

Clase 1a

Introducción a conceptos de error y estadística

<http://materias.df.uba.ar/mtb2019c1/laboratorio-martes-jueves/>

Repaso: Magnitudes

- **Magnitud física:** atributo cuantificable susceptible de ser medido
- **Medición:** Instrumento + método + sistema de unidades (SI)



Metrología

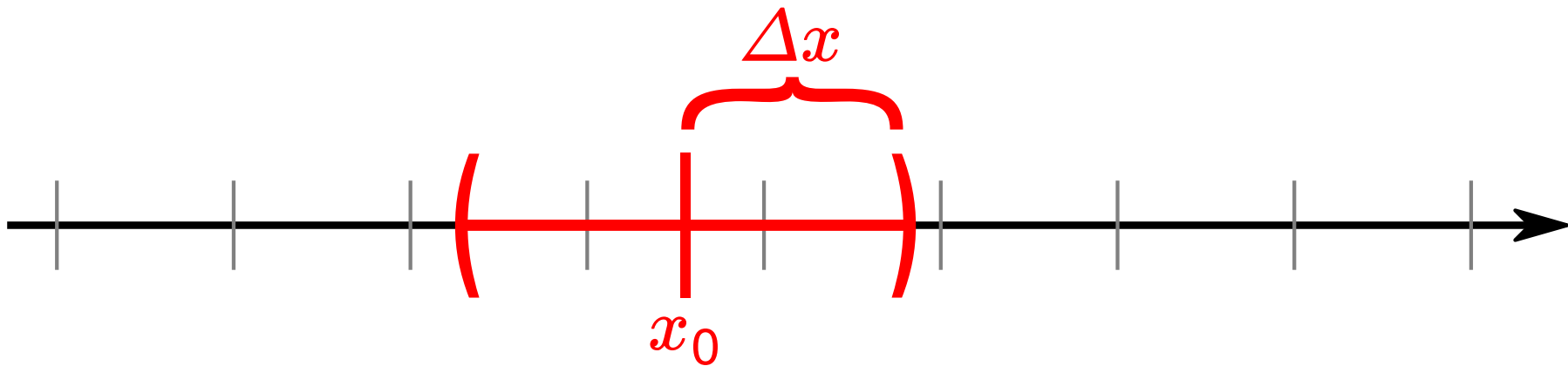
- Definición de unidades atadas a constantes físicas
- Realización de unidades mediante "experimentos"
- Comparación
- Calibración de instrumentos de medición

Repaso: Incertezas

→ Resultado de una medición:

- Valor (con unidad)
- Medida de mi desconocimiento: Incerteza

Intervalo



¿Cómo se reporta el error?

$$\rightarrow (x_0 \pm \Delta x) \text{ u}$$

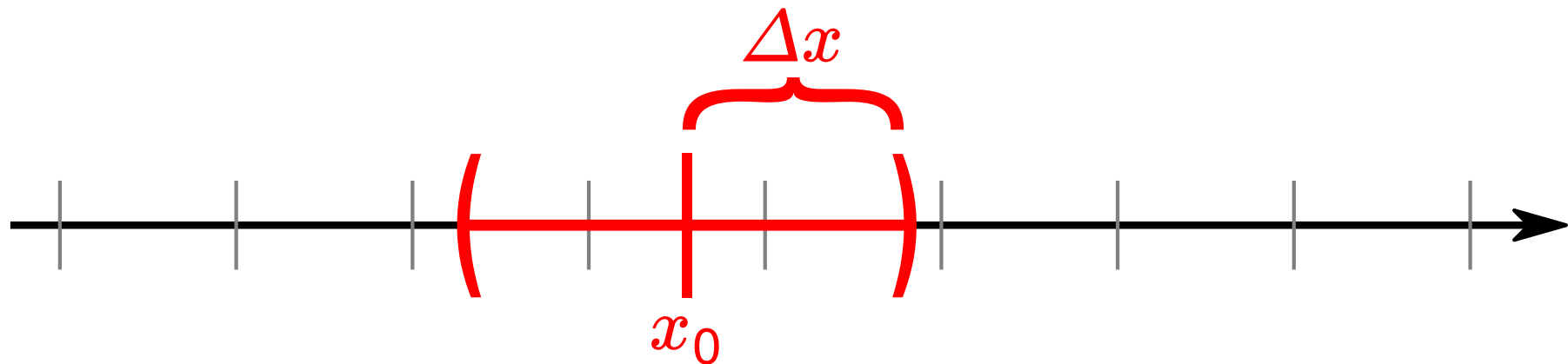
$$\rightarrow (2.23 \pm 0.12) \text{ m}$$

