

Estimación paramétrica

Alvaro José Álvarez Arranz

04/01/2024

Modelos estadísticos

La muestra

Sea X_1, \dots, X_n un conjunto de n variables (o vectores, o funciones) aleatorias. Vamos a suponer, de aquí en adelante, X_1, \dots, X_n i.i.d. con distribución F .

Modelo

Formular un modelo estadístico es especificar un conjunto de distribuciones al que pertenece F .

- Un modelo es **paramétrico** si cada distribución del conjunto es totalmente conocida salvo por el valor de un conjunto de parámetros $\theta \in \mathbb{R}^d$.

$$F \in \{F_\theta : \theta \in \Theta \subset \mathbb{R}\}$$

- **Identificabilidad:** si $\theta \neq \theta'$, entonces $F_\theta \neq F_{\theta'}$ ***Modelos no paramétricos:** por ejemplo,

$$F \in \{F : F \text{ tiene función de densidad } f\}$$

Estimación puntual

Sea X_1, \dots, X_n una muestra aleatoria simple de una v.a. X cuya distribución de probabilidad depende de un parámetro θ desconocido que toma valores en el espacio paramétrico Θ .

Un **estimador puntual**, $\hat{\theta}$, de un parámetro θ es una función real de la muestra (un estadístico), $T(X_1, \dots, X_n)$, que aproxima el valor de θ .

Por tanto, es una variable aleatoria y podemos estudiar su esperanza, su varianza o, en general, su distribución de probabilidad. Su distribución de probabilidad dependerá de X y recoge todas las posibles estimaciones del parámetro θ y sus respectivas probabilidades de ocurrencia.

Una **estimación (puntual)** es el valor numérico concreto que toma un estimador al ser aplicada a una realización muestral x_1, \dots, x_n concreta observada.

Es el valor más plausible para $\theta \in \Theta$, teniendo en cuenta la distribución de probabilidad supuesta y los datos observados.

Tanto el estimador puntual como la estimación se denotan utilizando el símbolo: $\hat{\cdot}$ (por ejemplo: $\hat{\mu}, \hat{\sigma}, \hat{p}, \hat{\lambda}$)

Planteamiento general

Disponemos de una muestra aleatoria simple X_1, \dots, X_n de una v.a. X :

- Las X_1, \dots, X_n son **independientes**.
- Todas ellas tienen **la misma distribución** que X .

Supongamos que la distribución de X es conocida salvo por el valor de un conjunto de *parámetros* que denotamos por θ .

Objetivo: Aproximar el valor de θ a partir de la muestra. Para ello se necesita calcular un *estimador* $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$.

Preguntas:

- ¿Qué propiedades debe tener un buen estimador?
- ¿Existen métodos generales para obtener estimadores?

Vamos a responder a estas preguntas.

Criterios para valorar un estimador

- Un estimador $\hat{\theta} = T(X_1, \dots, X_n)$ es también una variable aleatoria cuya distribución se denomina **distribución en el muestreo** del estimador.
- Esta distribución informa de los valores que podemos esperar que tome $\hat{\theta}$ si dispusiéramos de muchas muestras de la misma población
- La distribución en el muestreo determina la calidad de un estimador.

Sesgo y varianza

Una buena propiedad de un estimador es que no tenga tendencia sistemática a infraestimar o sobreestimar el parámetro.

Se dice que un estimador $\hat{\theta}$ es **insesgado** si $E(\hat{\theta}) = \theta$, para todo $\theta \in \Theta$.

En el caso de que esto no ocurra, el **sesgo** se define como $Sesgo(\hat{\theta}) = E(\hat{\theta}) - \theta$. Si el sesgo es positivo, hay una tendencia sistemática a sobreestimar el parámetro.

Otra buena propiedad de un estimador es no dar resultados muy diferentes para las distintas posibles muestras. Esto significa que es bueno que la **varianza** del estimador, $Var(\hat{\theta})$, sea lo menor posible.

El **error cuadrático medio** tiene en cuenta tanto el sesgo como la varianza simultáneamente:

$$ECM(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] = \dots = Sesgo(\hat{\theta})^2 + Var(\hat{\theta})$$

Demostración:

$$\begin{aligned} ECM(\hat{\theta}) &\equiv E[(\hat{\theta} - \theta)^2] = E[(\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta)^2] = \\ &= E[(\hat{\theta} - E(\hat{\theta}))^2 + 2((\hat{\theta} - E(\hat{\theta}))(E(\hat{\theta}) - \theta)) + (E(\hat{\theta}) - \theta)^2] = \\ &= E[(\hat{\theta} - E(\hat{\theta}))^2] + 2E[(\hat{\theta} - E(\hat{\theta}))(E(\hat{\theta}) - \theta)] + E[(E(\hat{\theta}) - \theta)^2] = \\ &= E[(\hat{\theta} - E(\hat{\theta}))^2] + 2(E(\hat{\theta} - \theta)E(\hat{\theta} - E(\hat{\theta}))) + E[(E(\hat{\theta}) - \theta)^2] \end{aligned}$$

Como $E(\hat{\theta} - E(\hat{\theta})) = E(\hat{\theta}) - E(\hat{\theta}) = 0$, se simplifica la expresión anterior y queda:

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

Ejemplo: Varianza y cuasi-varianza con R

1. Iniciamos los parámetros:

```
lambda = 4
sigma = 1/lambda^2
n = 20
m = 1000
```

2. Generamos los datos:

```
set.seed(2345)
muestras = matrix(rexp(n*m, lambda), n)
```

3. Calculamos los estimadores:

```
cuasi = apply(muestras, 2, var)
varianza = cuasi * (n-1)/n
estimador=gl(2,m,labels=c('cuasi', 'var'))
df = data.frame(estimaciones=c(cuasi, varianza), estimador=estimador)
```

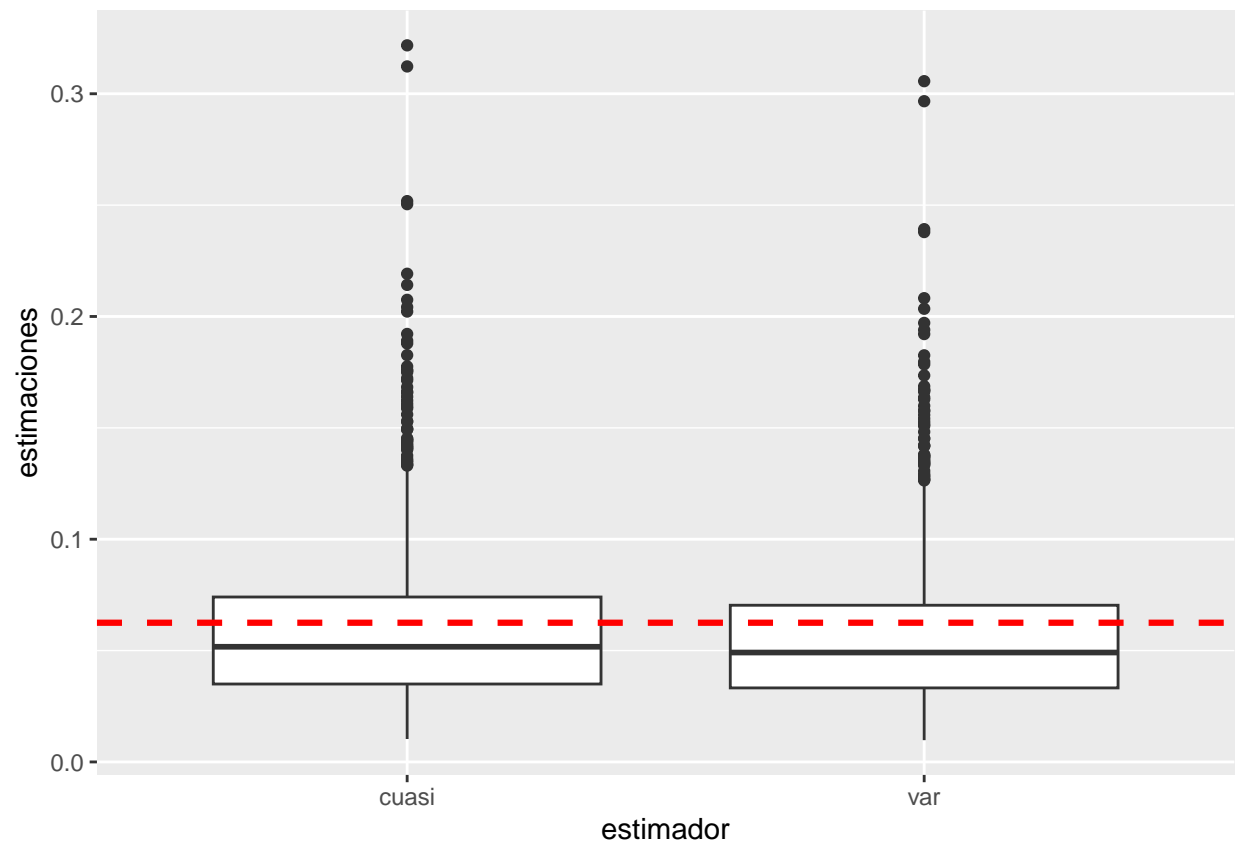
4. Gráficos

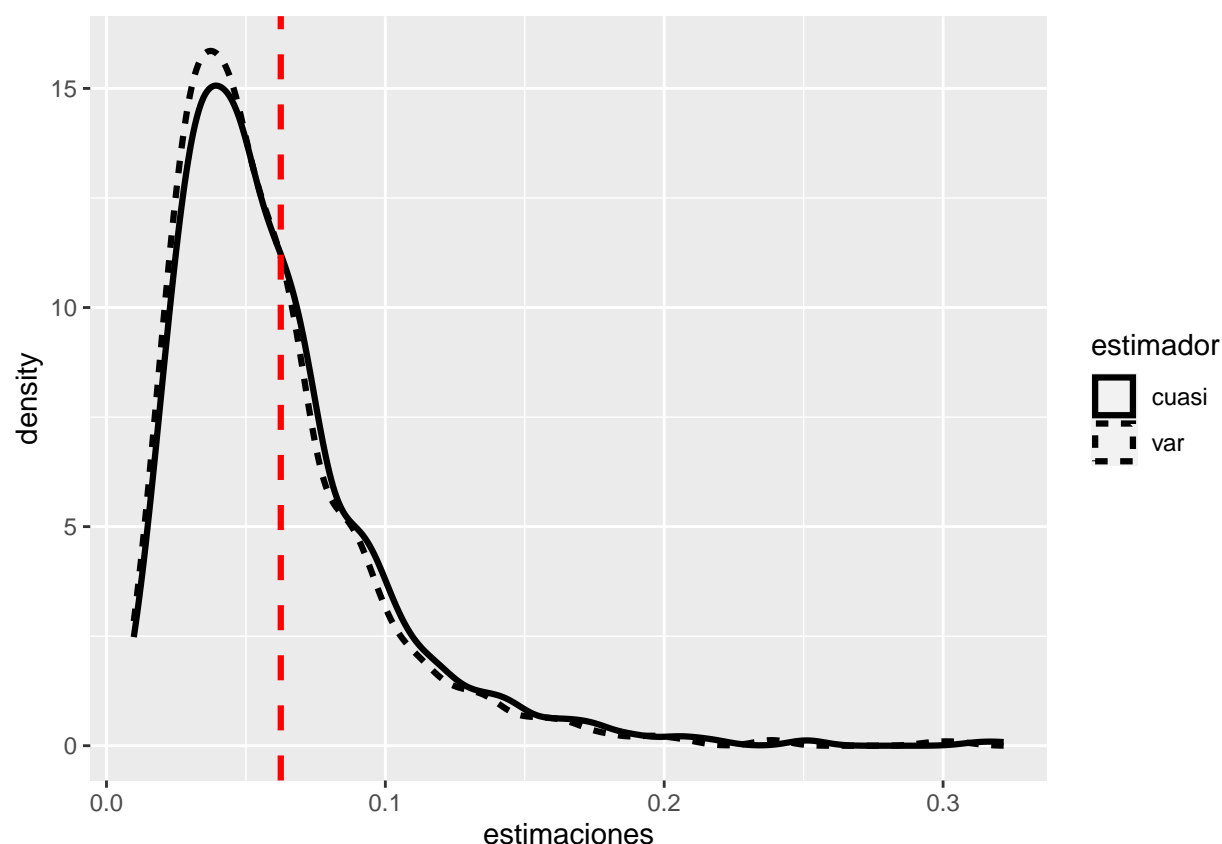
```
library(ggplot2)
cajas = ggplot(df) +
  geom_boxplot(aes(x=estimador, y=estimaciones)) +
  geom_hline(yintercept=sigma, col='red', size=1.1, linetype=2)
```

```
## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```

```
densidades = ggplot(df) +
  geom_density(aes(x=estimaciones, linetype=estimador), size=1.1) +
  geom_vline(xintercept = sigma, col='red', size=1.1, linetype=2)

