Repaso

Estimación en una población infinita

Población infinita: un ejemplo

Una máquina fabrica barras de acero.

En tanto, a un estadístico le piden estimar la media de las piezas producidas por la máquina.

Las unidades no tienen “identificador”, son de alguna manera intercambiables

Carácter a estudiar aleatorio

Volvamos a la población infinita de las barras de acero

Objetivo marcado por el usuario: ¿Cuál es la media de la longitud de las piezas producidas por la máquina?

¿De qué herramientas estadísticas disponemos para dar una respuesta al usuario?

Precisemos el objetivo en términos estadísticos

Estimar la media μ de la variable aleatoria X« longitud de las piezas producidas por la máquina »

Nota: Una **variable aleatoria** (v.a.) X es una función real definida en el espacio muestral asociado a un experimento aleatorio Ω.

La funciones de distribución/densidad F_X y f_X indican la probabilidad de tomar un valor en un intervalo.

En el caso de las barras, X es una variable aleatoria de ley o función de probabilidad $\mathcal{L}$ (de forma normal, exponencial, uniforme, etc., no se sabe y no importa), esperanza μ, que es el parámetro poblacional que se debe estimar.

Se modeliza la longitud de las barras por la variable X

Observar una muestra de barras, es decir, una muestra de realizaciones de la variable X. Extraer una muestra MAS de tamaño *n*= seleccionar las barras, de tal manera que sean las observaciones independientes e idénticamente distribuídas (IID=MAS)

- medir su longitud= trabajo de campo

- Muestra, $(X_{1,}X_2,\ldots,X_n)$ conjunto de variables iid de distribución F_X

- Muestra observada= conjunto de valores observados $(x_{1,}x_2,\ldots,x_n)$

- Un estimador del parámetro θ es una función de la muestra,

$$\hat{\theta} = \left(X_1,\ldots,X_i,\ldots,X_n\right)$$

cuyo valor es función de las observaciones

$$\hat{\theta} = \hat{\theta}\left(x_1, x_2, \ldots, x_n\right)$$

- Sería bueno que $\theta \approx \hat{\theta}$ ….Precisemos …

Ahora, nos interesamos al parámetro media poblacional= esperanza de **X**

$$\theta = \mu$$

Se escoge el siguiente estimador:

$$\hat{\theta} = \overline{\mathrm{X}}_n = \frac{\sum_{i=1}^{n} \mathrm{X}_i}{n}$$

Media muestal

Sobre la muestra, toma el valor :

$$\hat{\theta} = \overline{x}_n = \frac{\sum_{i=1}^{n} x_i}{n}$$

Qué podemos decir a partir de lo estudiado en cálculo de probabilidades?

Qué herramientas tenemos?

- $\overline{\mathrm{X}}_n$, estimador de μ es una variable aleatoria, como todos los estimadores
- Las leyes de los grandes números muestran que la media muestral de una muestra al azar de gran tamaño de una población tenderá a estar cerca de la media de la población completa

  Por qué nos importa aquí?
  $\bar{x}_n$, valor puntual del estimador escogido, no estará "demasiado alejado" de μ

- Cuando la variable muestreada tiene varianza finita, el teorema central del límite dice que la distribución de la media muestral converge a la distribución de una variable aleatoria normal.

  Por qué nos importa aquí?
  Nos permite dar una respuesta al usuario del tipo $\left[\bar{x}_n - d \, , \, \bar{x}_n + d\right]$

- $\overline{X}_n$, estimador de μ es una variable aleatoria, como todos los estimadores, cuya esperanza es

$$E\left(\overline{X}_n\right)=E\left(\frac{\sum_{i=1}^{n} X_i}{n}\right)=\sum_{i=1}^{n} E\left(\frac{X_i}{n}\right)=\frac{1}{n}\sum_{i=1}^{n} E\left(X_i\right)$$

$$=\frac{1}{n}\sum_{i=1}^{n}\mu=\mu$$

Estimador <u>sin sesgo</u>

$\boxed{V\left(\overline{X}_n\right)}$

por la independencia de los X_i

$$V\left(\overline{X}_n\right) = V\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \sum_{i=1}^{n} V\left(\frac{X_i}{n}\right) = \frac{1}{n^2}\sum_{i=1}^{n} V\left(X_i\right)$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2 = \frac{\sigma^2}{n}$$

La varianza de la media muestral disminuye con el tamaño de la muestra

Por qué nos importa la varianza?

