\right) - E^2(X) = \int x^2\, f(x)\, dx - \left(\frac{2}{3}\, b\right)^2 = \int_{-1}^{1} x^2\left(\frac{1}{2} + bx\right) dx - \frac{4}{9}\, b^2 \\ &= \frac{1}{3} - \frac{4}{9}\, b^2 \end{aligned}$$

3. Comprueba que la media no es un estadístico suficiente. Busca un estadístico suficiente minimal.

Para $-1 \leq x \leq 1$,

$$f(x, b) = 0{,}5 + b{\cdot}x \quad \Longrightarrow \quad \mathcal{L}(\vec{x}, b) = \prod_{i=1}^{n}(0{,}5 + b{\cdot}x_i)$$

Para ver que la media muestral no es suficiente para b, consideramos dos muestras $\vec{x}$ y $\vec{y}$ tales que $\bar{x} = \bar{y}$, por ejemplo $\vec{x} = (0, 0)$ y $\vec{y} = (-1/2, 1/2)$:

$$\frac{\mathcal{L}(\vec{x}, b)}{\mathcal{L}(\vec{y}, b)} = \frac{\prod_{i=1}^{n}(0{,}5 + b{\cdot}x_i)}{\prod_{i=1}^{n}(0{,}5 + b{\cdot}y_i)} = \frac{0{,}5 \cdot 0{,}5}{(0{,}5 - b/2) \cdot (0{,}5 + b/2)} = \frac{0{,}5^2}{0{,}25 - b^2/4} \text{ depende de } b$$

De hecho, el numerador $\prod_{i=1}^{n}(0{,}5 + b{\cdot}x_i)$ y el denominador $\prod_{i=1}^{n}(0{,}5 + b{\cdot}y_i)$ son sendos polinomios en b de grado $\leq n$, de forma que su cociente sería independiente de b si y sólo si coincidieran sus grados y sus raíces,

$$\left\{\frac{-0{,}5}{x_i} \;\middle|\; x_i \neq 0; i = 1, \ldots, n\right\} = \left\{\frac{-0{,}5}{y_i} \;\middle|\; y_i \neq 0; i = 1, \ldots, n\right\}$$

luego si $\vec{x}_{(\cdot)} = \vec{y}_{(\cdot)}$. Por tanto, un estadístico minimal suficiente sería la muestra ordenada, $\vec{X}_{(\cdot)}$.

4. Genera al azar un valor de b y genera una muestra aleatoria de tamaño $n = 500$.

Usaremos el método de la trasformación inversa. La función de distribución para $-1 \leq x \leq 1$ es

$$F(x) = \int_{-\infty}^{x} f = \frac{bx^2 + x - b + 1}{2}$$

y su inversa

$$u = F(x) \quad \Longrightarrow \quad x = F^{-1}(u) = \frac{\sqrt{8bu + 4b^2 - 4b + 1} - 1}{2b}$$

porque, de las dos ramas de la parábola F, hay que considerar la creciente, que es la de la izquierda si $b < 0$ y la de la derecha si $b > 0$ (si $b = 0$, entonces se trata de una recta y hay una sola "rama"). Por tanto, de las dos soluciones de la ecuación cuadrática,

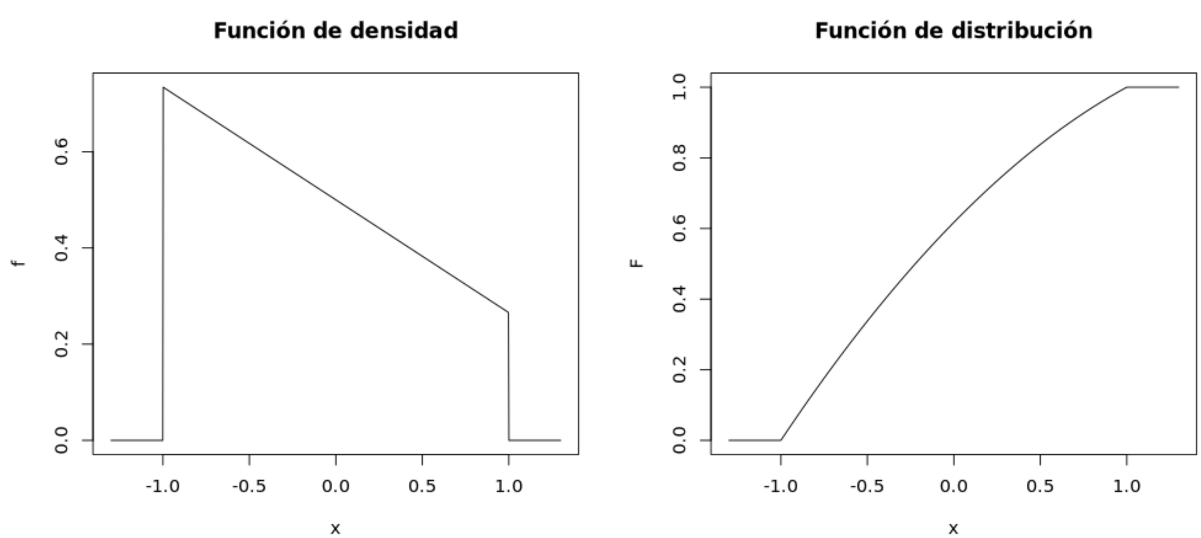
$$x_1 = \frac{-1 + \sqrt{8bu + 4b^2 - 4b + 1}}{2b} \qquad x_2 = \frac{-1 - \sqrt{8bu + 4b^2 - 4b + 1}}{2b}$$

