

# XS3310 Teoría Estadística

I Semestre 2021

Escuela de Estadística

2021-05-03

## ¿Qué hemos visto hasta ahora?

Todo sobre estimadores puntuales.

## ¿Qué vamos a discutir hoy?

Intervalos de confianza clásicos

---

class: center, middle

## ¿Qué es un intervalo de confianza?

---

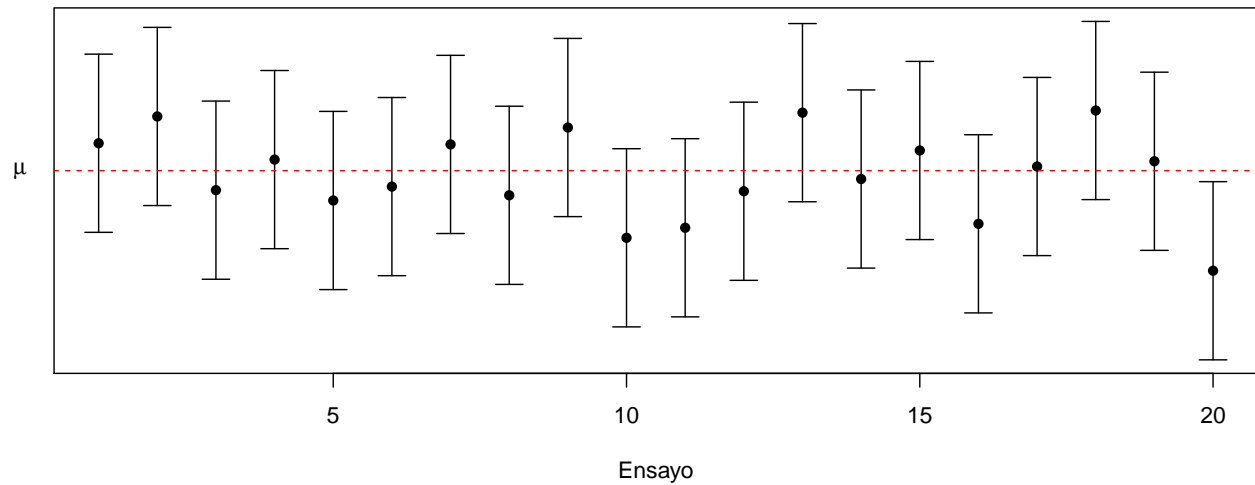
## Estimación por Intervalo

- Una estimación por intervalo para un parámetro desconocido  $\theta$  viene dada por un rango o intervalo que posee cierta probabilidad  $1 - \alpha$  de contener a dicho parámetro.
- Usualmente este concepto de *confianza* se interpreta diciendo que si obtenemos 100 intervalos entonces  $\alpha\%$  de ellos no van a incluir el valor del parámetro. Es decir, si tenemos 100 intervalos del 95% de confianza entonces se espera que 5 de ellos no incluyan el valor del parámetro, o que de 20 intervalos haya uno que no lo incluya.

Visualmente: <https://seeing-theory.brown.edu/frequentist-inference/es.html>

---

## Estimación por Intervalo



## Estimación por Intervalo

Existen dos tipos de estimaciones por intervalos:

- Intervalos bilaterales
- Intervalos unilaterales

En el primer caso se tienen dos valores  $a$  y  $b$  que constituyen los límites inferior y superior del intervalo. Es decir, el intervalo  $[a, b]$  contiene el valor de  $\theta$  con una confianza del  $(1 - \alpha) \cdot 100\%$ .

En el segundo caso, se determina solamente un límite que puede ser inferior ( $a$ ) o superior ( $b$ ), dependiendo del interés del estudio. Se tiene así intervalos de la forma  $[a, +\infty[$  o  $]-\infty, b]$ . Los valores  $\pm\infty$  pueden ser sustituidos por el mínimo o máximo valor del dominio del parámetro de interés.

---

## Técnica para estimaciones por intervalo

Existen varias técnicas para obtener estimaciones por intervalo, sin embargo en este curso nos enfocaremos en la **técnica del pivote**.

En la versión en línea verán un video [acá](#)

---

## Técnica del pivote

Esta técnica consiste en definir una variable aleatoria,  $U$ , que llamaremos **pivote**. El pivote debe cumplir con las siguientes condiciones:

- Para una muestra aleatoria  $X_1, X_2, \dots, X_n$  y un parámetro desconocido  $\theta$ ,  $U$  debe estar definido en términos de los elementos de la muestra aleatoria y del parámetro desconocido, donde el parámetro desconocido  $\theta$  sea la única cantidad desconocida.
- La distribución de probabilidad de  $U$  debe ser conocido y no depender de  $\theta$ .

