

UNIVERSIDAD NACIONAL AGRARIA LA MOLINA

DEPARTAMENTO DE ESTADÍSTICA E INFORMÁTICA
FACULTAD DE ECONOMÍA Y PLANIFICACIÓN



TRABAJO INTEGRADOR

Inferencia Estadística 2025-II

Integrante	Código
Montúfar Paiva Yeraldi Mercedes	20230400
Kay Daniela L. Zavala Malpartida	20230420
Rojas Taco Fabiana	2023
Castillo Ruiz Mauricio Gabriel	20230384
Gómez Vigo Héctor Estefano	20230397

Docente: Fernando Miranda Villagómez

LA MOLINA - LIMA - PERÚ 2025

Distribución Pareto:

$$f(x) = \frac{\beta\alpha^\beta}{x^{\beta+1}} I_{(\alpha, \infty)}(x)$$

$$E(X) = \frac{\alpha\beta}{\beta-1} ; \quad \text{VAR}(X) = \left(\frac{\alpha}{\beta-1}\right)^2 \frac{\beta}{\beta-2}$$

1.1 EMV para parametro alfa:

Usaremos el Método de máxima verosimilitud para poder hallar el primer estimador para alfa:

$$L(\alpha, \beta) = \prod_{i=1}^n f(x_i; \alpha, \beta) = \prod_{i=1}^n \frac{\beta\alpha^\beta}{x_i^{\beta+1}} I_{[\alpha, \infty)}(x_i) = \beta^n \alpha^{n\beta} \left(\prod_{i=1}^n x_i^{-(\beta+1)} \right) \times \prod_{i=1}^n I_{[\alpha, \infty)}(x_i)$$

Analizando la indicadora:

$$\begin{aligned} \prod_{i=1}^n I_{[\alpha, \infty]}(x_i) &= I\left(\bigcap_{i=1}^n \{x_i \geq \alpha\}\right) \\ &= I(\alpha \leq x_1, \alpha \leq x_2, \dots, \alpha \leq x_n) \\ &= I(\alpha \leq \min\{x_1, \dots, x_n\}) = I(\alpha \leq y_1) \end{aligned}$$

donde $y_1 = X(1)$ es el mínimo de la muestra:

El EMV de $\theta = \alpha$ es $T = Y_1$.

1.1.2 Propiedades:

a) **Insesgamiento:** Distribución del mínimo:

$$\begin{aligned} X_1, \dots, X_n &\sim \text{Pareto } (\alpha, \beta) \text{ i.i.d} \\ F(x) &= 1 - \left(\frac{\alpha}{x}\right)^\beta, x \geq \alpha \\ y &= x_i \\ Fx_i(y) &= n[1 - F(y)]^{n-1} \mathcal{F}(y) \rightarrow n \left[\left(\frac{\alpha}{y}\right)^\beta \right]^{\beta-1} \cdot \frac{\beta\alpha^\beta}{y^{\beta+1}} \\ &= n \left[\frac{\alpha^{\beta(n-1)}}{\alpha^{\beta(n-1)}} \right] \cdot \frac{\beta\alpha^\beta}{y^{\beta+1}} = \frac{(n\beta)\alpha^{n\beta}}{y^{n\beta+1}}, y \geq \alpha \\ \rightarrow x_1 &\sim \text{Pareto}(\alpha, n\beta) \end{aligned}$$

