UNIDAD 4

Distribuciones Muestrales

Mat. Fernando Sandoya, PhD

FCNM
Facultad de Ciencias Naturales y Matemáticas

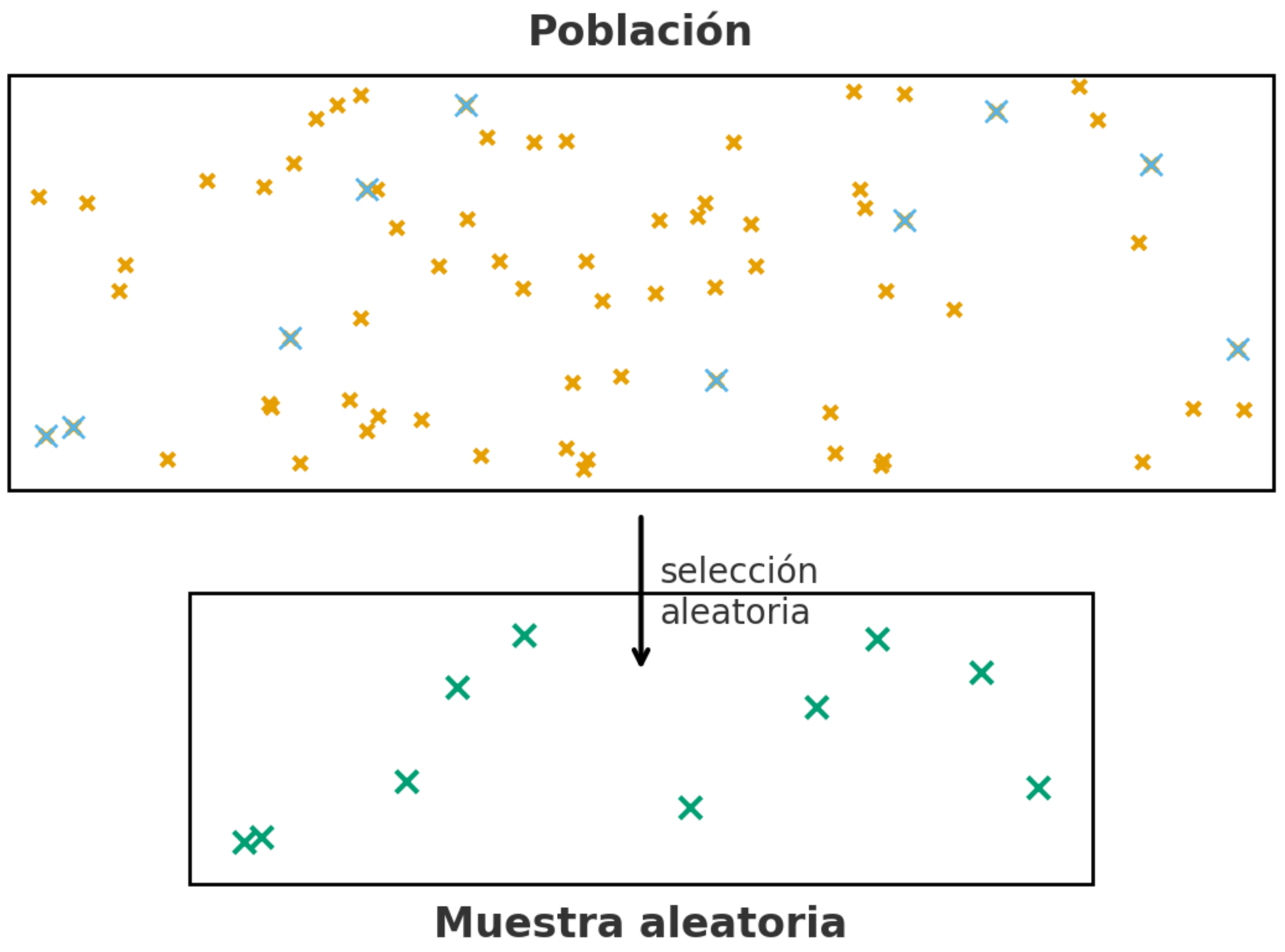
Contenido

Poblaciones y muestras aleatorias

Una muestra para que sea aleatoria tiene que ser considerada en dos contextos:

Población Finita

SI tenemos una población finita, una **muestra** es una selección de *n* elementos de la misma.
Sea una muestra $X_1, X_2, \ldots, X_n$ que es tomada de una ***población finita*** *de tamaño N,* que tiene distribución *f*, tal muestra es denominada **Muestra Aleatoria**, si y solo si al tomarla, cualquier subconjunto de tamaño *n* de la población, tiene igual probabilidad de ser seleccionado.



Poblaciones y muestras aleatorias (2)

Una muestra para que sea aleatoria tiene que ser considerada en dos contextos:

Población Infinita (X)

Si la ***población es infinita,*** y tiene una distribución o densidad *f(x)*, una muestra de tamaño *n*, $X_1, X_2, \ldots, X_n$ es **Aleatoria**, si la *n* variables que la conforman son **variables** i**ndependientes e Idénticamente distribuidas (*iid*)**.

Que sean idénticamente distribuidas las n variables significa que tienen la misma función de Densidad

Poblaciones y muestras aleatorias (3)

Seleccionar una muestra aleatoria no es fácil… es necesaria una representación simbólica de la población objetivo que desea investigarse. A esta representación se le conoce como **Marco Muestral**.

El principal insumo para este Marco Muestral son los registros administrativos, bases de datos, cartografía.

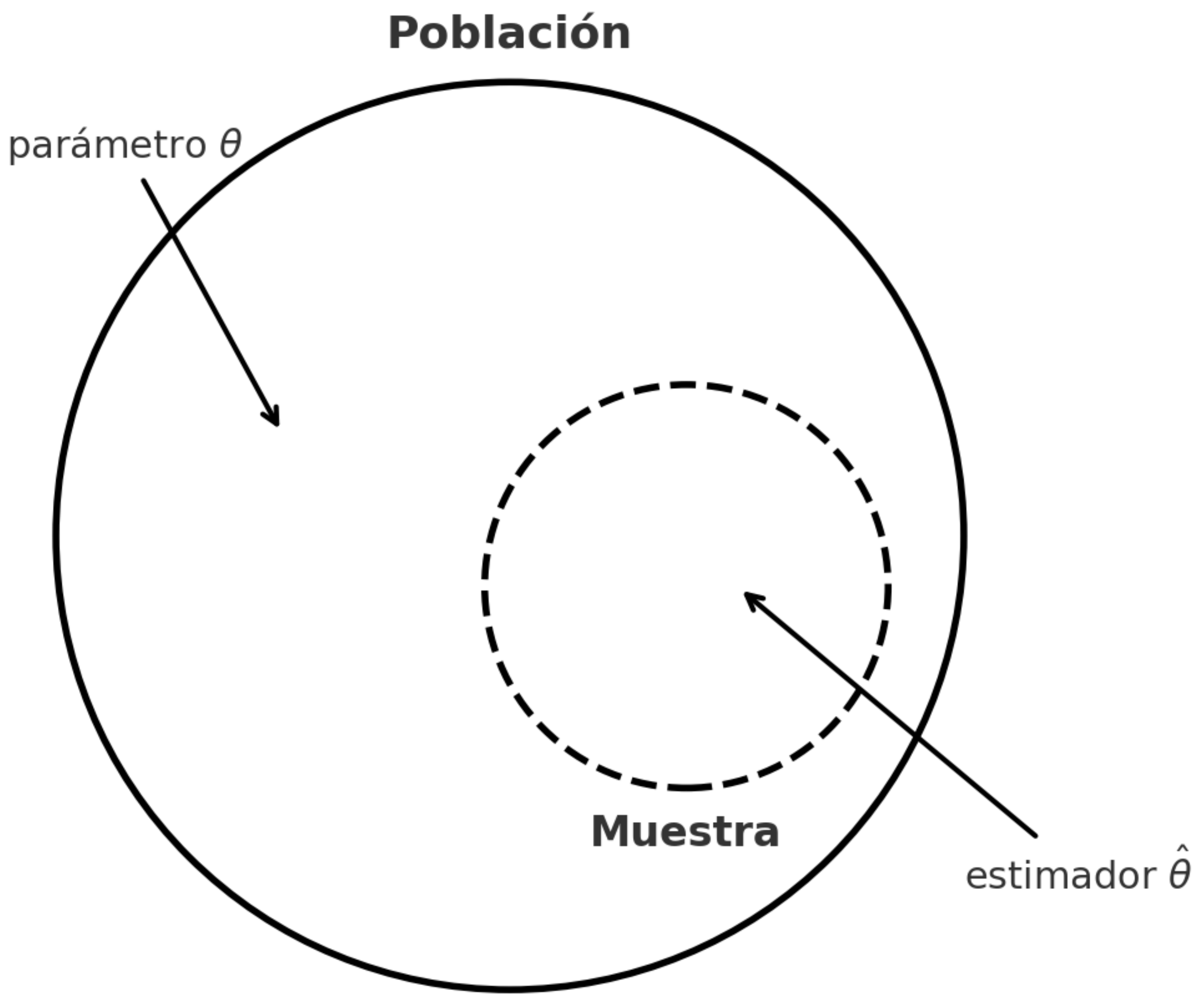
Poblaciones y muestras aleatorias (4)