Volvamos a nuestro problema: calcular un intervalo de confianza para la media de una variable de ley cualquiera…En qué nos ayuda el TLC?

$X \sim Ley(\mu,\sigma)$.

Se desconoce μ. Pero, se supone σ conocido

TLC:

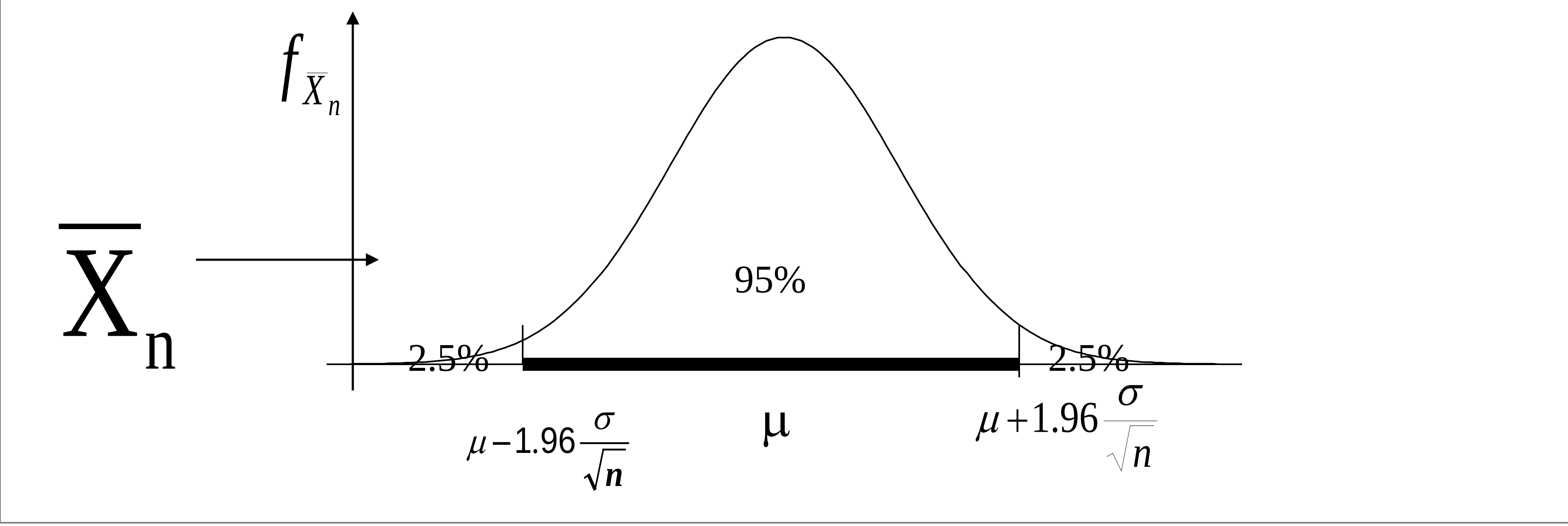
Si n > 30

$$\overline{\mathrm{X}}_{\mathrm{n}} \approx Normal$$

$$E\left(\overline{X}_n\right) = \mu$$

$$V\left(\overline{X}_n\right) = \frac{\sigma^2}{n}$$

Vemos cómo la normalidad de la media muestral permite calcular *d*.



Con una probabilidad del 95%, el valor observado de la media muestral estará en el intervalo indicado