**Ejemplo:** Sea  $X_1, X_2, \dots, X_n$  una muestra aleatoria tal que  $X_j \sim N(\mu, \sigma^2)$ , donde  $\mu$  es desconocido y  $\sigma^2$  es conocido. Obtenga una variable aleatoria que funcione como un pivote.

**Solución:** Debemos encontrar una variable aleatoria que esté definida en términos de la muestra y de  $\mu$ .

- Una forma de resumir toda la información de la muestra en una sola variable aleatoria. Lo ideal sería utilizar un estadístico suficiente:  $\sum_{j=1}^n X_j$  es suficiente para estimar  $\mu$ , por lo que podríamos usar alguna función de este estadístico en el pivote.
- $\bar{X}$  resume la información de la muestra aleatoria, pues está en términos de un estadístico suficiente y además sabemos que se distribuye  $N(\mu, \frac{\sigma^2}{n})$ . Ahora debemos encontrar alguna transformación de  $\bar{X}$  que incluya a  $\mu$  y cuya distribución no dependa de  $\mu$ .

- $U = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ , se distribuye  $N(0, 1)$ .  $U$  cumple con las propiedades de un pivote pues está definida en términos de la muestra y del parámetro desconocido  $\mu$  al igual que su distribución ya no depende de  $\mu$ .

**NOTA:**  $U$  no es la única opción de pivote. Por ejemplo,  $\bar{X} - \mu$  es otra variable aleatoria que también cumple las condiciones para ser un pivote. Existe una infinidad de variables aleatorias que cumplen esta condición, pero como veremos más adelante algunas son más útiles que otras.

Ahora debemos seguir con el procedimiento para encontrar un intervalo de confianza por medio de la técnica del pivote. Hay dos caminos:

- Caso bilateral
- Caso unilateral

- **Caso bilateral:** se requiere encontrar valores  $a$  y  $b$  tales que  $P(a \leq U \leq b) = 1 - \alpha$  y por lo tanto también se cumple que  $P(U < a) + P(U > b) = \alpha$ . Sin embargo, estas condiciones generan un número infinito de posibilidades para  $a$  y  $b$  por lo que se tiene que limitar de alguna manera. Para ello, se acostumbra utilizar la condición de cada una de las probabilidades de las “colas” sean iguales, es decir  $P(U < a) = P(U > b) = \frac{\alpha}{2}$ .
- **Caso unilateral:** podemos tener un límite inferior o superior. Dependiendo del que nos interese y de la forma del pivote debemos encontrar un valor  $a$  o  $b$  que cumpla  $P(U \geq a) = 1 - \alpha$  o  $P(U \leq b) = 1 - \alpha$ . Finalmente, como  $U$  está en términos de la muestra aleatoria y del parámetro  $\theta$  entonces despejamos  $\theta$  dentro de la probabilidad y con esto obtenemos los límites inferiores y superiores.

**Ejemplo:** Sea  $Y$  una variable aleatoria con distribución Exponencial con media  $\beta$  desconocida. Utilice este valor para construir un intervalo bilateral para  $\beta$  con una confianza del 90%.

**Solución:** En este caso podemos utilizar como pivote  $U = \frac{Y}{\beta}$  ya que está en términos de la variable aleatoria y de  $\beta$  al igual que se puede demostrar que su distribución no depende de  $\beta$ . En este caso  $U \sim \text{Exp}(1)$ .

Nuestro siguiente paso es buscar los valores  $a$  y  $b$  tales que:

$$P(a \leq U \leq b) = 0.9 \quad y \quad P(U < a) = P(U > b) = 0.05$$

Recordemos que en el caso de una variable aleatoria continua  $P(U < a) = P(U \leq a) = F_U(a)$ . Podemos usar nuestro conocimiento de la función de distribución de la exponencial para encontrar el valor de  $a$ :

$$P(U < a) = F_U(a) = 1 - e^{-a} = 0.05, \quad \Rightarrow a = -\ln(0.95)$$

Podemos hacer lo mismo para encontrar el valor de  $b$ :

$$P(U > b) = 1 - P(U \leq b) = 1 - F_U(b) = e^{-b} = 0.05, \quad \Rightarrow b = -\ln(0.05).$$