¿Qué es un Parámetro Poblacional?

Es una cantidad numérica calculada sobre una población que representa alguna característica o atributo de la misma.

¿Qué es un Estimador Muestral?

Es un valor $\hat{\theta}$ calculado a partir de las observaciones muestrales $X_1, X_2, \ldots, X_n$ que componen una muestra aleatoria. Su objetivo es estimar un parámetro poblacional θ



La población tiene un parámetro desconocido θ;
la muestra genera el estimador $\hat{\theta}$.

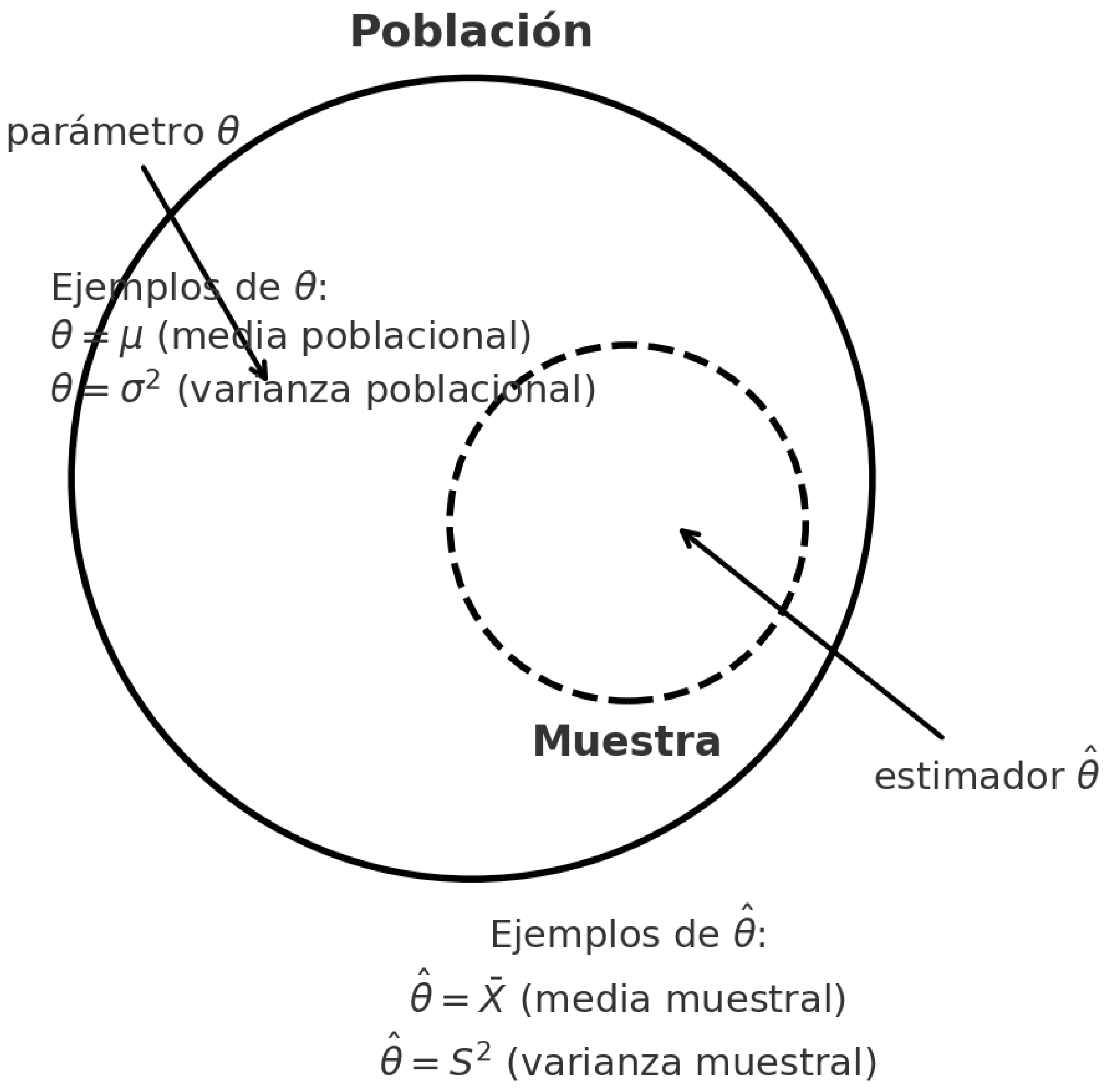
Mat. Fernando Sandoya, PhD

Distribuciones muestrales

Si $x_1, x_2, \dots, x_n$ es una muestra aleatoria (m.a.) de una población, se denominan distribución muestral a la distribución de probabilidades de cualquier estimador muestral $\hat{\theta}$.

Las principales distribuciones muestrales son:

- Distribución muestral de la media
- Distribución muestral de la varianza



Distribuciones muestrales de la media muestral

Si $x_1, x_2, \ldots, x_n$ es una m.a. de una población con media $\mathbf{E(X)=}\boldsymbol{\mu}$ y $\mathbf{Var(X)=}\boldsymbol{\sigma^2} < \infty$

La media muestral dada por: $\overline{\boldsymbol{X}} = \frac{\sum_{i=1}^{n} X_i}{n}$ **tiene la siguiente esperanza y varianza:**