$$\rightarrow (123 \pm 4) \text{ s}$$

$$\rightarrow x_0(\Delta x) \text{ u}$$

$$\rightarrow 2.23(12) \text{ m}$$

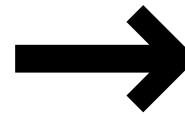
$$\rightarrow 123(4) \text{ s}$$



→ Cifras significativas

$$x_0 = 3745.12845 \text{ m}$$

$$\Delta x = 0.04932 \text{ m}$$



$$x_0 = 3745.13 \text{ m}$$

$$\Delta x = 0.05 \text{ m}$$

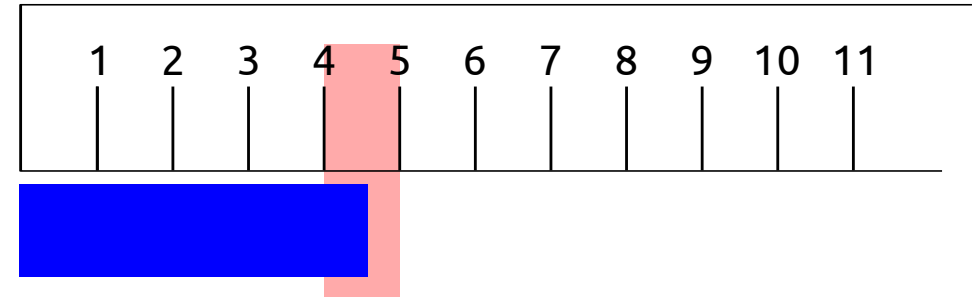
$$x_0 = (3745.13 \pm 0.05) \text{ m}$$

Repaso: Fuentes de error instrumental

→ Error de apreciación σ_{ap}

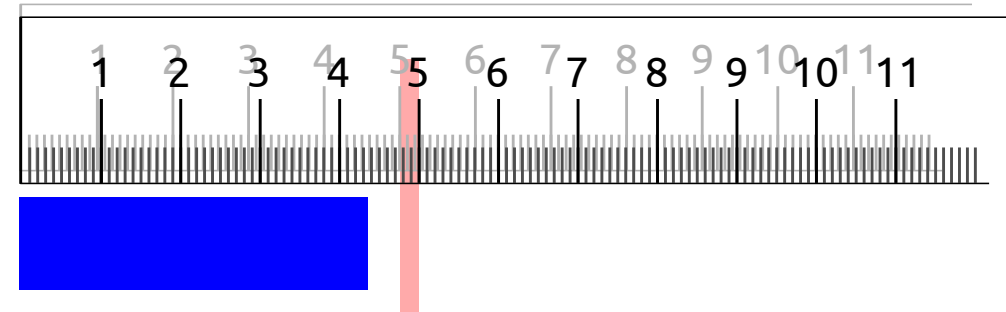
Resolución del instrumento

Presición



→ Error de exactitud σ_{ex}

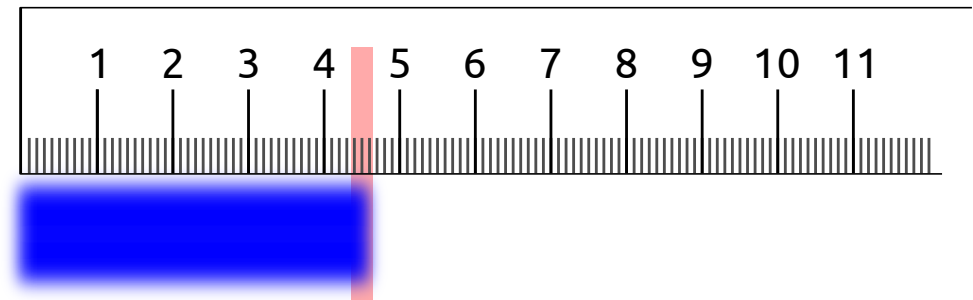
Calibración del instrumento



→ Error de interacción σ_{int}

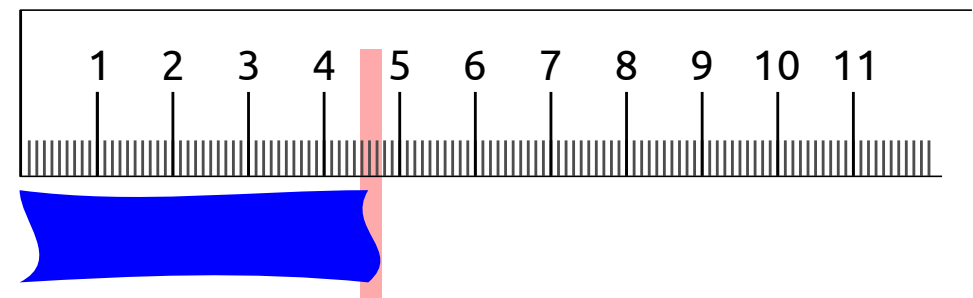
Método de medición

Observador



→ Error de definición σ_{def}

Naturaleza del objeto



Repaso: Fuentes de error instrumental

→ Error de apreciación σ_{ap}

Resolución del instrumento

Presión

→ Error de exactitud σ_{ex}

Calibración del instrumento

$$\sigma_{nom}^2 = \sigma_{ap}^2 + \sigma_{ex}^2 + \sigma_{int}^2 + \sigma_{def}^2$$

→ Error de interacción σ_{int}

Método de medición

Observador

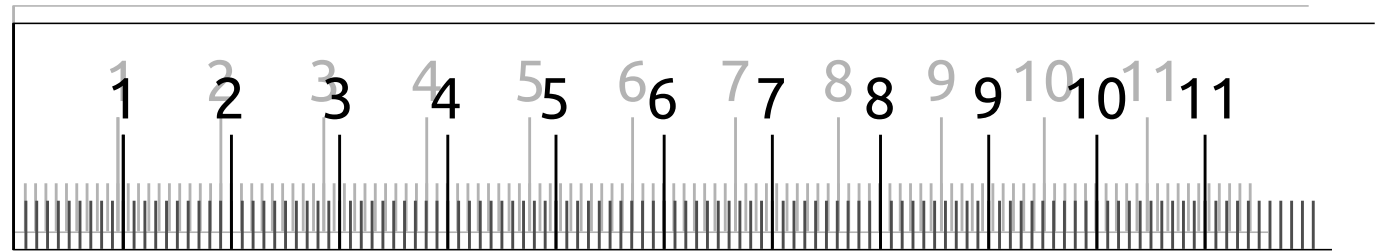
→ Error de definición σ_{def}

Naturaleza del objeto

Repaso: Tipos de error

→ Error Sistemático

- Ej: error instrumental de calibración
- Siempre aporta en un mismo sentido



→ Error Estadístico

- Producto del azar
 - Intrínseco (naturaleza)
 - Desconocimiento



Medición, incertidumbre y estadística (una breve introducción)

Luciano A. Masullo

Laboratorio 1 (1er Cuatrimestre 2018)
Departamento de Física
Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires



Probabilidad

- Variable aleatoria: es una cantidad cuyo valor no es fijo sino que puede tomar diferentes valores como resultados de un experimento aleatorio.
 - La variable aleatoria toma una serie de valores asociados a un conjunto de eventos posibles
 - No se puede predecir el próximo resultado (o valor de la variable aleatoria) pero sí se puede, en principio, conocer su **probabilidad**.
- Probabilidad: es una función que va de un conjunto (de eventos posibles) y les asigna un valor entre 0 (imposible) y 1 (seguro).
 - La suma de las probabilidades de todos los elementos del conjunto de eventos debe ser igual a 1.
 - El conjunto de eventos posibles puede ser discreto (ej: “resultados de tirar un dado”) o continuo (ej: “resultados de la medición del lado de una mesa”).