Nuestro último paso ahora es despejar el valor de  $\beta$  de la probabilidad que acumula 0.90:

$$\begin{aligned} P(a \leq U \leq b) &= P\left(-\ln(0.95) \leq \frac{Y}{\beta} \leq -\ln(0.05)\right) \\ &= P\left(\frac{-1}{\ln(0.95)} \geq \frac{\beta}{Y} \geq \frac{-1}{\ln(0.05)}\right) = P\left(\frac{-Y}{\ln(0.95)} \geq \beta \geq \frac{-Y}{\ln(0.05)}\right) \\ &= 0.90 \end{aligned}$$

Por lo tanto tenemos que  $a = \frac{-Y}{\ln(0.05)}$  y  $b = \frac{-Y}{\ln(0.95)}$  por lo que podemos concluir que con una confianza del 90% el intervalo  $\left[\frac{-Y}{\ln(0.05)}, \frac{-Y}{\ln(0.95)}\right]$  incluye el valor de  $\beta$ .

**Ejemplo:** Sea  $X$  una variable aleatoria Uniformemente distribuida en el intervalo  $(0, \theta)$ , con  $\theta$  desconocido. Determine un intervalo de confianza unilateral inferior para  $\theta$  del 95%.

**Solución:** Primero debemos encontrar una cantidad pivote satisfactoria. En este caso podemos utilizar  $U = \frac{X}{\theta}$  como pivote ya que cumple ambas condiciones. Se puede demostrar que  $U \sim \text{Unif}(0, 1)$  por medio de la segunda técnica de transformaciones:

$$\text{Sea } U = h(X) = \frac{X}{\theta}, \text{ por lo tanto } h^{-1}(U) = U\theta \quad \Rightarrow \quad \frac{dh^{-1}(u)}{du} = \theta$$

Aplicando la técnica tenemos:

$$f_U(u) = f_X(u\theta) \cdot \theta = \frac{1}{\theta} \cdot \theta = 1$$

Al ser el dominio original  $0 < x < \theta$  al sustituir  $x$  por  $u\theta$  obtenemos el nuevo dominio:  $0 < u < 1$ . Por lo tanto concluimos que  $U \sim \text{Unif}(0, 1)$ .

Ahora podemos usar esto para encontrar el valor de  $b$  tal que  $P(U \leq b) = 0.95$ :

$$P(U \leq b) = F_U(b) = 0.95, \Rightarrow b = 0.95$$

Con este valor podemos proceder a despejar el valor de  $\theta$  de esta probabilidad y obtener así el límite inferior de nuestro intervalo:

$$P(U \leq b) = P\left(\frac{X}{\theta} \leq 0.95\right) = P\left(\theta \geq \frac{X}{0.95}\right) = 0.95$$

Por lo tanto podemos concluir que con un 95% de confianza el intervalo  $\left[\frac{X}{0.95}, +\infty\right)$  incluye el valor de  $\theta$ .

**Ejemplo:** Sea  $X_1, X_2, \dots, X_{20}$  una muestra aleatoria tal que  $X_j \sim \text{Exp}(\beta)$ . Determine un intervalo de confianza bilateral para  $\beta$  con una confianza del  $(1 - \alpha)\%$ . Utilice como pivote una función de  $\sum_{j=1}^{20} X_j$ .

**Solución:** Sabemos de antemano que  $T = \sum_{j=1}^{20} X_j \sim \text{Gamma}(\alpha = 20, \beta)$ .

$T$  no sirve como pivote pues no depende de  $\beta$  y su distribución todavía depende de  $\beta$ . Debemos encontrar alguna transformación de  $T$  que haga que se cumplan las dos condiciones.

**Primera opción:**

Recordemos que la función generadora de momentos de una Gamma  $(\alpha, \beta)$  es  $(1 - \beta t)^{-\alpha}$ . Por lo tanto, si  $U = \frac{T}{\beta}$  entonces,

$$m_U(t) = m_T\left(\frac{t}{\beta}\right) = (1 - t)^{-20}$$

Esto quiere decir que  $U \sim \text{Gamma}(20, 1)$  y por lo tanto se puede emplear como un pivote pues cumple las dos condiciones. Ahora procedemos a encontrar los valores  $a$  y  $b$  tales que  $P(U < a) = P(U > b) = \frac{\alpha}{2}$ . Como la Gamma no tiene una función de distribución en forma cerrada debemos proceder a integrar la función de densidad:

$$P(U < a) = \int_0^a \frac{u^{19} e^{-u}}{\Gamma(20)} du = \frac{\alpha}{2}$$

Esta integral es posible de calcular a mano, pero su desarrollo necesita de 19 veces de “por partes” que hacen su cálculo extremadamente largo y tedioso. Sin embargo, podemos hacer uso de tablas estadísticas que nos dan cuantiles para ciertas probabilidades. En este caso no tenemos una tabla para una Gamma con estos parámetros, pero sí tenemos para otra Gamma: la  $\chi^2$ -cuadrado.