Se desconoce μ

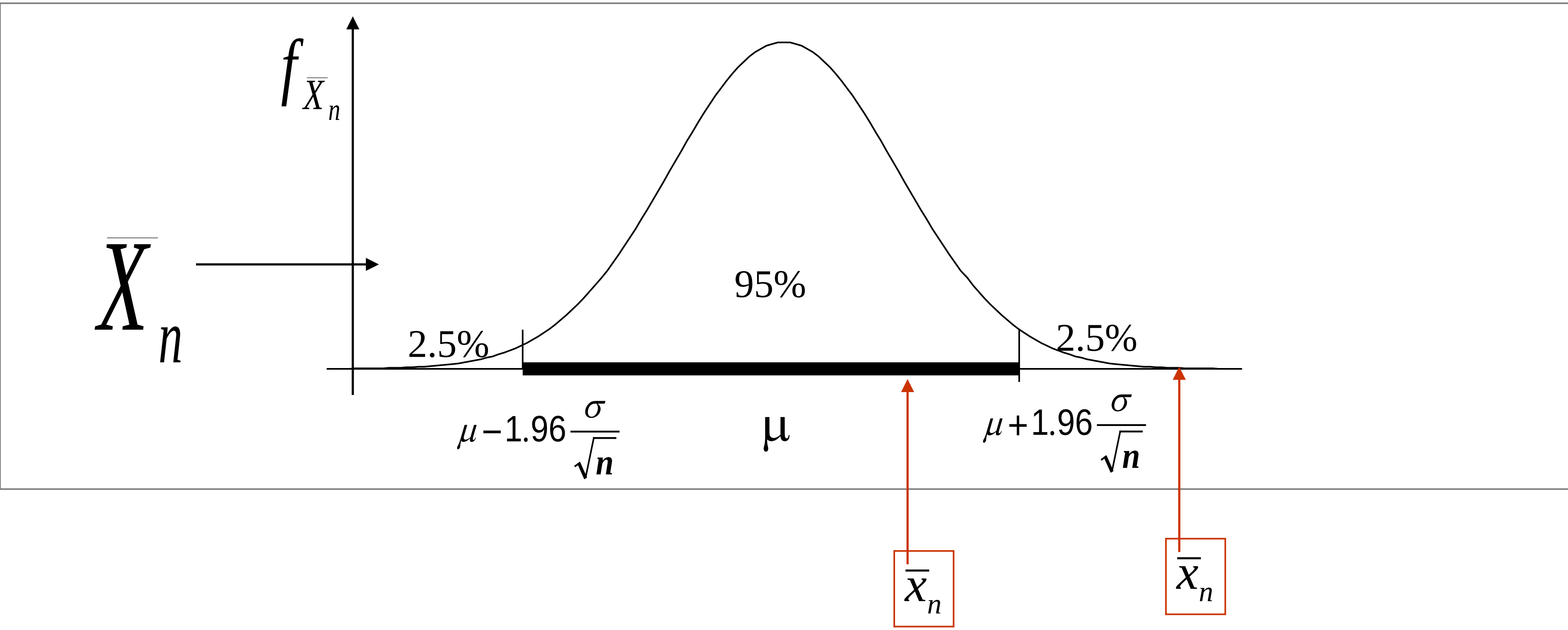
Se observa una muestra, se calcula $\bar{x}_n$

y se desea dar un intervalo de confianza para μ

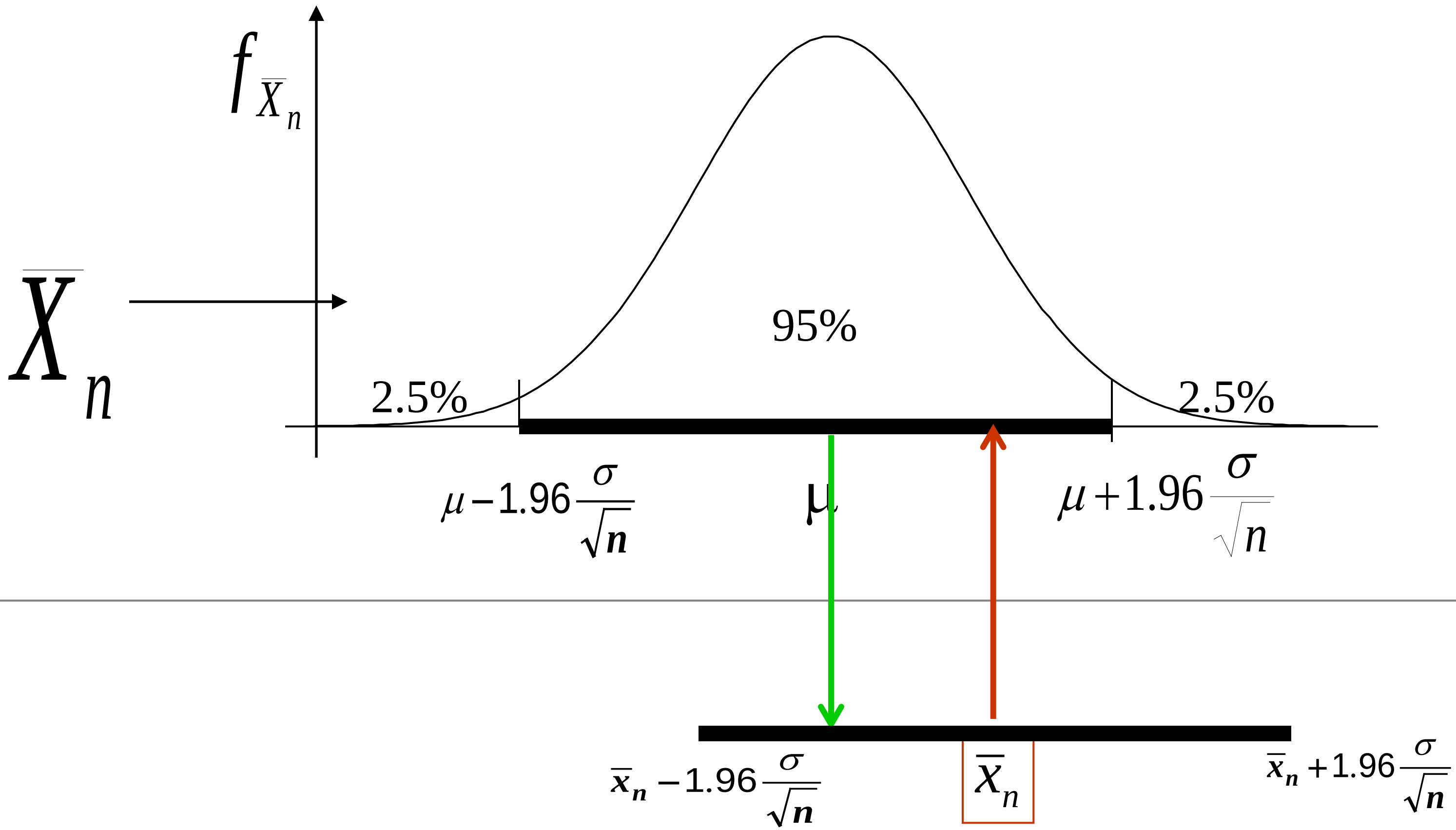
$$\left[\bar{x}_n - d\,,\, \bar{x}_n + d\right]$$