Resultados de un experimento

Tirar un dado $N = 100$ veces

Medición #	Cara del dado
1	2
2	6
3	1
...	...
99	4
100	1



Medir el período de un faro $N = 100$ veces

Medición #	Tiempo (s)
1	1,02
2	0,98
3	1,07
...	...
99	1,22
100	1,10



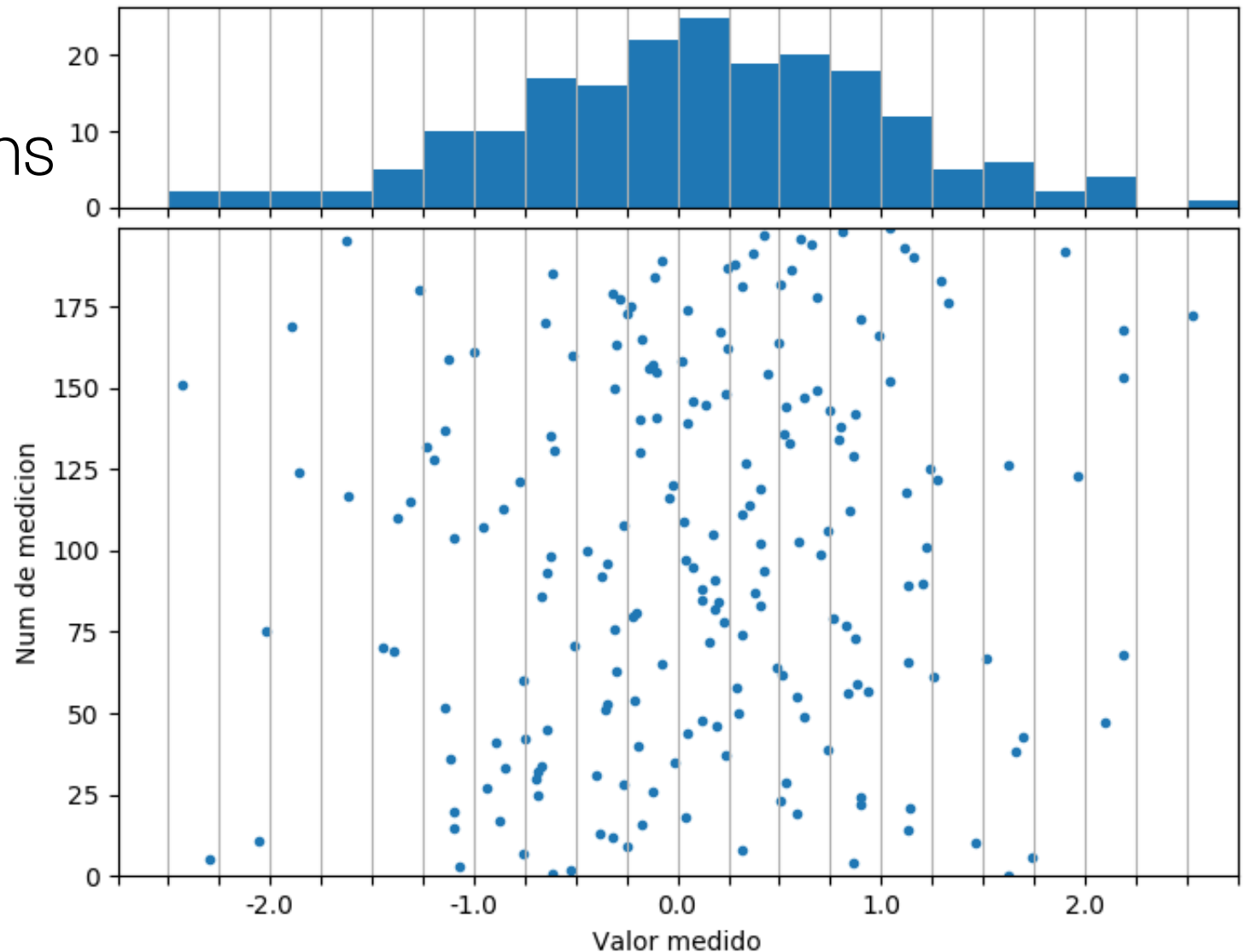
Histogramas

→ **Relevar variable aleatoria**

→ bin size

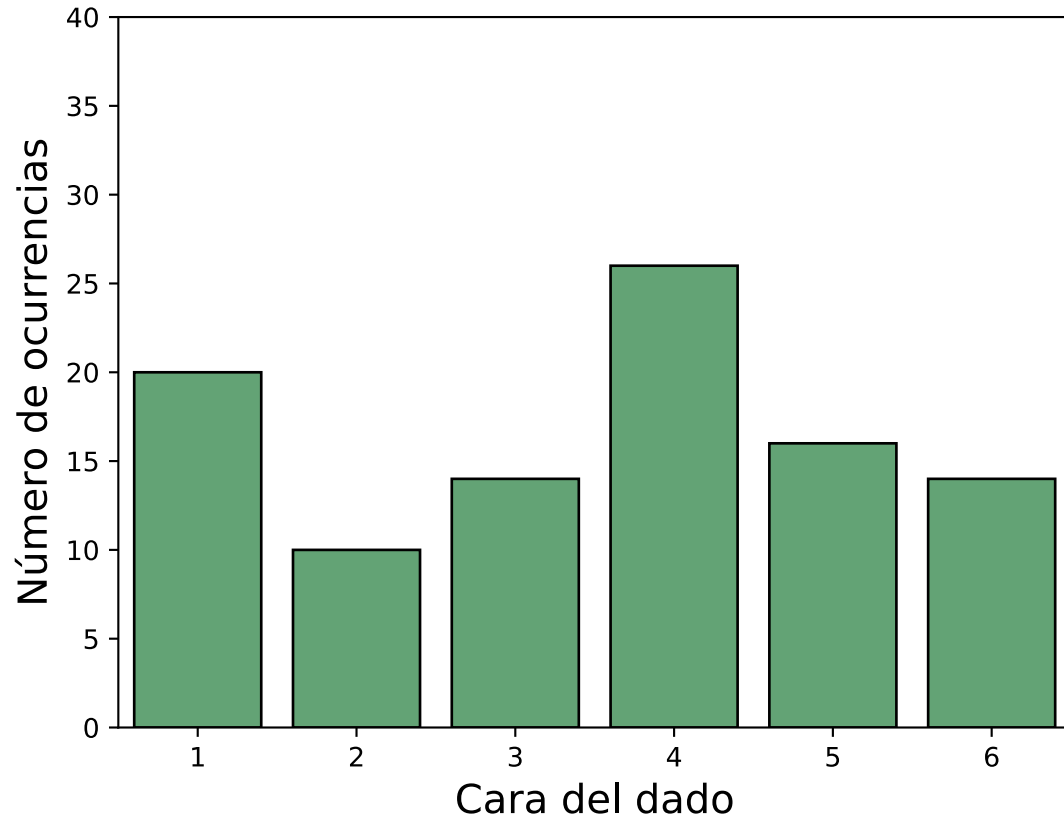
→ Límites

→ Numero de bins

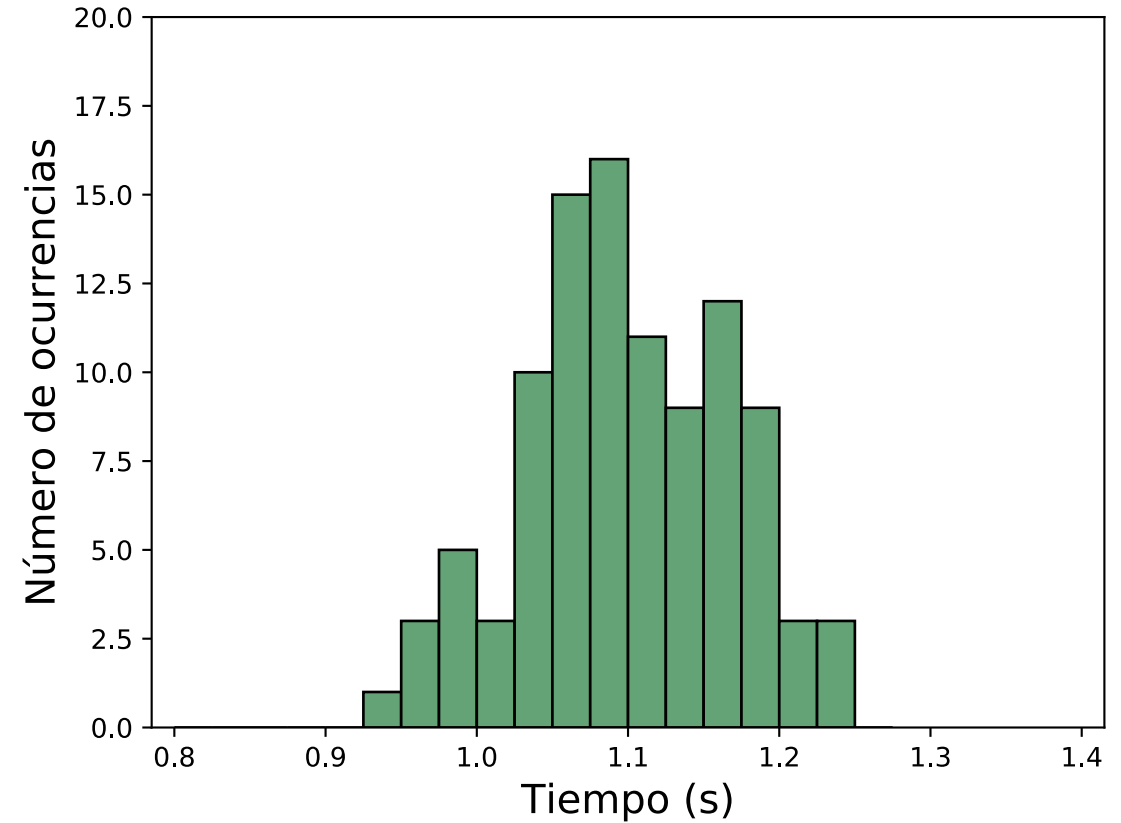


Resultados de un experimento: histograma

Tirar un dado N = 100 veces



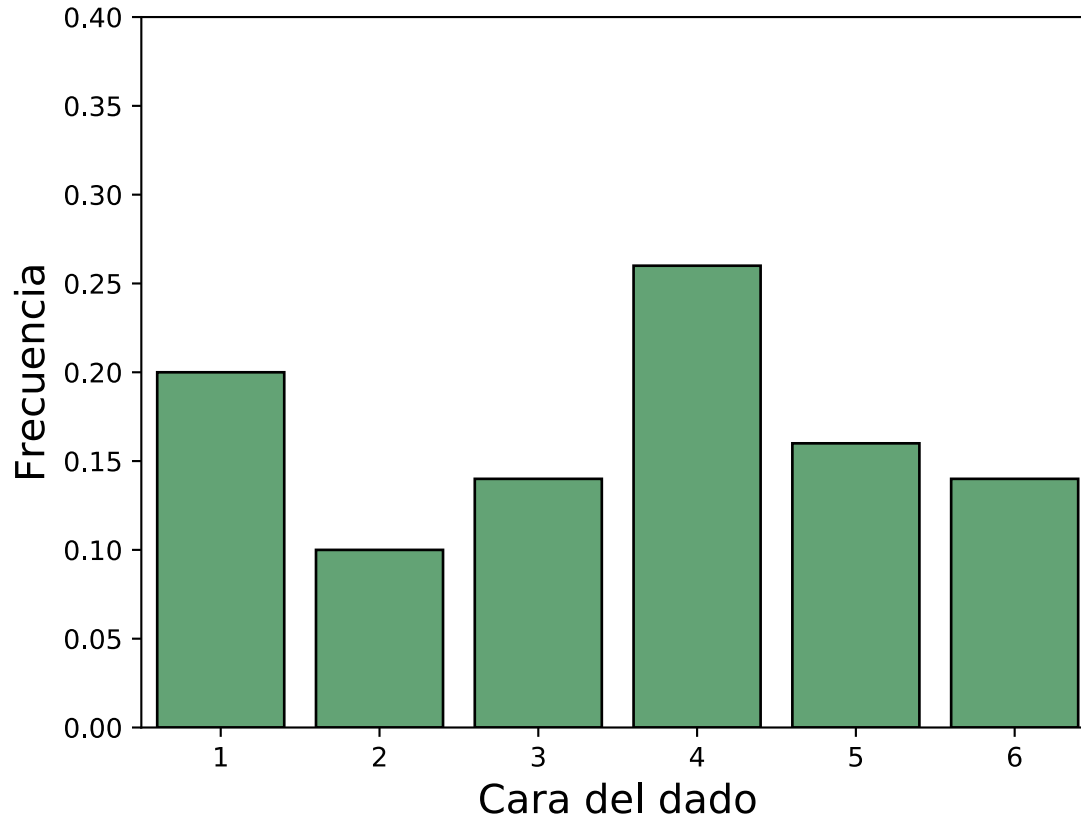
Medir el período de un faro N = 100 veces



Histograma normalizado

F_k = frecuencia del resultado k-ésimo
 n_k = número de resultados en el bin k-ésimo
 d_k = densidad de probabilidad del bin k-ésimo
 a = tamaño del bin (bin size)
 N = número total de eventos