cajas
```





Otros criterios: consistencia, normalidad asintótica,...

- Los criterios asintóticos para evaluar un estimador se refieren a su comportamiento límite a medida que disponemos de más y más datos.
- Se dice que un estimador es **consistente** si su valor converge al del parámetro al aumentar el tamaño muestral. Dato que $\hat{\theta}$ es una sucesión de v.a. debemos considerar algún tipo de convergencia estocástica.
- Otra buena propiedad que puede tener un estimador es la **normalidad asintótica**, es decir, la distribución límite del estimador sea aproximadamente normal para muestras grandes.
- Otros criterios: un estimador es **robusto** si no se ve muy afectado por la presencia de datos atípicos en la muestra.

Convergencia en probabilidad

Sea X_n una sucesión de variables aleatorias. Se dice que X_n **converge en probabilidad** a otra variable X y se denota $X_n \xrightarrow{P} X$ si, para todo $\epsilon > 0$ se cumple:

$$\lim_{n \rightarrow \infty} P\{|X_n - X| > \epsilon\} = 0$$

Dado un estimador $\hat{\theta}$, se dice que es debilmente **consistente** para θ si $\hat{\theta} \xrightarrow{P} \theta$.

Proposición. Si tanto el sesgo como la varianza de un estimador convergen a cero cuando $n \rightarrow \infty$, entonces el estimador es consistente.

Convergencia en distribución

Sea X_n una sucesión de variables aleatorias con funciones de distribución F_n . Se dice que X_n **converge en distribución** a otra variable aleatoria X con función de distribución F y se denota $X_n \xrightarrow{d} X$ si, para todo $x \in \text{Cont}(F)$, se verifica $\lim_{n \rightarrow \infty} F_n(x) = F(x)$, donde $\text{Cont}(F)$ es el conjunto de puntos en los que F es continua.

Relación entre las convergencias

$$X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{d} X$$

Si la distribución límite es degenerada,

$$X_n \xrightarrow{P} \theta \iff X_n \xrightarrow{d} \theta$$

Teorema de la aplicación continua. Sea X_n una sucesión de v.a. tal que $X_n \xrightarrow{P} X$ y sea $g : \mathbb{R} \rightarrow \mathbb{R}$ una función continua. Entonces, $g(X_n) \xrightarrow{P} g(X)$.

Ley débil de los grande números (LDGN)

LDGN. Sea X_n una sucesión de v.a.i.i.d. con media μ , entonces:

$$\bar{X}_n = \frac{X_1 + \dots + X_n}{n} \xrightarrow{P} \mu$$

La demostración cuando $\text{Var}(X_i) = \sigma^2 < \infty$ se reduce a una aplicación elemental de la desigualdad de Chebychev:

$$P(|\hat{X}_n - \mu| > \epsilon) \leq \frac{\text{Var}(\hat{X}_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \rightarrow 0$$