la correcta es x_1 porque

$$b > 0 \Rightarrow x_1 > x_2 \qquad b < 0 \Rightarrow x_1 < x_2$$

En R podemos representar las funciones así:

```
set.seed(1)            # para reproducibilidad
b <- runif(1, -.5, .5) # valor teórico de b
## función de densidad
f <- function (x) (.5 + b*x) * (x >= -1) * (x <= 1)
## función de distribución
F <- function (x) (b*x^2 + x - b + 1) / 2 * (x >= -1) * (x <= 1) +
                  0 * (x < -1) + 1 * (x > 1)
plot (f, -1.3, 1.3, n=1000, main="Función de densidad")
plot (F, -1.3, 1.3, n=1000, main="Función de distribución")
## otra forma
F <- function (x) integrate (f, -Inf, x) $ value
```



En este caso, el b generado ha resultado negativo. La función de distribución se corresponde con la rama izquierda de una parábola cuyo coeficiente cuadrático es negativo.

La implementación del método para generar la muestra aleatoria sería:

```
F1 <- function(u) (sqrt(8*b*u + 4*b^2 - 4*b + 1) - 1) / (2*b)
n  <- 500
u  <- runif(n)
x  <- F1(u)
summary(x)
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -0.9975 -0.6263 -0.2645 -0.1648  0.2483  0.9853
2/3 * b       # -0.15632 media teórica
var(x)        # 0.302887 cuasivarianza muestral
1/3 - 4/9*b^2 # 0.308895 varianza teórica
```

5. Calcula las estimaciones por el método de los momentos (EMM) y el de máxima verosimilitud (EMV).

El valor teórico del parámetro simulado es

```
> b
[1] -0.2344913
```

5.1. Método de los momentos

$$E(X) = \frac{2}{3}\,b \quad \Longrightarrow \hat{b}_{\text{MM}} = \frac{3}{2}\,\bar{X}$$

En nuestra muestra:

```
bMM <- 3/2 * mean(x) # -0.2471665
```

5.2. Máxima verosimilitud

Para $-1 \leq x \leq 1$,

$$\mathcal{L}(\vec{x}, b) = \prod_{i=1}^{n}(0{,}5 + b{\cdot}x_i) \quad \Longrightarrow \quad \ln \mathcal{L}(\vec{x}, b) = \sum_{i=1}^{n} \ln(0{,}5 + b{\cdot}x_i)$$

Obtenemos la estimación de forma numérica, bien maximizando directamente la logverosimilitud,

```
# defino logL con sapply para vectorizarla en b porque x es un vector
logL <- function(b) sapply(b, function (bj) sum(log(.5 + bj*x)))
bMV  <- optimize (logL, c(-.5, .5), maximum=TRUE) $ maximum # -0.244692
```

bien hallando un cero de la ecuación de verosimilitud:

```
derL <- function(b) sapply(b, function (bj) sum(x / (.5 + bj*x)))
uniroot (derL, c(-.5,.5)) $ root # -0.244697
```

6. Compara el EMV con el EMM.

Haremos 1000 repeticiones de cada procedimiento y calcularemos el error cuadrático medio (ECM) de cada uno.