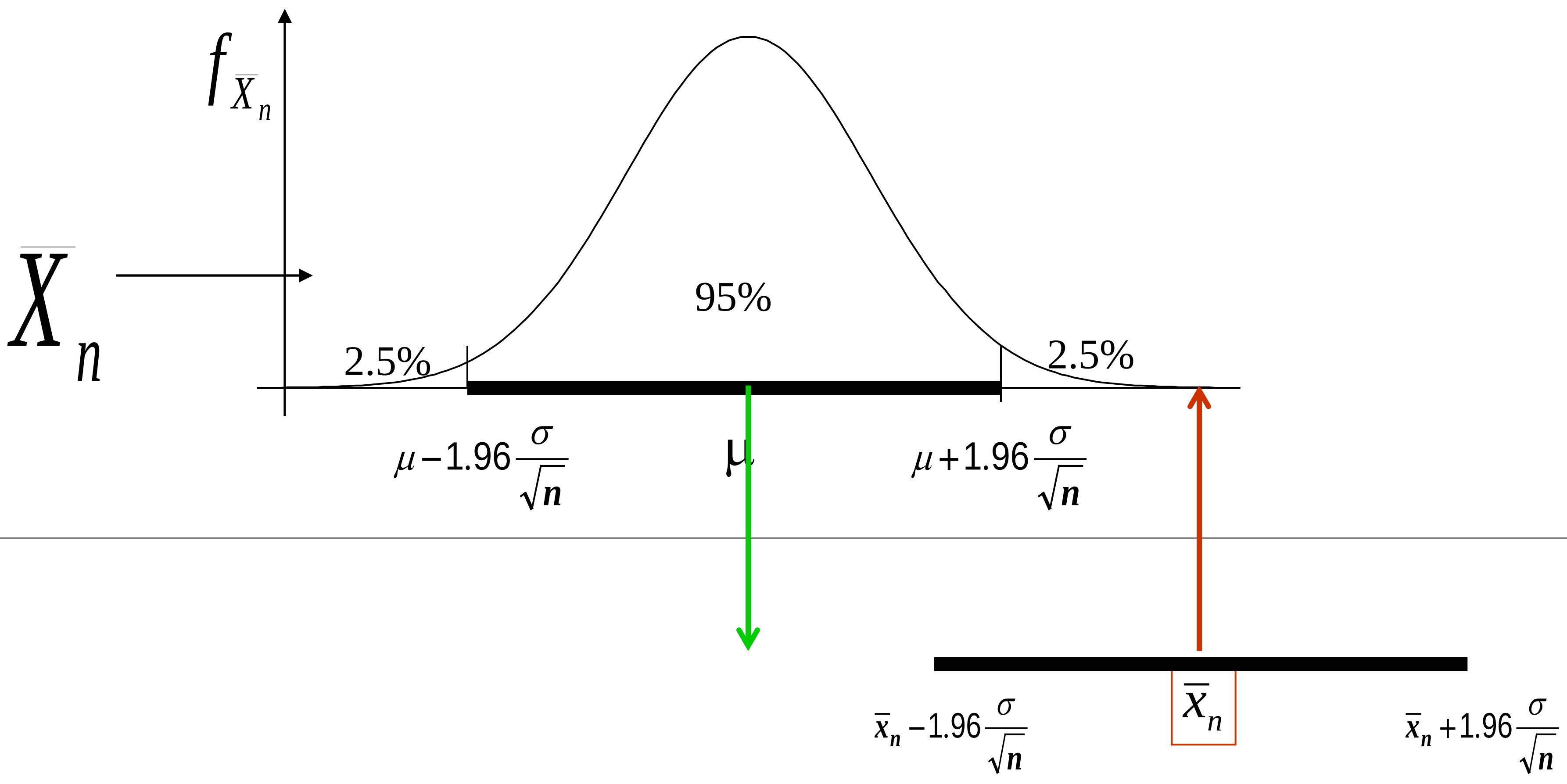
$$\bar{x}_n - d \qquad\qquad \bar{x}_n \qquad\qquad \bar{x}_n + d$$