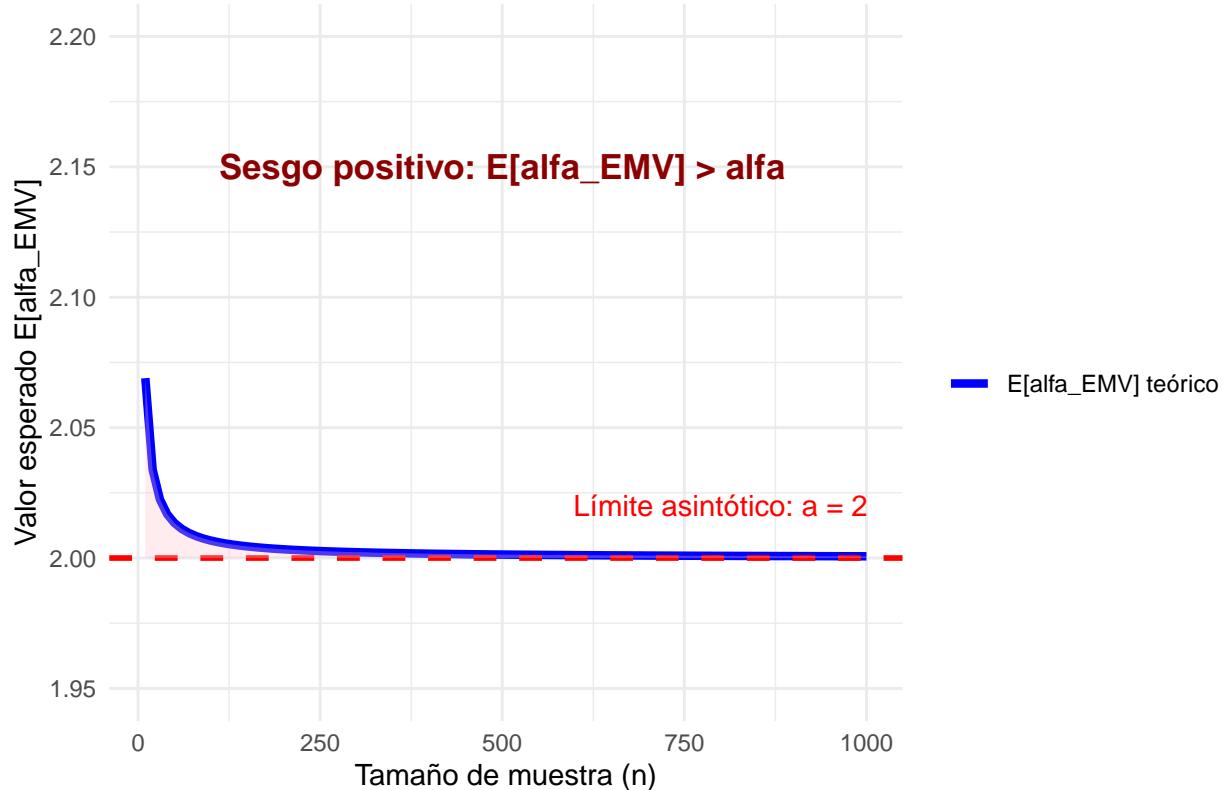
Hay sesgo positivo, es decir que sobreestima al parámetro, no cumple propiedades de Insesgamiento. Ahora lo veremos mejor con los valores para nuestra data:

```
# 1. Valor observado del mínimo
alpha_EMV_observado <- min(reclamos)
# 2. Valor esperado teórico del mínimo
E_alpha_EMV_teorico <- (n * beta_true * alpha_true) / (n * beta_true - 1)
```

```
## [1] "El valor observado es: 2.00031032898679 y el valor esperado es: 2.00066688896299"
```

El valor observado en los datos fue de 2.00031, mientras que el valor esperado teórico es 2.00067, ambos ligeramente por encima del valor real del parámetro ($\alpha = 2.00000$). Esto confirma que, para muestras finitas, el estimador tiende a sobreestimar el parámetro verdadero, aunque en este caso la magnitud del sesgo es muy pequeña (del orden de 0.0003 a 0.0007). La cercanía entre el valor observado y el teórico valida la expresión matemática del sesgo y respalda la propiedad de insesgamiento asintótico, ya que a medida que el tamaño de muestra aumenta, el sesgo tiende a cero.

Sesgo Positivo y Convergencia Asintótica de $E[\alpha_{EMV}]$



de nuestra base de datos, tomamos submuestras que van incrementando su valor para verificar que tendiendo al infinito (tamaños suficientemente grande) podemos ver como el valor tiende al parámetro de $\alpha = 2$.