Teorema central del límite (TCL)

TCL. Sea X_n una sucesión de v.a.i.i.d. con media μ y varianza σ^2 , entonces

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0, 1)$$

La conclusión del TCL es equivalente a

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$$

El TCL da una aproximación de la distribución en el muestreo de la media:

$$\bar{X}_n \approx N\left(\mu, \frac{\sigma^2}{n}\right)$$

Método de momentos

Sean X_1, \dots, X_n v.a.i.i.d. con distribución F_θ , donde $\theta = (\theta_1, \dots, \theta_d)$ es un vector d -dimensional. Los momentos *poblacionales* son funciones de θ :

$$g_k(\theta) := E_\theta(X^k) = \int x^k dF_\theta(x), k = 1, \dots, d$$

Los momentos *muestrales* no dependen del parámetro:

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

Si θ es el valor verdadero, $m_k \approx g_k(\theta)$ y se estima θ mediante aquel valor que hace que los momentos poblacionales coincidan con los correspondientes momentos muestrales.

El estimador de momentos es el vector $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_d)$ que verifica

$$m_1 = g_1(\hat{\theta}), \dots, m_d = g_d(\hat{\theta})$$

Método de máxima verosimilitud

- Sea x_1, \dots, x_n una realización de una muestra X_1, \dots, X_n con función de densidad o de probabilidad conjunta $g(x_1, \dots, x_n; \theta)$ en (x_1, \dots, x_n) , donde $\theta \in \Theta \subset \mathbb{R}^d$
- La **función de verosimilitud** $L : \Theta \rightarrow \mathbb{R}$ se define como:

$$L(\theta) = L(\theta; x_1, \dots, x_n) = g(x_1, \dots, x_n; \theta)$$

- Si X_1, \dots, X_n son v.a.i.i.d. con densidad o probabilidad $f(x; \theta)$, entonces $L(\theta) = \prod_{i=1}^n f(x_i; \theta)$

Un **estimador de máxima verosimilitud** de θ es un valor $\hat{\theta} \in \Theta$ tal que $L(\hat{\theta}) = \sup_{\theta \in \Theta} L(\theta)$

- En vez de maximizar directamente $L(\theta)$ suele ser más conveniente maximizar $l(\theta : x) = l(\theta) = \log L(\theta) = \sum_{i=1}^n \log f(x_i, \theta)$

Ejemplo: distribución uniforme con R

1. Parámetros:

```
theta = 10
n = 20
m = 1000
```

2. Generación de datos:

```
set.seed(1234)
muestras <- matrix(runif(n * m, min = 0, max = theta), n)
```

3. Cálculo de estimadores

```

emv = apply(muestras, 2, max)
momentos = 2 * apply(muestras, 2, mean)
metodo = gl(2, m, labels = c('emv', 'momentos'))
df = data.frame(estimadores = c(emv, momentos), metodo=metodo)

```

4. Gráficas

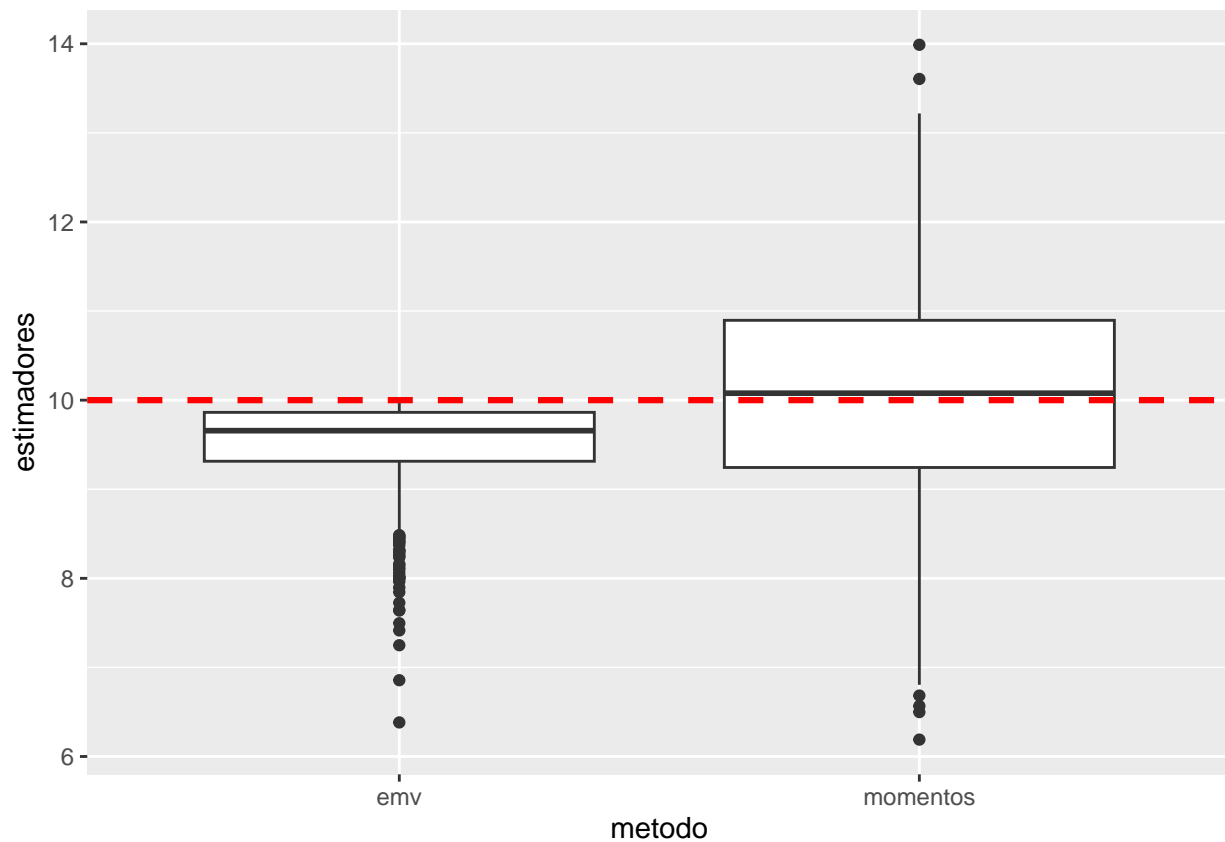
```