¿cómo calcular d? ¿qué es la confianza?

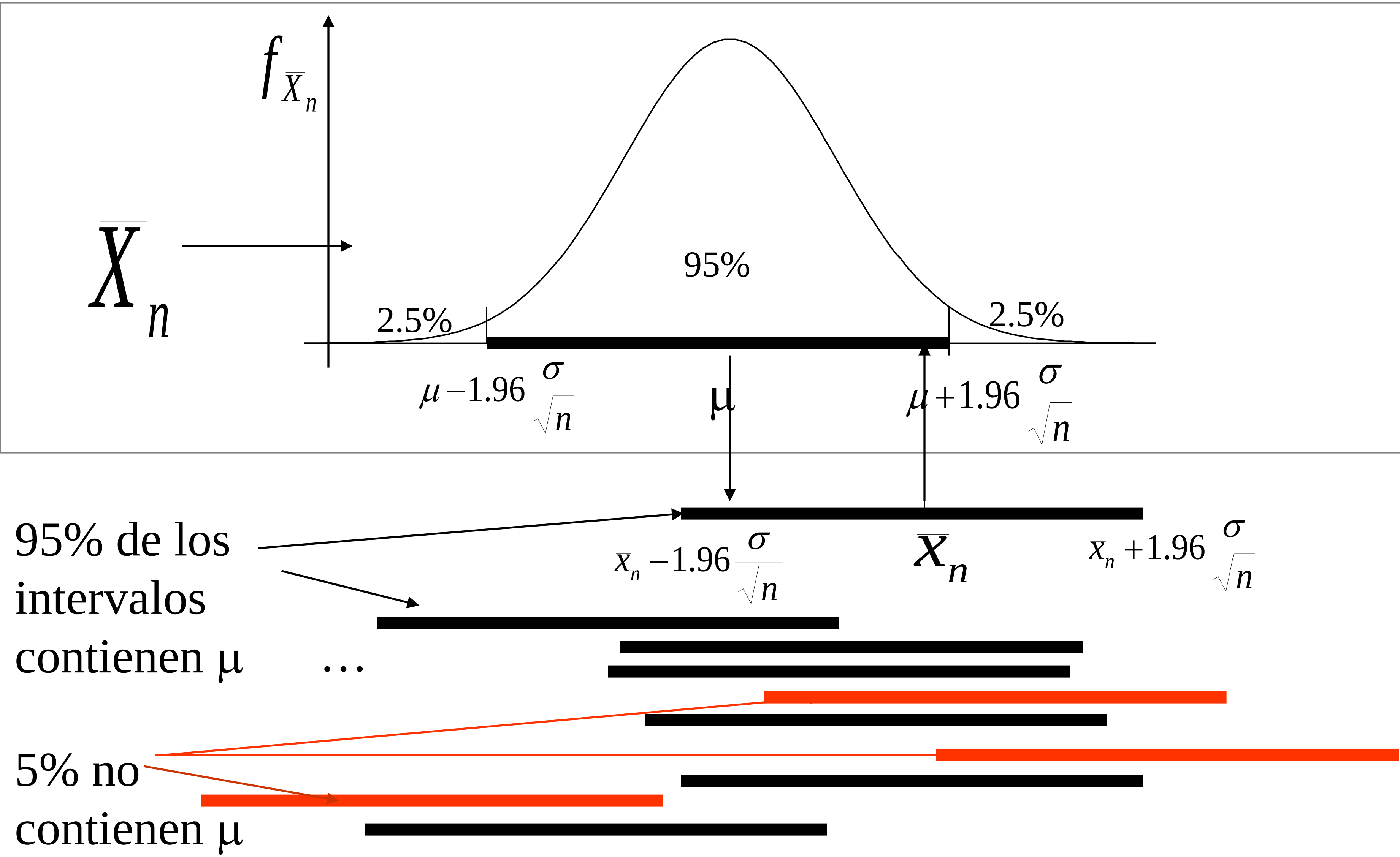


$f_{\bar{X}_n}$
$\bar{X}_n$
2.5%
95%
2.5%
$\mu - 1.96\frac{\sigma}{\sqrt{n}}$
μ
$\mu + 1.96\frac{\sigma}{\sqrt{n}}$
$\bar{x}_n$
$\bar{x}_n$


$f_{\overline{X}_n}$
$\overline{X}_n$
95%
2.5%
2.5%
$\mu - 1.96\frac{\sigma}{\sqrt{n}}$
μ
$\mu + 1.96\frac{\sigma}{\sqrt{n}}$
$\overline{x}_n - 1.96\frac{\sigma}{\sqrt{n}}$
$\overline{X}_n$
$\overline{x}_n + 1.96\frac{\sigma}{\sqrt{n}}$


$f_{\overline{X}_n}$
$\overline{X}_n$
2.5%
95%
2.5%
$\mu - 1.96\frac{\sigma}{\sqrt{n}}$
μ
$\mu + 1.96\frac{\sigma}{\sqrt{n}}$
$\bar{x}_n - 1.96\frac{\sigma}{\sqrt{n}}$
$\overline{X}_n$
$\bar{x}_n + 1.96\frac{\sigma}{\sqrt{n}}$

Intervalo de confianza para μ, σ conocido



En la práctica:

Después de observar la muestra, se puede afirmar: con una confianza de nivel (1-α), la verdadera media pertenece al intervalo:

$$\left(\bar{x}_n - \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha/2} , \bar{x}_n + \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha/2} \right)$$

Forma de la distribución del estimador + nivel de confianza

$$\left(\bar{x}_n - \sqrt{V\left(\bar{X}_n\right)} \cdot z_{1-\alpha/2} , \bar{x}_n + \sqrt{V\left(\bar{X}_n\right)} \cdot z_{1-\alpha/2} \right)$$

Valor observado del estimador