**Segunda opción:**

Afortunadamente existe una forma para pasar de cualquier Gamma a una  $\chi^2$ :

Sea  $V = \frac{2T}{\beta}$  entonces,

$$m_V(t) = m_T\left(\frac{2t}{\beta}\right) = (1 - 2t)^{-20} = (1 - 2t)^{-\frac{40}{2}}$$

Esta es la función generadora de momentos de una  $\chi^2$ -cuadrado con 40 grados de libertad, de la cuál sí tenemos tablas para encontrar cuantiles. Ahora si podemos proceder a encontrar los valores de  $a$  y  $b$ , sabiendo que gráficamente lucen de la siguiente manera:

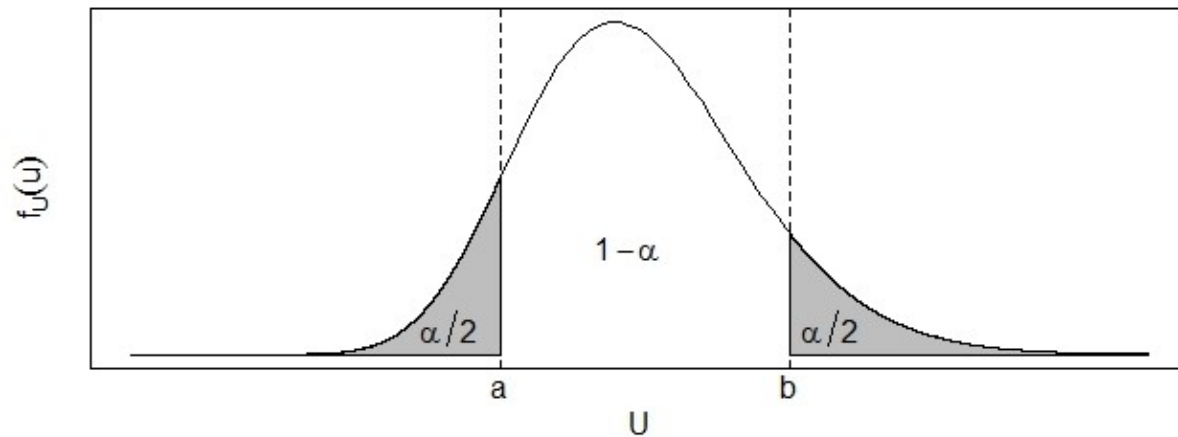


Figure 1: Cuantiles de una  $\chi^2$ -cuadrado con 40gl

---

Como la tabla de la  $\chi^2$ -cuadrado acumula hacia la derecha (a más de), esto significa que  $a$  es el valor de la tabla de una  $\chi^2$ -cuadrado con 40gl que acumula a su derecha una probabilidad de  $1 - \frac{\alpha}{2}$ . Por consiguiente  $b$  es el valor de la tabla que acumula a su derecha una probabilidad de  $\frac{\alpha}{2}$ . Para simplicidad vamos a llamar estos valores  $\chi^2_{1-\frac{\alpha}{2},40}$  y  $\chi^2_{\frac{\alpha}{2},40}$ , respectivamente.

Ya encontrados los valores de  $a$  y  $b$  podemos proceder a despejar  $\beta$  de la probabilidad que acumula  $1 - \alpha$ :

$$P(a \leq V \leq b) = P\left(\chi^2_{1-\frac{\alpha}{2},40} \leq \frac{2T}{\beta} \leq \chi^2_{\frac{\alpha}{2},40}\right) = P\left(\frac{2T}{\chi^2_{\frac{\alpha}{2},40}} \leq \beta \leq \frac{2T}{\chi^2_{1-\frac{\alpha}{2},40}}\right) = 1 - \alpha$$

Por lo tanto, con una confianza de  $(1 - \alpha)\%$  el intervalo  $\left[\frac{2\sum_{j=1}^{20} X_j}{\chi^2_{\frac{\alpha}{2},40}}, \frac{2\sum_{j=1}^{20} X_j}{\chi^2_{1-\frac{\alpha}{2},40}}\right]$  incluye el valor de  $\beta$ .

---

class: center, middle

## ¿Qué discutimos hoy?

Estimación por intervalos, método del pivote