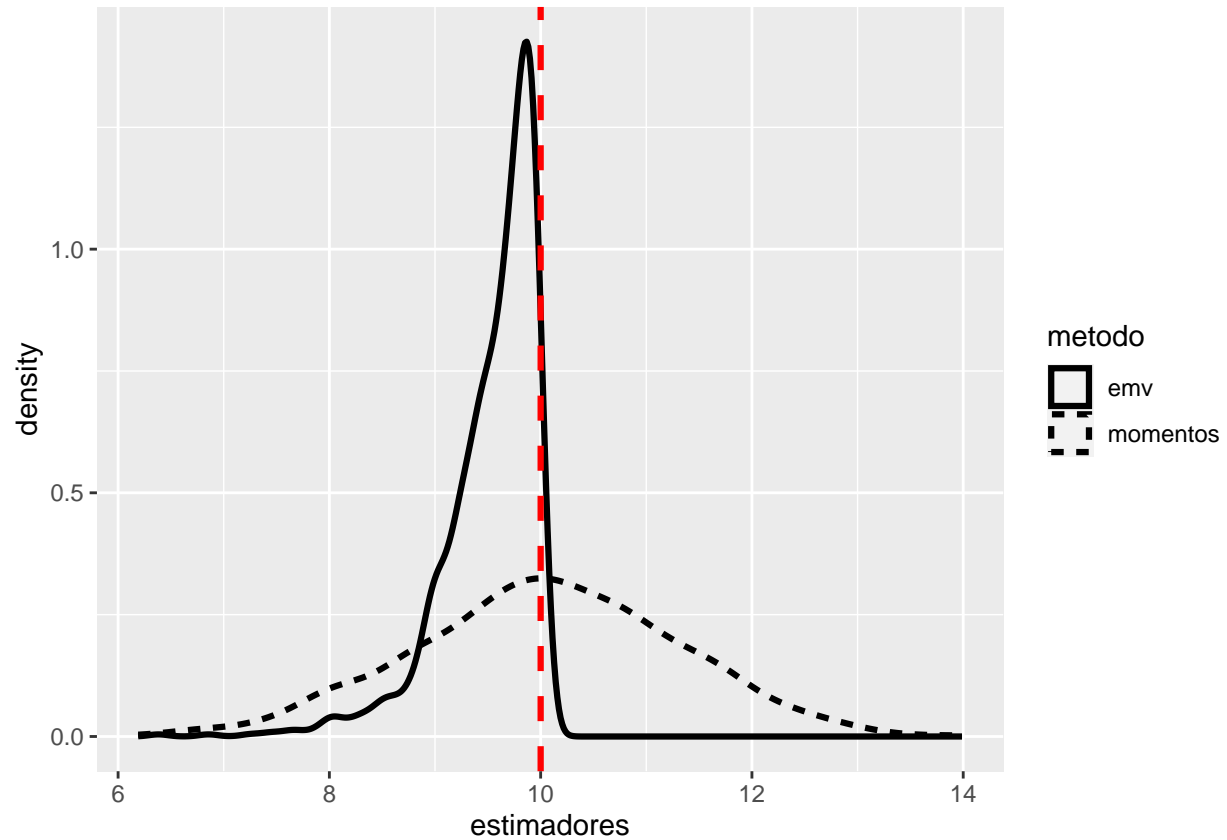
cajas = ggplot(df)+
  geom_boxplot(aes(x=metodo, y=estimadores))+
  geom_hline(yintercept = theta, col='red', size=1.1, linetype=2)