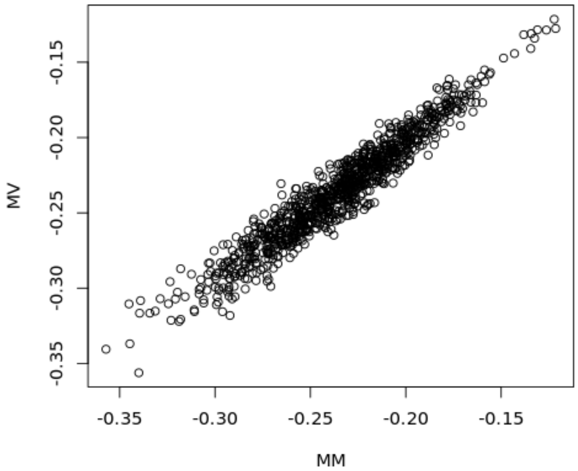
```
set.seed(1)              # para reproducibilidad
n  <- 500                # tamaño muestral
b <- runif(1, -.5, .5)   # valor teórico de b
F1 <- function(u) (sqrt(8*b*u + 4*b^2 - 4*b + 1) - 1) / (2*b) # F inversa
estimas <- replicate (1000,
{
  x  <- F1(runif(n))
  bMM <- 3/2 * mean(x)
  logL <- function(b) sapply(b, function (bj) sum(log(.5 + bj*x)))
  bMV  <- optimize (logL, c(-.5, .5), maximum=TRUE) $ maximum
  c (MM = bMM, MV = bMV)
})
ECM <- rowMeans ((estimas-b)^2) # MM 0.00135 ; MV 0.00126
```

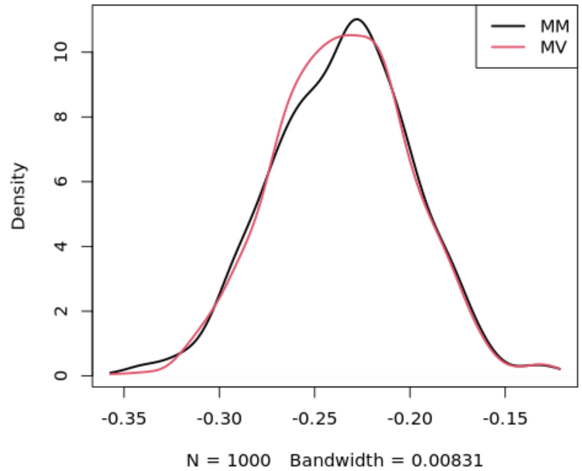
Los ECM son parecidos, pero el del método MV es inferior al del MM. La correlación entre ambas estimaciones es muy alta:

```
> cor(t(estimas))
          MM        MV
MM 1.0000000 0.9624636
MV 0.9624636 1.0000000
```

Gráficamente, también las densidades estimadas se parecen mucho.

```
> plot (t(estimas)) # nube de puntos
> plot (densidad (estimas["MM",]), lwd=2, main="")
> lines (densidad (estimas["MV",]), lwd=2, col=2)
> legend ("topright", c("MM","MV"), lwd=2, col=1:2)
> savePlot ("densidades.png")
```