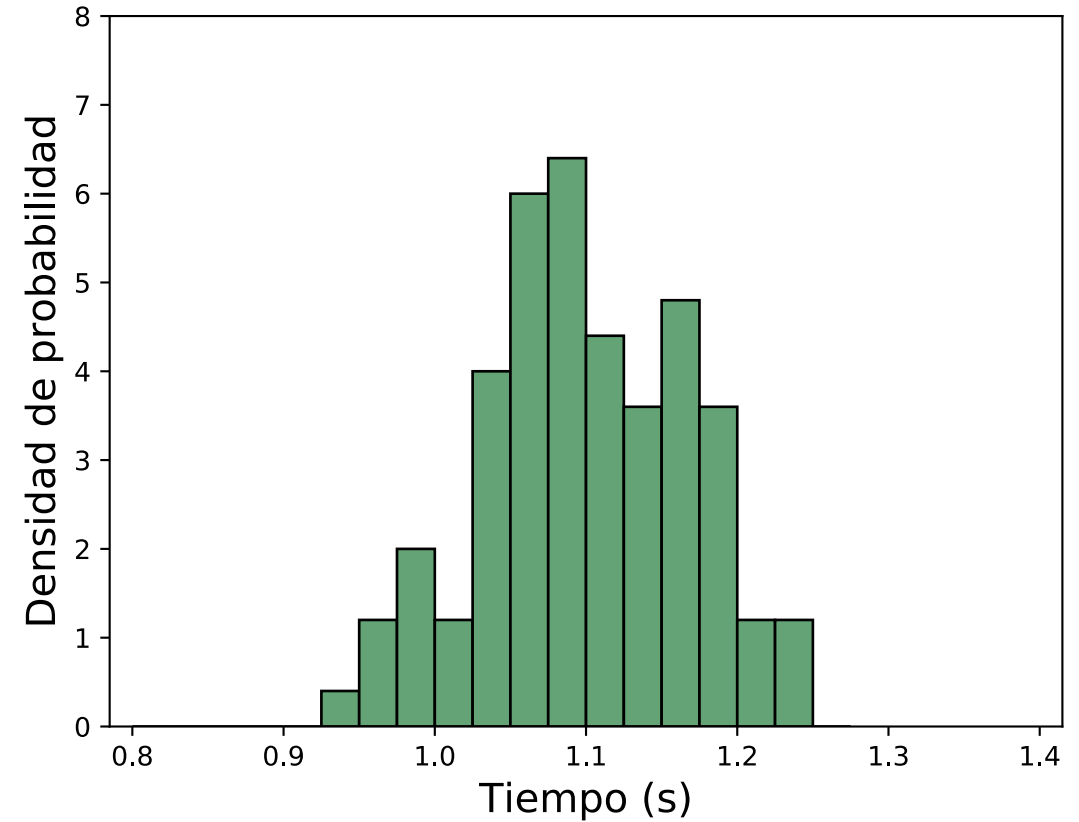
Tirar un dado N = 100 veces



$$F_k = \frac{n_k}{N}$$

Normalización: $\sum_k F_k = 1$

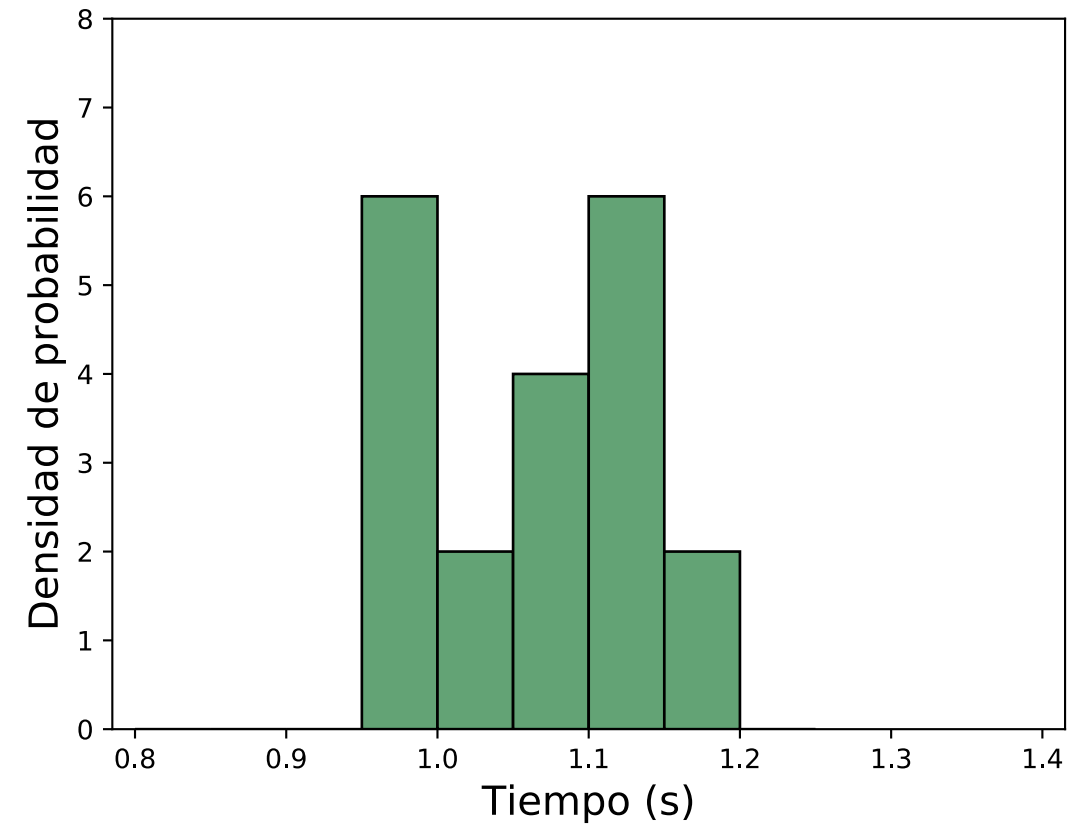
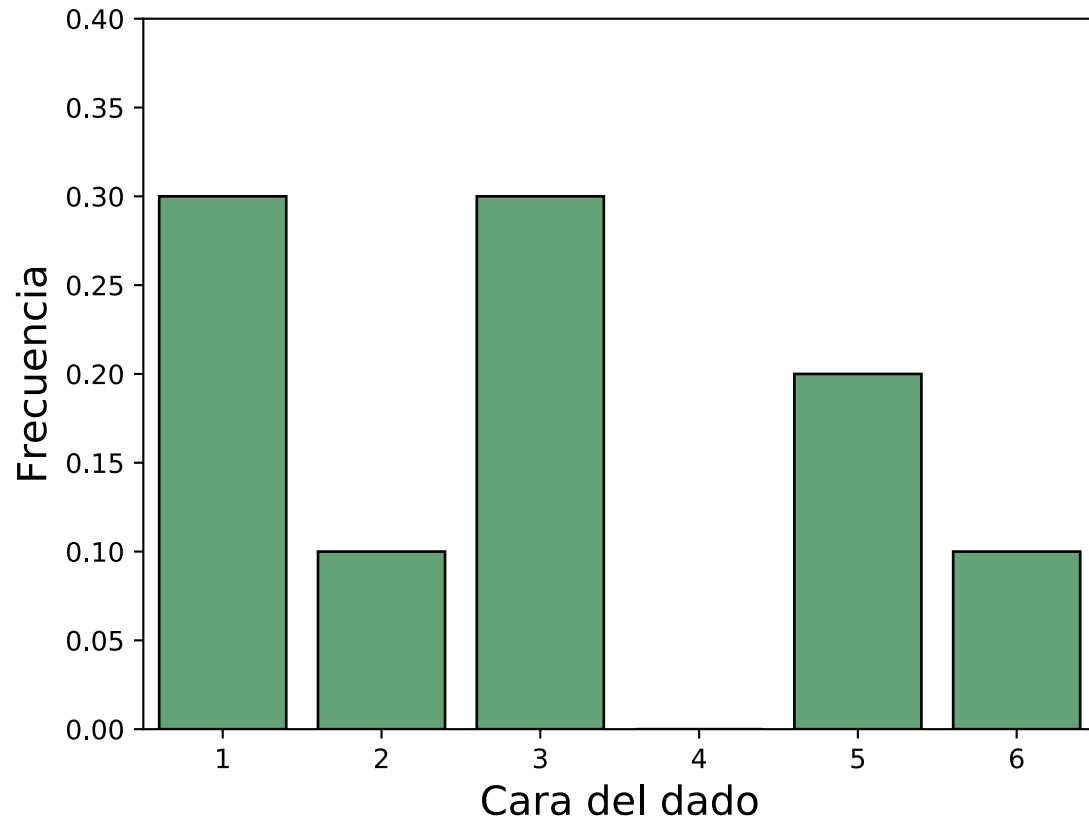