$$E(\bar{X}) = E\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{\sum_{i=1}^{n} E(X_i)}{n} = \frac{n\boldsymbol{\mu}}{n} = \boldsymbol{\mu}$$

$$\mathrm{Var}(\bar{X}) = Var\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{\sum_{i=1}^{n} Var(X_i)}{n^2} = \frac{\boldsymbol{\sigma^2}}{\boldsymbol{n}}$$

OBSERVACIÓN:

- $\boldsymbol{E(X) = E(\overline{X})}$
- $\boldsymbol{\sigma^2_{\bar{x}} \leq \sigma^2_x}$

Distribución muestral de la varianza muestral

Si $x_1, x_2, \ldots, x_n$ es una m.a. de una población con media $\mathbf{E(X)}= \boldsymbol{\mu}$ y $\mathbf{Var(X)}=\boldsymbol{\sigma^2} < \infty$

La varianza muestral está dada por: $\boldsymbol{S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}}$

Si descomponemos la varianza tenemos:

$$(\mathrm{n}-1)\mathrm{S}^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n}(X_i - \boldsymbol{\mu} + \boldsymbol{\mu} - \bar{X})^2$$

$$= \sum_{i=1}^{n}(X_i - \boldsymbol{\mu})^2 + n(\bar{X} - \boldsymbol{\mu})^2 + 2(\bar{X} - \boldsymbol{\mu})\mathrm{n}(\bar{X} - \boldsymbol{\mu})$$

$$= \sum_{i=1}^{n}(X_i - \boldsymbol{\mu})^2 - n(\bar{X} - \boldsymbol{\mu})^2$$

La variabilidad respecto a la media poblacional se descompone en: variabilidad respecto a la media poblacional y la variabilidad entre la media poblacional y muestral