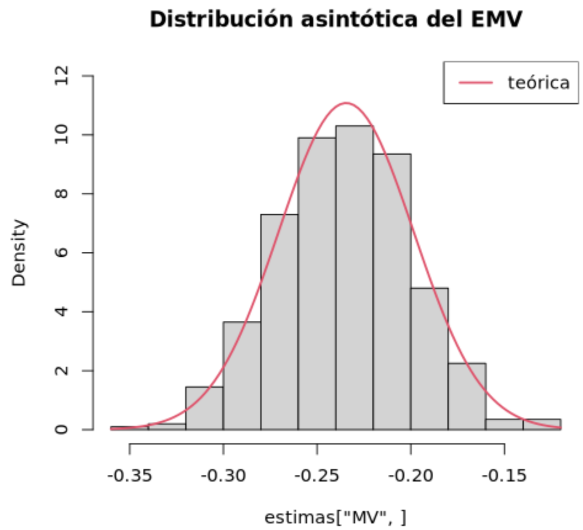


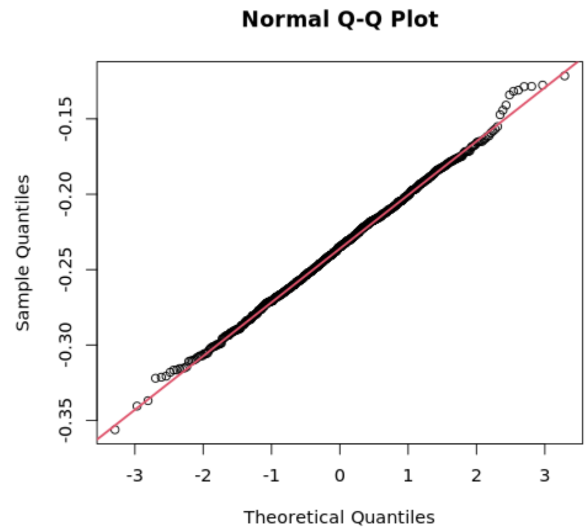
7. Calcula la distribución asintótica del EMV.

La distribución de $\hat{b}_{\text{MV}}$ converge en ley a una $N\left(b, \frac{1}{\sqrt{n\,I(b)}}\right)$ donde

$$I(b) = E\left[\left(\frac{\partial}{\partial b}\ln f(X,b)\right)^2\right] = -E\left[\frac{\partial^2}{\partial b^2}\ln f(X,b)\right] = \frac{\ln(2\,b+1) - 4\,b - \ln(1-2\,b)}{4\,b^3}$$

```
hist (estimas["MV",], prob=TRUE, ylim=c(0,12),
  main="Distribución asintótica del EMV")
i <- function (b) (log(2*b+1) - 4*b - log(1-2*b)) / (4 * b^3) # info fisher
sigma <- function (b) 1/sqrt(n*i(b))
curve (dnorm(x,b,sigma(b)), add=TRUE, col=2, lwd=2)
legend ("topright", "teórica", col=2, lwd=2)
qqnorm (estimas["MV",])
qqline (estimas["MV",], col=2, lwd=2)
```





8. Calcula una expresión que aproxime el EMV.

$$\ell(b) = \ln \mathcal{L}(\vec{x}, b) = \sum_{i=1}^{n} \ln(0{,}5 + b{\cdot}x_i)$$

Sea b_0 cercano a $\hat{b}_{\mathrm{MV}}$, por ejemplo, $b_0 = \hat{b}_{\mathrm{MM}}$. Sea $\beta = \hat{b}_{\mathrm{MV}} - b_0$, luego $\hat{b}_{MV} = b_0 + \beta$.

$$0 = \frac{\partial \ell}{\partial b}\left(\hat{b}_{\mathrm{MV}}\right) = \sum \frac{x_i}{0{,}5 + \hat{b}_{\mathrm{MV}} x_i} = \sum \frac{x_i}{0{,}5 + (b_0 + \beta)\, x_i} = \sum \frac{x_i}{\underbrace{0{,}5 + b_0 x_i}_{=:f_i} + \beta x_i} = \sum \underbrace{\frac{x_i}{f_i + \beta x_i}}_{=:g_i(\beta)}$$

Aproximando

$$g_i(\beta) \approx g_i(0) + \beta\, g_i'(0) = \frac{x_i}{f_i} - \beta \left(\left. \frac{x_i^2}{(f_i + \beta x_i)^2} \right|_{\beta=0} \right) = \frac{x_i}{f_i} - \beta\, \frac{x_i^2}{f_i^2} \quad \Rightarrow \quad 0 \approx \sum_i \left(\frac{x_i}{f_i} - \beta\, \frac{x_i^2}{f_i^2} \right)$$

luego

$$\hat{b}_{\mathrm{MV}} \approx b_0 + \hat{\beta} = b_0 + \frac{\sum x_i / f_i}{\sum x_i^2 / f_i^2}$$

Con nuestra muestra simulada:

```
> bMM
[1] -0.2471665
> b0 <- bMM
> fi <- .5 + b0*x
> b0 + sum(x/fi) / sum(x^2/fi^2) # aproximación de bMV
[1] -0.2447067
> bMV               # obtenido por maximización directa
[1] -0.244692
```

MAXIMA

Podemos emplear Maxima (`http://maxima.sourceforge.io`) como ayuda en los cálculos.