Medir el período de un faro N = 100 veces



$$F_k = d_k a \quad \text{con} \quad d_k = \frac{n_k}{a N}$$

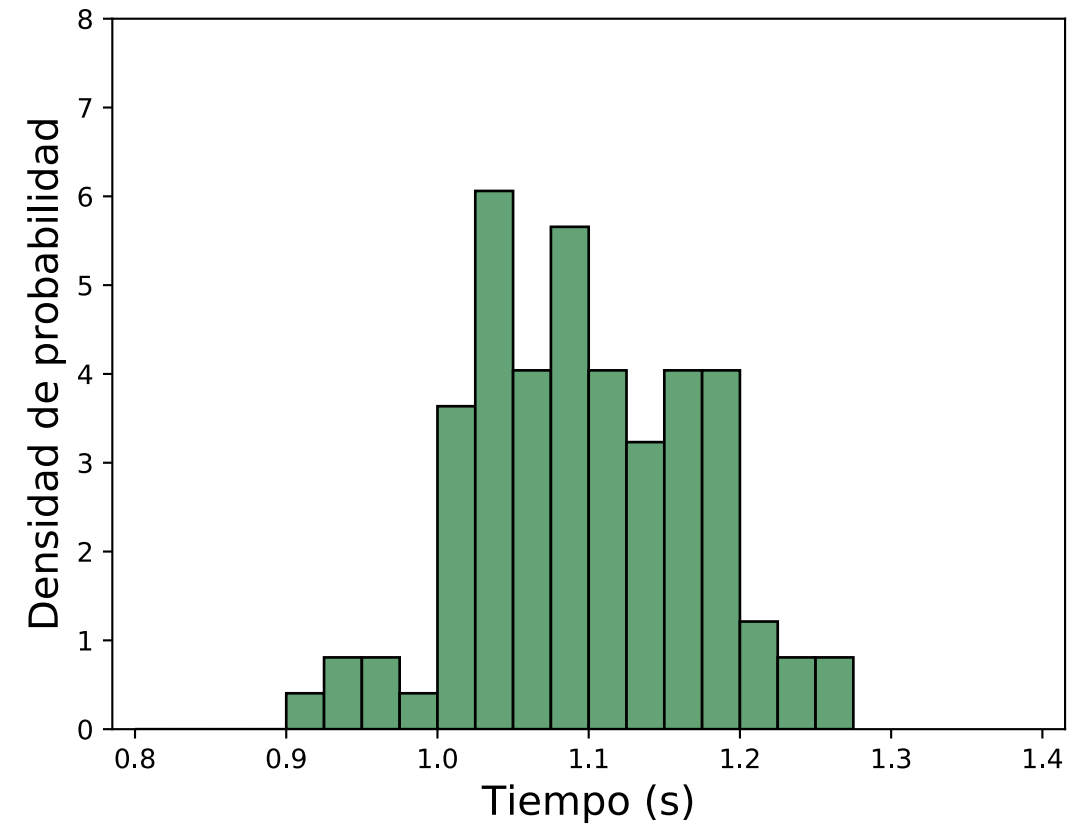
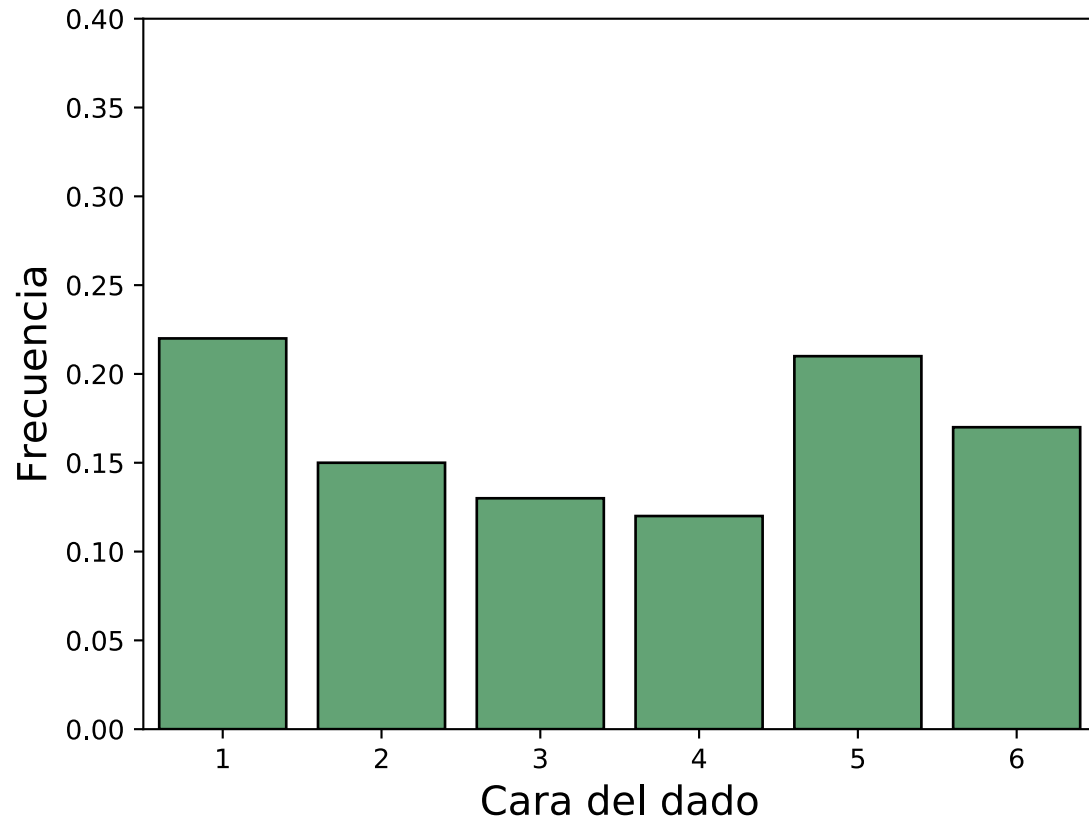
Distribución de probabilidad