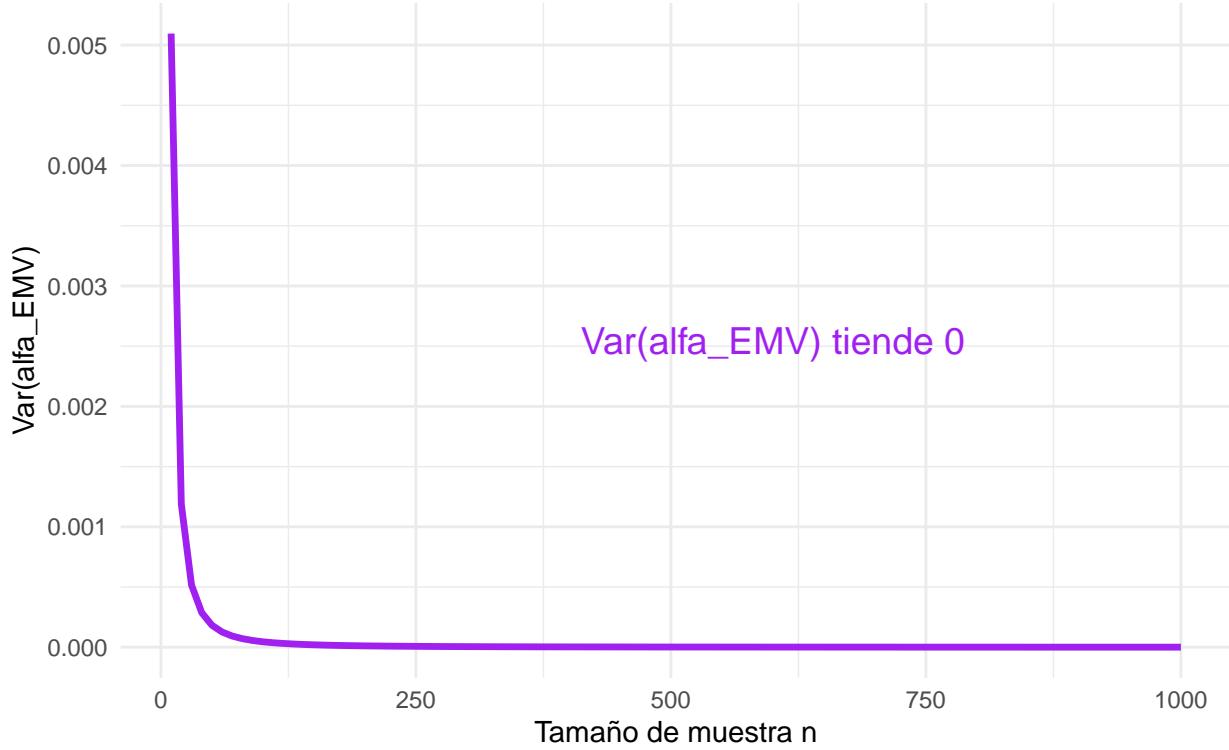
b) Consistencia: Podemos hacer uso del Teorema 2, por la propiedad anterior del insesgamiento asintótico

- $\lim_{n \rightarrow \infty} E(y_1) = \alpha$ ya que $\lim_{n \rightarrow \infty} \frac{n\beta \cdot \alpha}{n\beta - 1} = \lim_{n \rightarrow \infty} \frac{\beta\alpha}{\beta - \frac{1}{n}} = \frac{\beta\alpha}{\beta} = \alpha$
- $\lim_{n \rightarrow \infty} \text{Var}(y_1) = \lim_{n \rightarrow \infty} \frac{n\beta\alpha^2}{(n\beta - 1)^2(n\beta - 2)} = \lim_{n \rightarrow \infty} \frac{n\beta\alpha^2}{n^3\beta^3} = \lim_{n \rightarrow \infty} \frac{\alpha^2}{n^2\beta^2} = 0$

Como podemos observar cumple ambas condiciones, por ende decimos que nuestro estimador es consistente n otras palabras el estimador converge en probabilidades al parámetro.

** Gráfico:** Hemos visto ya un gráfico para solidar la primera demostración, ahora veamos uno que consolide la segunda condición:

Convergencia de la Varianza de alfa_EMV hacia Cero
Var(alfa_EMV) tiende 0 cuando n tiende inf



c) **Suficiencia:**

$$P(x = x/t = T) = \frac{P(x_1 = x_1 \dots x_n = x_n, T = y_1)}{P(T = y_1)}$$

$$\begin{aligned} & \frac{\left(\frac{\beta\alpha^\beta}{x_1^{\beta+1}}\right) \left(\frac{\beta\alpha^\beta}{x_2^{\beta+1}}\right) \cdots \left(\frac{\beta\alpha^\beta}{x_n^{\beta+1}}\right)}{\frac{n\beta\alpha^{n\beta}}{y^{n\beta+1}}}, \quad x_i \geq \alpha, \quad i = 1, 2, \dots, n \\ & = \frac{\beta^n \cdot \alpha^{n\beta} \cdot \prod_{i=1}^n x_i^{-(\beta+1)}}{\frac{n\beta\alpha^{n\beta}}{y^{n\beta+1}}} = \frac{\beta^{n-1} \cdot y^{n\beta+1} \cdot \prod_{i=1}^n x_i^{-(\beta+1)}}{n} \end{aligned}$$

Dada la dsistribución condicional dado $Y = X_1$, no depende de alfa, entonces podemos decir que es una estadística suficiente para alpha, cumple la propiedad.

d) **Ancilaridad:** La distribución del mínimo $x_{(1)}$ es:

$$\begin{aligned} x_{(1)} & \sim \text{Pareto}(\alpha, n\beta) \\ f_{x_{(1)}}(t) & = \frac{n\beta\alpha^{n\beta}}{t^{n\beta+1}}, \quad t \geq \alpha \end{aligned}$$

Como vemos su distribución depende de alfa, no cumple con la propiedad de ancilaridad.

e) **Completitud:**

INSERTAR DESARROLLO DEL CUADERNOP Por lo tanto $T = Y(1)$ es una estadística completa para alfa cuando tomamos a beta como conocido.