1. Comprueba que $a = 0{,}5$ y $|b| \leq 0{,}5$.

```
(%i8) f(x) := a+b*x $

(%i9) integrate (f(x), x, -1, 1) ;
                              b + 2 a   b - 2 a
(%o9)                         ------- - -------
                                 2         2
(%i10) solve (%=1, a) ;
                                       1
(%o10)                            [a = -]
                                       2
(%i11) a : 1/2 $

(%i12) load (solve_rat_ineq) $

(%i14) solve_rat_ineq (f(1) >= 0) ;
                                        1
(%o14)                         [[b >= - -]]
                                        2
(%i15) solve_rat_ineq (f(-1) >= 0) ;
                                       1
(%o15)                          [[b <= -]]
                                       2
```

2. Calcula su media y su varianza.

```
(%i16) esperanza : integrate (x * f(x), x, -1, 1) ;
                              4 b + 3   4 b - 3
(%o16)                        ------- + -------
                                12        12

(%i17) expand(%) ;
                                     2 b
(%o17)                               ---
                                      3
```

```
(%i18) varianza : integrate (x^2 * f(x), x, -1, 1) - esperanza^2 ;
                      4 b + 3   4 b - 3 2    3 b + 2   3 b - 2
(%o18)           (- (------- + -------) ) + ------- - -------
                        12        12           12        12
(%i19) expand(%) ;
                                            2
                                     1   4 b
(%o19)                               - - ----
                                     3    9
```

3. **Comprueba que la media no es un estadístico suficiente. Busca un estadístico suficiente minimal.**

4. **Genera al azar un valor de b y genera una muestra aleatoria de tamaño $n = 500$.**

obtenibles mediante Maxima así:

```
(%i20) F(x) := integrate (f(t), t, -1, x) ;
(%o20)               F(x) := integrate(f(t), t, - 1, x)

(%i21) F(x) ;
                                    2
                                 b x  + x   b - 1
(%o21)                           -------- - -----
                                    2         2

(%i22) expand(%) ;
                                    2
                                 b x    x   b   1
(%o22)                           ---- + - - - + -
                                  2     2   2   2

(%i23) ratsimp(%) ;
                                    2
                                 b x  + x - b + 1
(%o23)                           ----------------
                                        2
```

```
(%i24) solve (u=F(x), x) ;
                                   2
               sqrt(8 b u + 4 b  - 4 b + 1) + 1
(%o24) [x = - --------------------------------,
                              2 b
                                                        2
                                       sqrt(8 b u + 4 b  - 4 b + 1) - 1
                                   x = --------------------------------]
                                                     2 b

(%i25) /* ejemplo */ %, b=.2, u=.3 ;
(%o25)         [x = - 4.791287847477919, x = - 0.20871215252208]
(%i26) /* vale la derecha porque -1 < x < 1 */
```

5. Calcula las estimaciones por el método de los momentos (EMM) y el de máxima verosimilitud (EMV).

6. Compara el EMV con el EMM.

7. Calcula la distribución asintótica del EMV.

```
(%i41) assume(b < .5, b > -.5);
(%o41)                      [b < 0.5, b > - 0.5]

(%i42) facts();
(%o42)                      [0.5 > b, b > - 0.5]

(%i43) f(x);
                                      1
(%o43)                          b x + -
                                      2

(%i44) diff(log(%),b);
                                   x
(%o44)                          -------
                                      1
                                b x + -
                                      2
```

```
(%i45) integrate(%^2 * f(x), x, -1, 1) ;
Is b positive or negative?

positive;
                               2            2
             log(2 b + 1) + 2 b  - 2 b   2 b  + 2 b + log(1 - 2 b)
(%o45)       ------------------------- - -------------------------
                          3                            3
                       4 b                          4 b

(%i46) expand(%);
                       log(2 b + 1)   1    log(1 - 2 b)
(%o46)                 ------------ - -- - ------------
                             3        2          3
                          4 b        b        4 b

(%i47) ratsimp(%);
                       log(2 b + 1) - 4 b - log(1 - 2 b)
(%o47)                 ---------------------------------
                                        3
                                     4 b

(%i48) tex(%);
$${{\log \left(2\,b+1\right)-4\,b-\log \left(1-2\,b\right)}\over{4\,b
 ^3}}$$
(%o48)                              false

(%i49) /* otra forma */
(%i50) -diff(log(f(x)),b,2);
                                      2
                                     x
(%o50)                           ----------
                                        1 2
                                 (b x + -)
                                        2
```

```
(%i51) ratsimp (integrate(% * f(x), x, -1, 1)) ;
Is b positive or negative?

positive;
                       log(2 b + 1) - 4 b - log(1 - 2 b)
(%o51)                 ---------------------------------
                                        3
                                     4 b
```

8. Calcula una expresión que aproxime el EMV.