N = 10



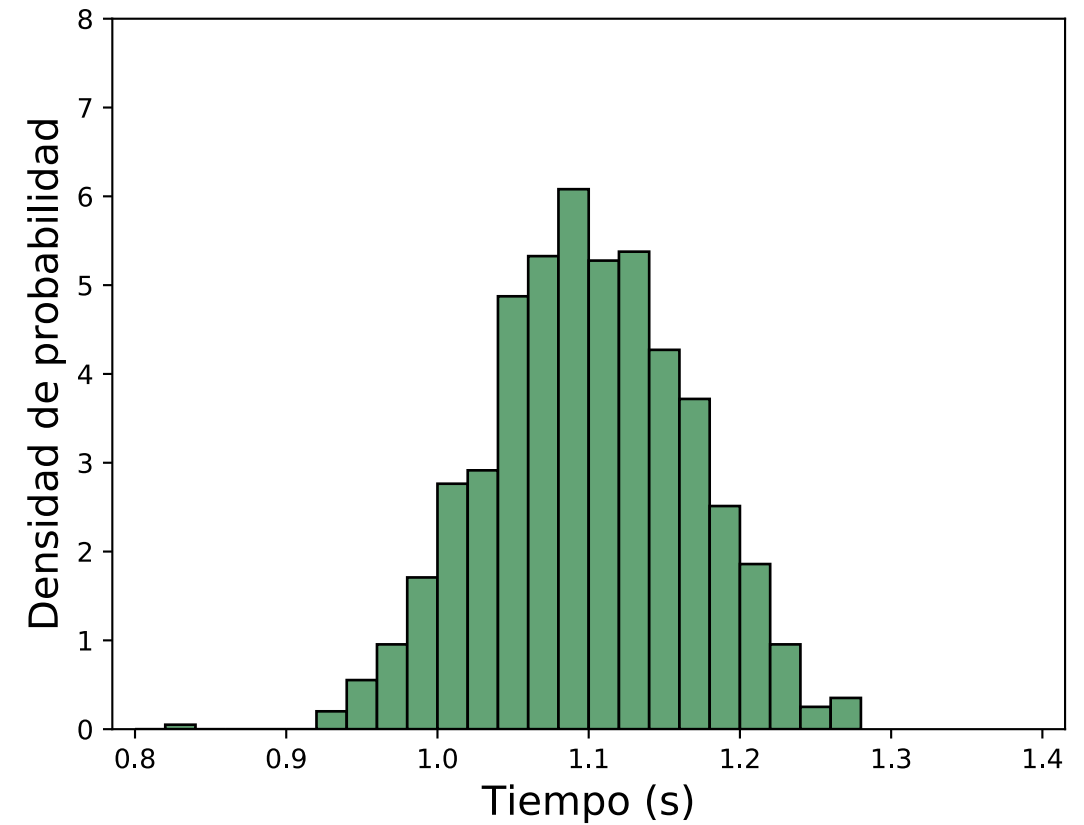
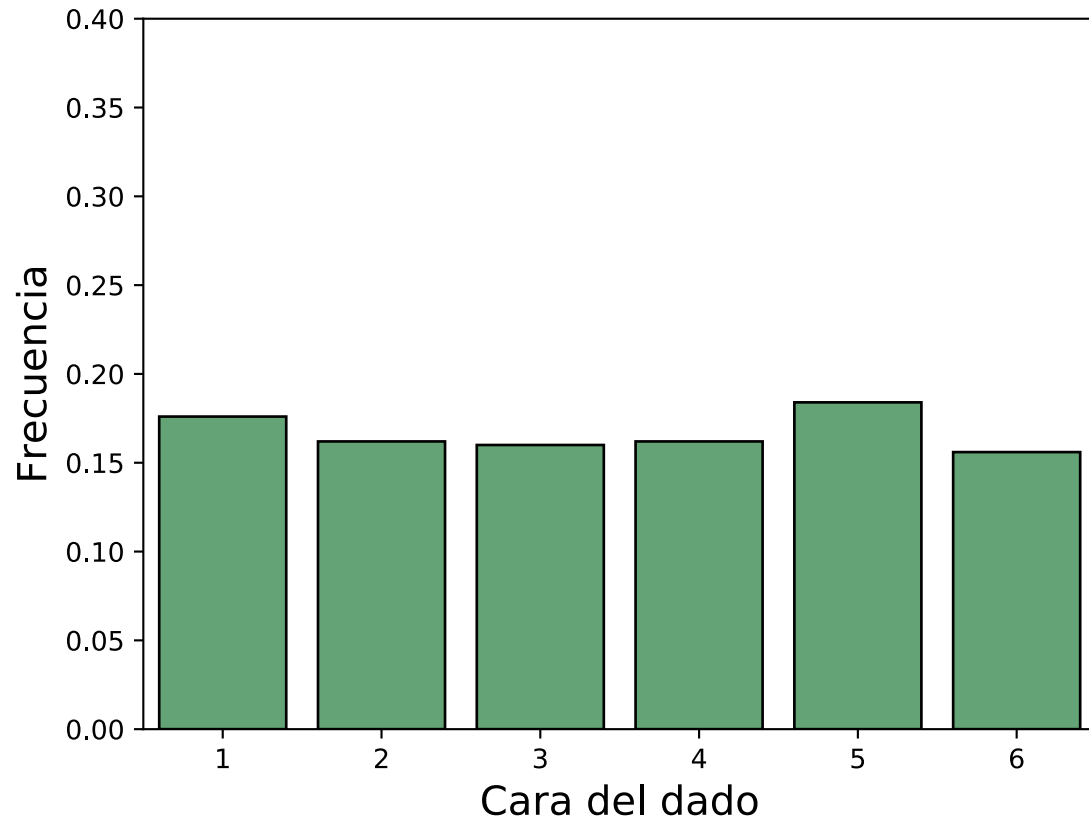
Distribución de probabilidad