densidades = ggplot(df)+
  geom_density(aes(x=estimadores, linetype=metodo), size=1.1)+
  geom_vline(xintercept = theta, col='red', size=1.1, linetype=2)

cajas

```





Cálculo de EMV: el algoritmo EM

Queremos calcular el EMV a partir de unos datos $x = (x_1, \dots, x_n)$

Denotamos por y a otros datos (posiblemente inobservables) que proporcionan información adicional.

- Log-verosimilitud para x :

$$l(\theta; x)$$

- Log-verosimilitud completa para (x, y) :

$$l_c(\theta; x, y)$$

Aunque y sea en realidad inobservable, este algoritmo puede ser útil si maximizar l_c es más fácil que maximizar l .

Las situaciones típicas son las mixturas, censura y datos faltantes entre otros.

Algoritmo

Sea θ_k el valor del parámetro en el paso k .

Paso E (Expectation): Se sustituye $l_c(\theta; x, y)$ (no observada) por su valor esperado dada la información disponible, suponiendo que θ_k sea el verdadero valor del parámetro.

$$Q(\theta; \theta_k, x) := E_{Y|x; \theta_k}[l_c(\theta; x, Y)]$$

Paso M (Maximization): Se maximiza en θ la función $Q(\theta; \theta_k, x)$, lo que da lugar al siguiente valor del parámetro, θ_{k+1}

Bajo condiciones de regularidad poco exigentes, en cada paso del algoritmo aumenta la verosimilitud: $l(\theta_{k+1}; x) \geq l(\theta_k; x)$

Ejemplo

Las observaciones x_1, \dots, x_n proceden con probabilidad $1 - p$ de una distribución con densidad f_0 y con probabilidad p de otra distribución f_1 . Tanto f_0 como f_1 son conocidas y el objetivo es calcular el EMP de p .

Verosimilitud que hay que maximizar:

$$l(p; x) = \sum_{i=1}^n \log [(1 - p)f_0(x_i) + pf_1(x_i)]$$

Info adicional: la observación x_i viene de f_1 (que denotamos como $y_i = 1$) o viene de f_0 (que denotamos como $y_i = 0$). La verosimilitud completa es:

$$l(p; x, y) = \sum_{i=1}^n [(1 - y_i)\log(1 - p) + y_i\log p] + C$$

donde C no depende del parámetro p .

Paso E:

Por la fórmula de Bayes,

$$y_{i,k} := E(Y_i|x_i; p_k) = \frac{p_k f_1(x_i)}{(1 - p_k)f_0(x_i) + p_k f_1(x_i)}$$

Salvo sumandos que no dependen de p ,

$$Q(p; p_k) = \sum_{i=1}^n [(1 - y_{i,k})\log(1 - p) + y_{i,k}\log p]$$

Paso M:

Maximizamos $Q(p; p_k)$. Para ello derivamos e igualamos a cero y obtenemos

$$p_{k+1} = \frac{1}{n} \sum_{i=1}^n y_{i,k}$$

Ejemplo del algoritmo en R

1. Generación de datos de una $N(0, 1)$ o de una $N(4, 1)$:

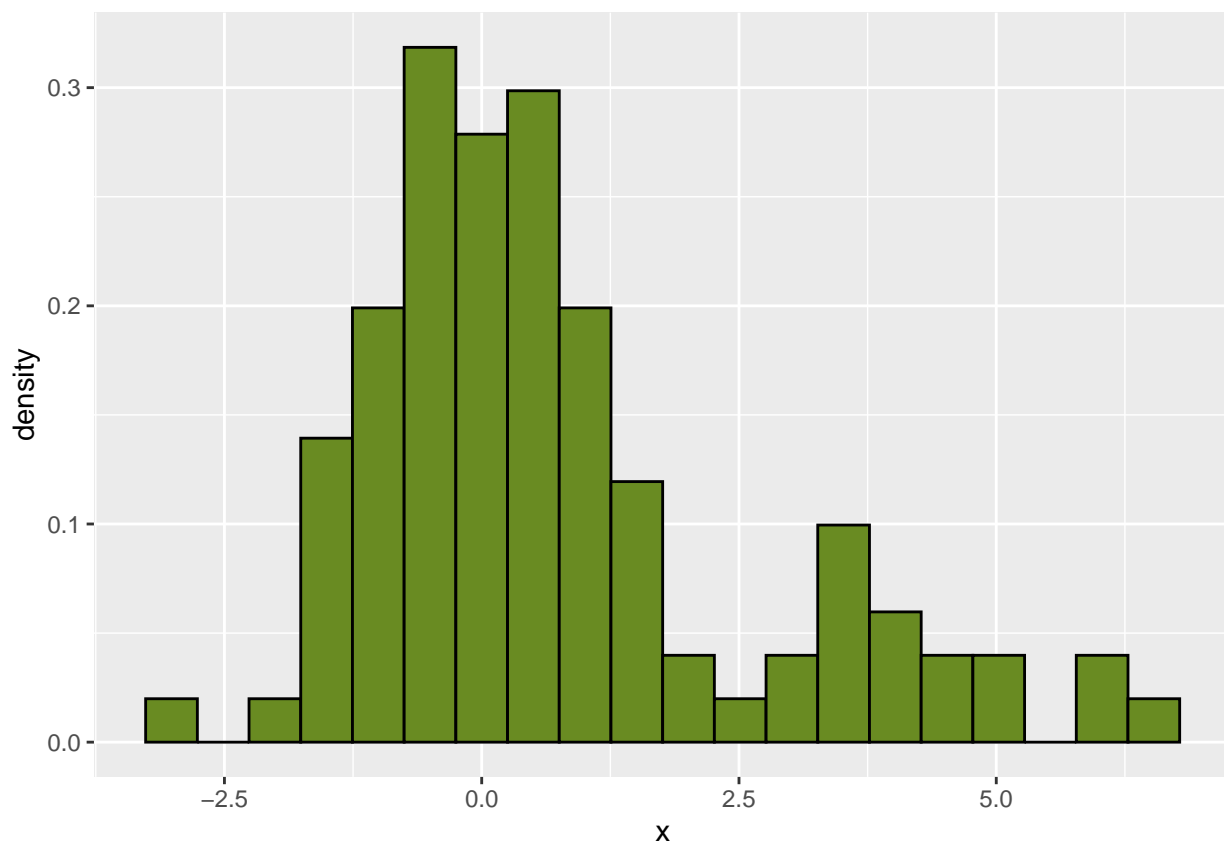
```
set.seed(156)
n=100
p=0.8
media=4