Distribución muestral de la varianza muestral (2)

Respecto al valor esperado y varianza se tiene que:

$$E(S^2) = E\left(\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1}\right)$$

$$= \sum_{i=1}^{n} E(X_i - \boldsymbol{\mu} + \boldsymbol{\mu} - \bar{X})^{\,2} = \frac{1}{n-1} E\left(\sum_{i=1}^{n}(X_i - \boldsymbol{\mu})^2 - n\,(\bar{X} - \boldsymbol{\mu}\,)^2\right)$$

$$= \frac{1}{n-1}\left(\sum_{i=1}^{n} \sigma^2 - n\,E((\bar{X} - \boldsymbol{\mu})^2)\right) = \frac{1}{n-1}\left(n\sigma^2 - \frac{n\sigma^2}{n}\right) = \sigma^2$$

$$\mathrm{Var}(S^2) = \ldots = \frac{2\sigma^4}{n-1}$$

S^2 es un estimador insesgado σ^2. La distribución de S^2 no es simétrica y su forma depende de n y la distribución de X_i. No puede ser aproximada por una distribución normal.

Teorema de Límite Central y aplicaciones

Sea $x_1, x_2, x_3, \ldots$ una muestra de una población con media E(X$) = \boldsymbol{\mu}$ **y** $\boldsymbol{Var(X_i) = \sigma^2}$ finita,

Entonces, la distribución de $\bar{X}$ converge a una distribución normal con media $\boldsymbol{\mu}$ y varianza $\boldsymbol{\sigma^2/n}$, es decir:

$$\bar{X} \xrightarrow{\mathrm{n}\to\infty} \mathcal{N}(\mu, \frac{\sigma^2}{n})$$

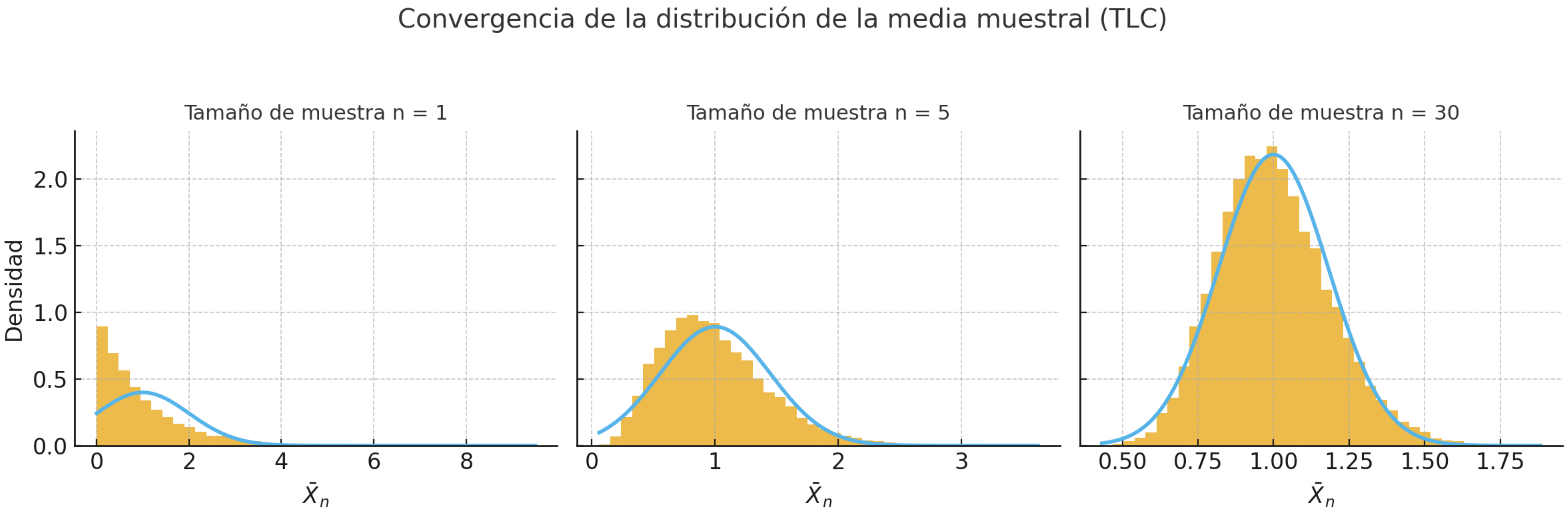
Es decir, bajo estas condiciones, $Z = \frac{(\bar{X}-\mu)}{\sigma/\sqrt{n}} \xrightarrow{\mathrm{n}\to\infty} \mathcal{N}(0,1)$