Varianza del estimador

Estimador sin sesgo, para que lo que se afirma sobre la confianza sea cierto

Si desconocido, se estima a partir de la muestra:

$$s^2 = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}_n\right)^2}{n-1}$$

Lo que permite estimar la varianza del estimador:

$$\hat{V}\left(\bar{X}_n\right) = \frac{s^2}{n}$$

La distribución ya no es normal, pero si *n* es muy grande….

$$\left(\bar{x}_n - \sqrt{\hat{V}\left(\bar{X}_n\right)} \cdot z_{1-\alpha/2}, \bar{x}_n + \sqrt{\hat{V}\left(\bar{X}_n\right)} \cdot z_{1-\alpha/2}\right)$$

Estimación de la varianza del estimador

Problema

Se desea estimar la verdadera longitud media (en metro) de las barras de acero fabricadas por una determinada máquina.

Se observa una muestra de tamaño 200. Sobre esta muestra, se observa una media de 1,02m y una varianza muestral igual a $0.01m^2$

Estimar la longitud media por punto y por intervalo *lim_sup* = 1,02 + 1,96*0,1/sqrt(200) = 1,03386

lim_inf = 1,02 - 1,96*0,1/sqrt(200) = 1,00614

IC de confianza de nivel 95%
para la media poblacional:

$$IC(\mu, 0.95) = [1.006, \quad 1.034]$$

Segundo estimador?

$$\hat{\theta} = Media_{truncada} = \frac{\sum_{i=2}^{n-1} X_i}{n-2}$$

Sobre la muestra, toma el valor :

$$\hat{\theta} = Media_{truncada} = \frac{\sum_{i=1}^{n} x_i}{n}$$