y=rbinom(n, 1, p)
x=rnorm(n)
n0 = sum(y==0)
x[y==0]=rnorm(n0, mean=media)
```

2. Representación del histograma de los datos

```
ggplot(data.frame(x))+
  geom_histogram(aes(x=x, y=..density..), col='black', fill='olivedrab4', bins=20)
```

```
## Warning: The dot-dot notation ('..density..') was deprecated in ggplot2 3.4.0.
## i Please use 'after_stat(density)' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```



3. Aplicación del algoritmo:

```
p_k=0.5
tol=0.1

while(tol > 0.001){
  p_0=p_k
  y_k=p_0*dnorm(x)/(p_0*dnorm(x) + (1-p_0)*dnorm(x, mean=media)) # Aplicación del paso E
  p_k=mean(y_k) # Aplicación del paso M
  tol=abs(p_k-p_0)
}

p_k

## [1] 0.8179144
```

Métodos bayesianos

Hay dos interpretaciones de la probabilidad de un suceso:

- **Frecuentista:** el límite de la frecuencia relativa de veces que ocurre el suceso cuando un experimento aleatorio se va repitiendo más y más veces.
- **Bayesiana:** grado de creencia subjetiva en que tal suceso ocurra.

Si se adopta un enfoque bayesiano, tiene sentido describir la incertidumbre sobre el parámetro mediante una distribución de probabilidad definida en el espacio paramétrico Θ .

El **método bayesiano** opera de la siguiente forma:

- Se establece una distribución a priori sobre Θ , $\pi(\theta)$, previa a la observación de la muestra que refleja la opinión de un experto sobre los valores del parámetro.
- La información sobre θ sostenida en $\pi(\theta)$ se combina con el modelo estadístico para los datos $f(x|\theta)$ mediante el teorema de Bayes para calcular la llamada **distribución a posteriori** $\pi(\theta|x)$:

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int_{\Theta} f(x|\theta)\pi(\theta)d\theta}$$

- Se toma como estimador de θ alguna medida numérica de posición que resuma la distribución a posteriori. Usualmente la media o la moda.

Ejemplo

El número de llegadas por hora a las emergencias de un hospital sigue una distribución de Poisson de parámetro λ . A priori se supone que el valor de λ toma alguno de los siguientes valores con las siguientes probabilidades:

Valores	3	3.5	4	4.5	5
Probabilidades	0.1	0.2	0.4	0.2	0.1

Table 1: Ejemplo de tabla en LaTeX

Se han registrado las visitas durante $n = 10$ periodos de una hora independientes y el número total de visitas ha sido 31. ¿Cuál es el estimador bayesiano de λ ?

Solución con R:

```
library(knitr)
lambda = c(3, 3.5, 4, 4.5, 5)
a_priori = c(0.1, 0.2, 0.4, 0.2, 0.1)
producto = a_priori*exp(-10*lambda)*lambda^31
a_posteriori = producto/sum(producto)

resultado = kable(round(cbind(lambda, a_priori, producto, a_posteriori), 3))
resultado
```

lambda	a_priori	producto	a_posteriori
3.0	0.1	5.780	0.241
3.5	0.2	9.265	0.386
4.0	0.4	7.837	0.327
4.5	0.2	1.017	0.042
5.0	0.1	0.090	0.004

- Como consecuencia,

$$\hat{\lambda} := \operatorname{argmax} \pi(\lambda|x_1, \dots, x_n) = 3.5$$

- No coinciden los valores más probables a priori y a posteriori.
- No es necesario calcular el denominador de la fórmula de Bayes para resolver el problema.