(Z converge en Distribución a una Variable Aleatoria Normal Estándar.)

Teorema de Límite Central y aplicaciones (2)

En la práctica se considera que la aproximación para $\bar{X}$ por medio de una normal es buena para "**muestras grandes**", por lo general la muestra es grande si **$n \geq 30$**

¿Si la muestra es pequeña?: En caso de que **$n < 30$**, se requiere que la población de la que procede la muestra sea aproximadamente normal…



Mat. Fernando Sandoya, PhD

Teorema de Límite Central y aplicaciones (3)

Ejemplo:

Una empresa de material eléctrico fabrica bombillas que tienen una duración que se distribuye aproximadamente en forma normal, con media de 800 horas y desviación estándar de 40 horas. Calcule la probabilidad de que una muestra aleatoria de 16 bombillas tenga una vida promedio de menos de 775 horas.

Teorema de Límite Central y aplicaciones (4)

Una aplicación muy importante del teorema del límite central consiste en determinar valores razonables de la media de la población μ.

Con frecuencia un investigador desea que los datos (en la forma de $\bar{X}$) respalden alguna conjetura predeterminada respecto al valor de μ

Teorema de Límite Central y aplicaciones (5)

Ejemplo (Inferencia sobre la media poblacional):

Un importante proceso de fabricación produce partes de componentes cilíndricos para la industria automotriz. Es importante que el proceso produzca partes que tengan un diámetro medio de 5.0 milímetros. El ingeniero implicado asume que la media de la población es de 5.0 milímetros. Se lleva a cabo un experimento donde se seleccionan al azar 100 partes elaboradas por el proceso y se mide el diámetro de cada una de ellas. Se sabe que la desviación estándar de la población es $\sigma = 0.01$ milímetros. El experimento indica un diámetro promedio muestral de $\bar{X} = 5{,}027$ milímetros. ¿Esta información de la muestra parece apoyar o refutar la suposición del ingeniero?

Ejercicios

Las puntuaciones en la Escala de Inteligencia para Adultos de Wechsler (WAIS) es un proceso de media 100 y desviación típica 16. Al extraer una muestra aleatoria simple de 35 individuos, calcular:

a) Probabilidad de que la media de esos individuos sea inferior a 95
b) Probabilidad de que la media esté comprendida entre 98 y 102.