Por el Teorema de Bahadur, como es suficiente y completo, podemos afirmar que nuestro estimador es minimal suficiente.

f) **Optimalidad:** En nuestro caso alfa_EMV no es óptimo porque es sesgado. Pero si existe un estimador óptimo basado en la estadística suficiente que se podría obtener mediante corrección del sesgo. (Teorema de Lehman-Scheffé)

1.2 EMV para parametro Beta:

Para hallar el estimador debemos remplazar nuestro primer estimador para el parametro alfa:

$$\hat{\alpha}_{MLE} = \min\{X_1, X_2, \dots, X_n\}$$

Como sabemos el máximo en $L(\alpha, \beta)$ es igual al máximo de $\ln(L(\alpha, \beta))$ entonces haremos uso del ln:

$$\begin{aligned} L(\alpha, \beta) &= \beta^n \cdot \alpha^{n\beta} \cdot \prod_{i=1}^n X_i^{-(\beta+1)} \\ \ln(L(\alpha, \beta)) &= n \cdot \ln(\beta) + n\beta \ln(\alpha) - (\beta + 1) \sum_{i=1}^n \ln(X_i) \\ \frac{d \ln(L(\alpha, \beta))}{d\beta} &= \frac{n}{\beta} + n \ln \alpha - \sum_{i=1}^n \ln X_i \\ &= \frac{n}{\beta} + n \ln \alpha - \sum_{i=1}^n \ln X_i = 0 \\ \Rightarrow \frac{n}{\beta} &= \sum_{i=1}^n \ln X_i - n \ln \alpha \\ \Rightarrow \frac{n}{\beta} &= \sum_{i=1}^n \ln \left(\frac{X_i}{\alpha} \right) \\ \Rightarrow \hat{\beta} &= \frac{n}{\sum_{i=1}^n \ln \left(\frac{X_i}{\alpha} \right)} \end{aligned}$$

*1.1. Propiedades:**

a) **Insesgamiento** Primero analizaremos una parte del denominador para determinar su distribución y se facilite el procedimiento.

$$\begin{aligned} F_Y(y) &= P(Y_i \leq y) = P \left(\ln \left(\frac{X_i}{\alpha} \right) \leq y \right) = P(X_i \leq \alpha e^y) \\ &= 1 - \left(\frac{\alpha}{\alpha e^y} \right)^\beta = 1 - e^{-\beta y}, \quad y \geq 0 \\ f_Y(y) &= \frac{d}{dy} F_Y(y) = \beta e^{-\beta y}, \quad y \geq 0 \end{aligned}$$

Analizando el resultado de la distribución, notamos que tiene la forma de la exponencial. Entonces el mínimo igual será exponencial con parametro n por beta. Ademas tomamos a T como la sumatoria, donde por propiedad de ln de una división podemos desplegarlo.

$$\begin{aligned}
Y_i &\sim \text{Exponencial}(\beta) \\
W = Y_{(1)} &= \min(Y_1, \dots, Y_n) \sim \text{Exp}(n\beta) \\
T &= \sum_{i=1}^n (Y_i - Y_{(1)}) \sim \text{Gamma}(n-1, \beta) \\
\hat{\beta} &= \frac{n}{T}, \quad T \sim \text{Gamma}(n-1, \beta) \\
E\left[\frac{1}{T}\right] &= \frac{\beta}{n-2}, \quad n > 2 \\
E[\hat{\beta}] &= n \cdot E\left[\frac{1}{T}\right] = \frac{n\beta}{n-2}
\end{aligned}$$

Acá vemos que hay un sesgo presente

$$\text{Sesgo} = E[\hat{\beta}] - \beta = \frac{n\beta}{n-2} - \beta = \frac{2\beta}{n-2}$$

Por ende, no es un estimador insesgado. Ademas al ser positivo podemos decir que sobreestima al parametro.