N = 100



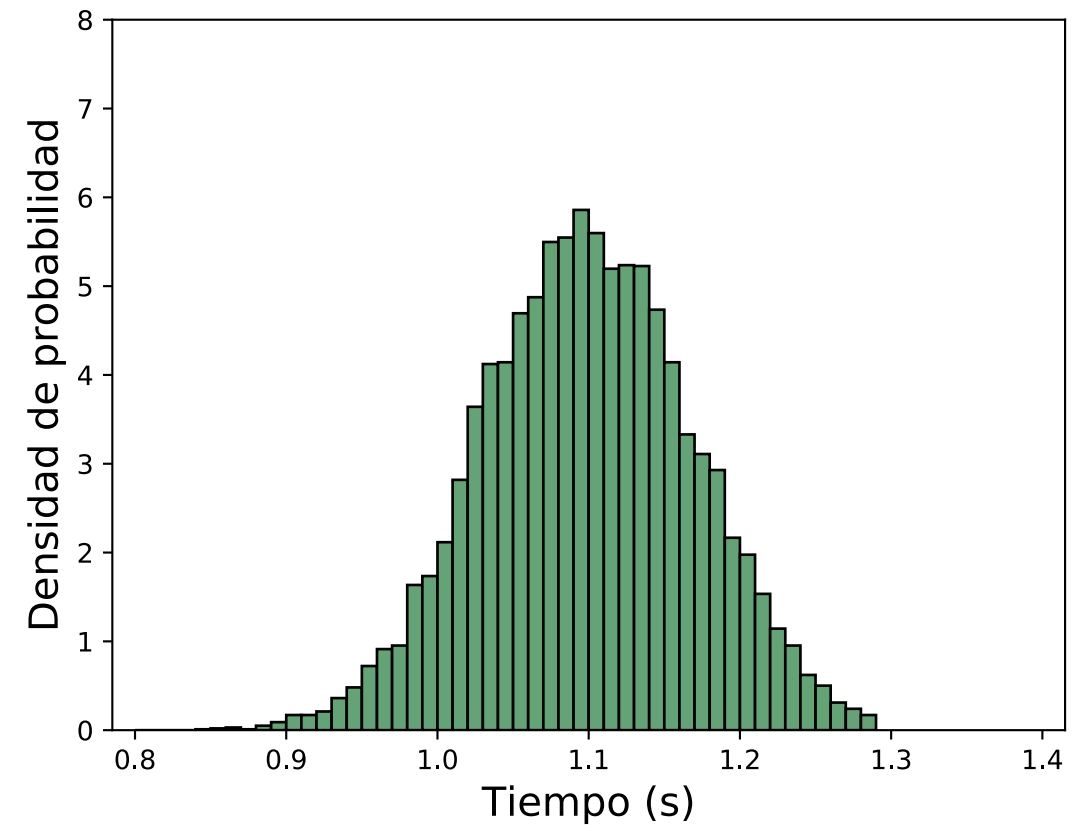
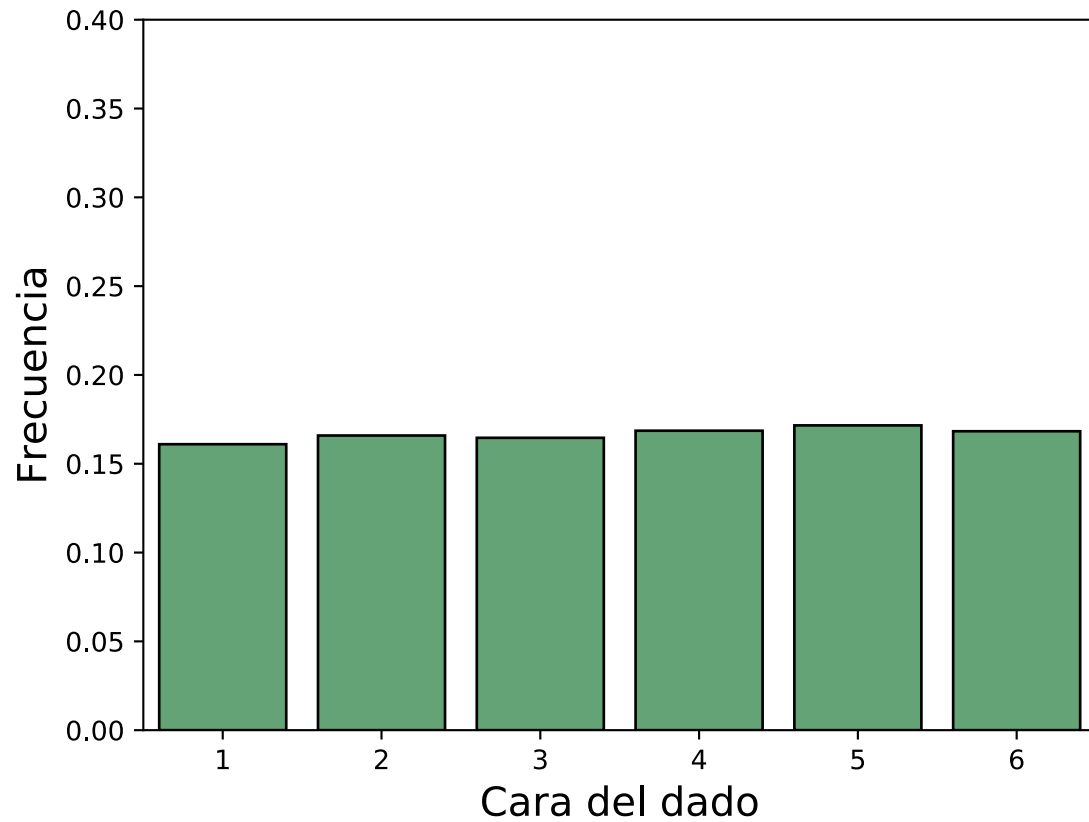
Distribución de probabilidad