DISTRIBUCIÓN MUESTRAL DE LA VARIANZA MUESTRAL

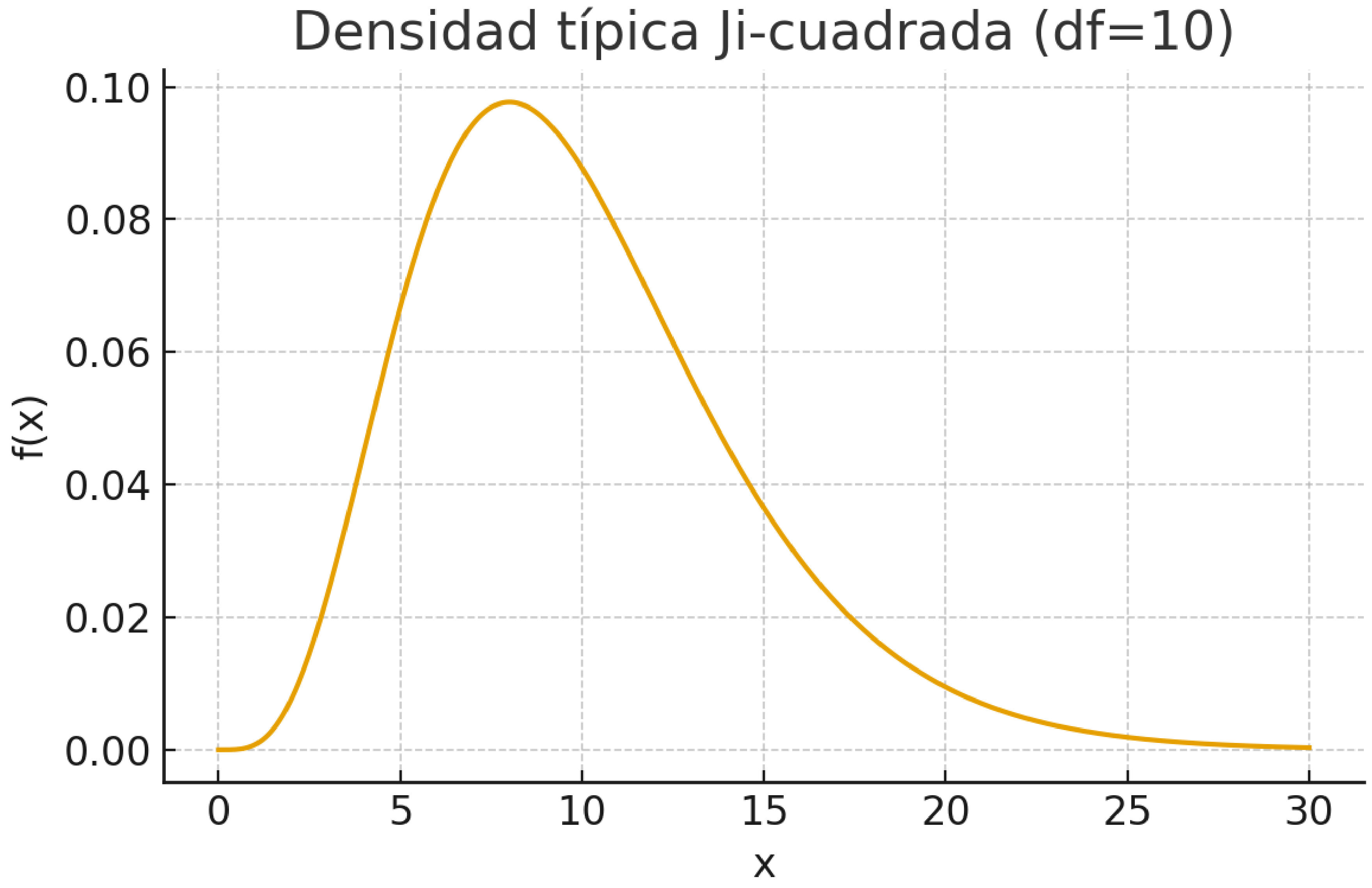
Dada una muestra de tamaño *n* tomada de una **población** que tiene una distribución normal con media μ y varianza σ^2. Entonces la distribución muestral de la varianza muestral S^2 se establece por la siguiente propiedad:

$$\frac{(n-1)}{\sigma^2}S^2 \text{ tiene una distribución } \textbf{Chi−cuadrada} \text{ con } (\text{n}-1) \text{ grados de libertad}$$

Lo cual se escribe como: $\boldsymbol{\frac{(n-1)}{\sigma^2}S^2 \sim \chi^2_{(n-1)}}$

DISTRIBUCIÓN MUESTRAL DE LA VARIANZA MUESTRAL

$$\frac{(n-1)}{\sigma^2} S^2 \sim \chi^2_{(n-1)}$$



Mat. Fernando Sandoya, PhD