Métodos bayesianos y computación

- Las *familias conjugadas* para un modelo dado son familias paramétricas de distribuciones a priori tales que la distribución a posteriori pertenece a la misma familia paramétrica, cuando los datos siguen ese modelo.
- Si las distribuciones no son conjugadas, los problemas computacionales que representa el cálculo de la distribución a posteriori y su esperanza pueden ser muy difíciles, especialmente si θ es un vector de alta dimensión.
- Se han desarrollado métodos numéricos basados en simulación de cadenas de Markov (**Gibbs Sampling** y, más en general, **métodos MCMC**(Markov Chain Monte Carlo)) que permiten extender la aplicación de los métodos bayesianos a modelos muy complejos con un número grande de parámetros.

Ejemplo de Métodos bayesianos.

- Observamos $X \equiv B(n, p_1)$ e $Y \equiv B(m, p_2)$. Estamos interesados en estimar la diferencia de proporciones $\theta = p_1 - p_2$.
- Suponemos a priori que p_1 y p_2 son independientes en $[0, 1]$, es decir, $\pi(p_1, p_2) = 1$
- La distribución a posteriori verifica:

$$\pi(p_1, p_2 | X, Y) \propto p_1^X (1 - p_1)^{n-X} p_2^Y (1 - p_2)^{m-Y}$$

- Como consecuencia:
 - p_1 y p_2 son independientes a posteriori.
 - $p_1 | X, Y \equiv \text{Beta}(X + 1, n - X + 1)$
 - $p_2 | X, Y \equiv \text{Beta}(Y + 1, m - Y + 1)$

Para aproximar la media a posteriori de θ :

- Generamos $(p_1^{*(j)}, p_2^{*(j)})$, $j = 1, \dots, R$, con las distribuciones beta marginales anteriores.
- Para cada par, se calcula $\theta^{*(j)} = p_1^{*(j)} - p_2^{*(j)}$
- Los valores anteriores son realizaciones de la distribución a posteriori de θ .
- Por la ley de grandes números,

$$\hat{\theta} = E(\theta | X, Y) \approx \frac{1}{R} \sum_{j=1}^R \theta^{*(j)}$$

Un ejemplo con R:

```
set.seed(100)

# Datos

X=8
Y=6
n=10
m=10

# Generación de números aleatorios

R=1000 # Valores generados
p1 = rbeta(R, X+1, n-X+1)
p2 = rbeta(R, Y+1, m-Y+1)
theta = p1-p2

# Media a posteriori
mean(theta)

## [1] 0.1595184
```

Otra técnica de inferencia: contrastes

- Si $\Theta_0 \cup \Theta_1 = \Theta$, $\Theta_0 \cap \Theta_1 = \emptyset$, podemos expresar en general las hipótesis nula y alternativa como $H_0 : \theta \in \Theta_0$ y $H_1 : \theta \in \Theta_1$
- Al aceptar o rechazar H_0 podemos cometer dos tipos de errores:
 - **Error de tipo I:** Rechazar H_0 cuando es cierta.
 - **Error de tipo II:** Aceptar H_0 cuando es falsa.
- La función de potencia del contraste definido por la región crítica R se define como $\beta(\theta) = P_\theta(R)$.
- El **tamaño o nivel de significación de un contraste** es la máxima probabilidad de error de tipo I:

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

- El **p-valor** de un contraste para una muestra dada es el **ínfimo de los valores α para los cuales se rechaza la hipótesis nula a un nivel de significación α** con esa muestra.

Ejemplo de contraste

Tenemos una muestra X_1, \dots, X_n una v.a.i.i.d. con distribución $N(\theta, 1)$. Queremos contrastar $H_0 : \theta \leq 0$ frente a $H_1 : \theta > 0$. Consideramos la región crítica $R = \{\bar{X} > c\}$.

1. Calcular la función de potencia $\beta(\theta)$ como función de c y determina el valor que debe tener c para que el tamaño del contraste sea un valor prefijado $\alpha \in (0, 1)$.

Solución:

$$\begin{aligned}\bar{X} &\sim N(\theta, 1/n) \Rightarrow \sqrt{n}(\bar{X} - \theta) \sim N(0, 1) \\ \beta(\theta) &= P_\theta(R) = P(\bar{X} > c) = P\left(\frac{\sqrt{n}(\bar{X} - \theta)}{Z} > \sqrt{n}(c - \theta)\right) = \\ &= P(z > \sqrt{n}(c - \theta)) = 1 - P(z < \sqrt{n}(c - \theta)) = 1 - \Phi(\sqrt{n}(c - \theta))\end{aligned}$$

2. ¿Cuál es la mayor de las probabilidades de error de tipo II para el contraste anterior?

Solución:

$$\begin{aligned}\alpha &= \sup_{\theta \in \Theta_0} \beta(\theta), \quad \theta \in \Theta_0 \\ \alpha &= \beta(\theta) = 1 - \Phi(\sqrt{n}c) \\ \Phi(\sqrt{n}c) &= 1 - \alpha \iff Z_\alpha\end{aligned}$$

Como:

$$\Phi(\sqrt{n}c) = P(Z \leq \sqrt{n}c) \Rightarrow c = \frac{Z_\alpha}{\sqrt{n}}$$