N = 1000



Distribución de probabilidad

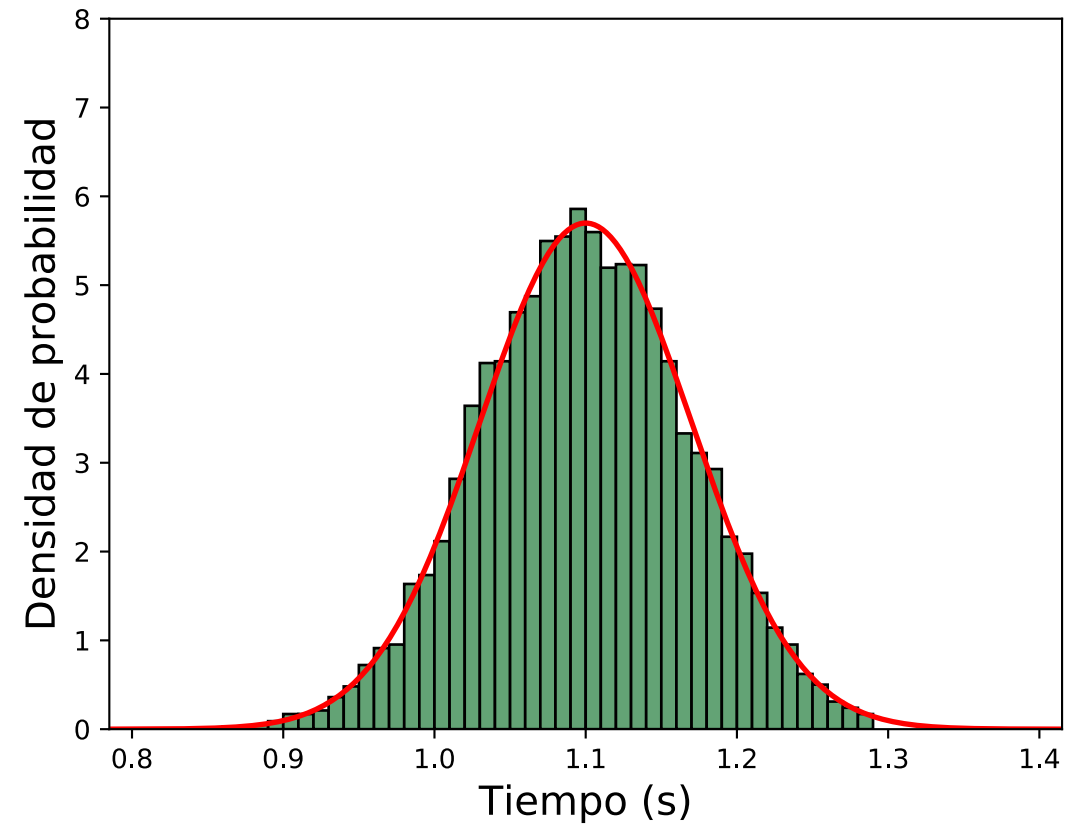
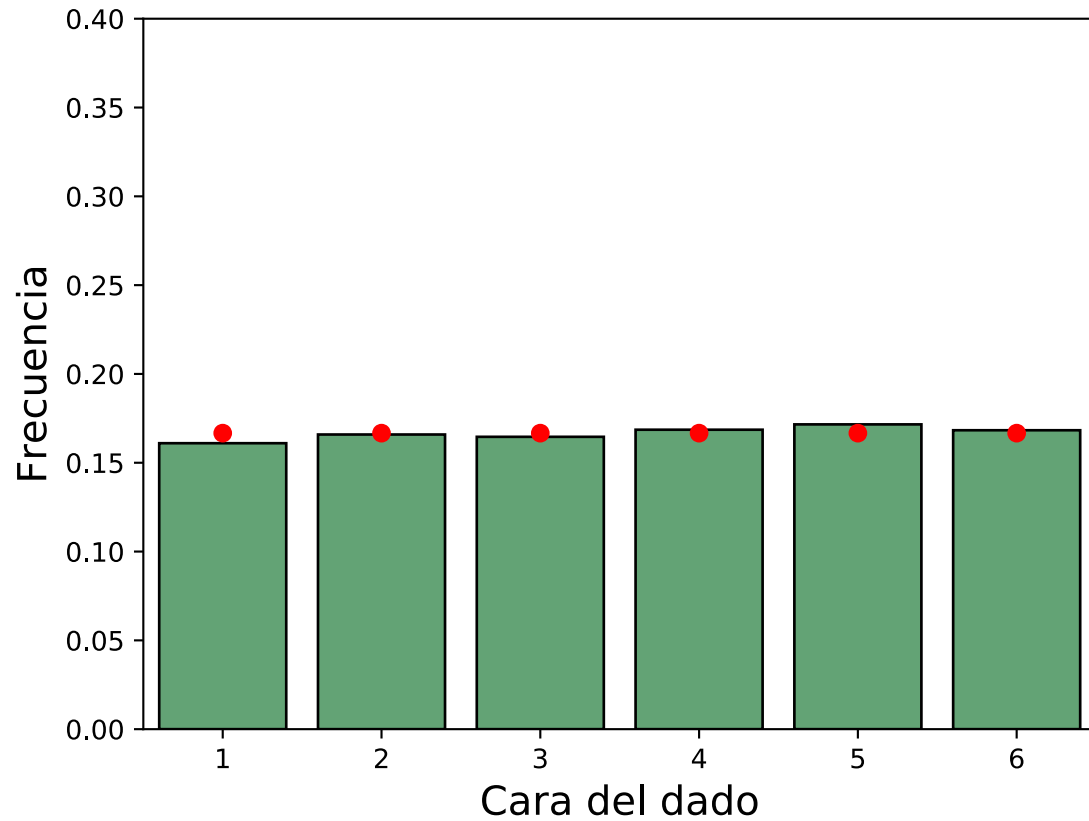
N = 10000



Distribución de probabilidad

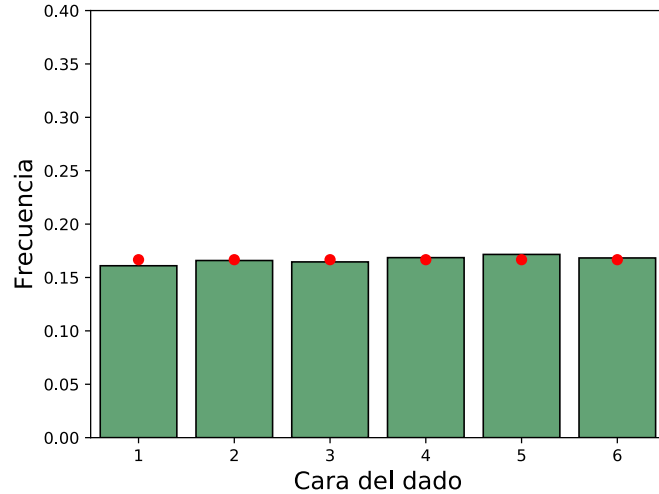
N = 10000

Distribuciones de probabilidad



Distribuciones de probabilidad discreta y continua

Discreto



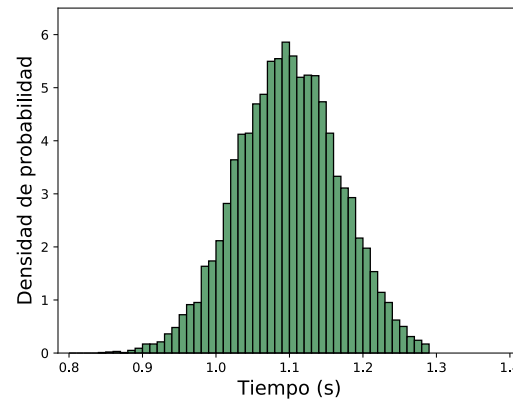
$$F_k \xrightarrow{N \rightarrow \infty} P_k$$

$$\sum_k P_k = 1$$

$$F_k = \frac{n_k}{N}$$

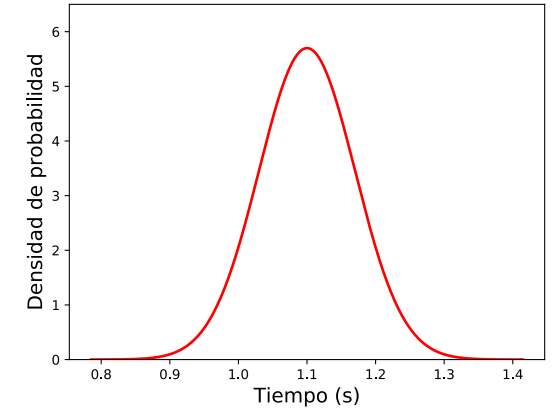
Condición de normalización

Continuo



$$N \rightarrow \infty$$

$$a \rightarrow 0$$



$$F_k \rightarrow f(x)dx$$

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

a : bin size

$$F_k = d_k a \quad \text{con} \quad d_k = \frac{n_k}{a N}$$

Valor medio, varianza y desviación estándar

Caso discreto

Caso continuo

Valor medio $E(X) = \frac{1}{N} \sum_k n_k x_k = \sum_k x_k F_k$

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx$$

Varianza $VAR(X) = \frac{1}{N} \sum_k n_k (x_k - \bar{X})^2 = \sum_k (x_k - \bar{X})^2 F_k$

$$VAR(X) = \int_{-\infty}^{+\infty} (x - \bar{X})^2 f(x) dx$$

Desviación estándar $SD(X) = \sqrt{VAR(X)}$

Notaciones

$$E(X) \equiv \langle X \rangle \equiv \bar{X}$$

$$VAR(X) \equiv \sigma^2$$

$$SD(X) \equiv \sigma$$

X: variable aleatoria (ej: resultado de una medición)

x_k : resultado k-ésimo

x es análogo a x_k

x: resultado continuo

F_k : frecuencia del resultado x_k $f(x)dx$ es análogo a F_k

$f(x)$: densidad de probabilidad del resultado x

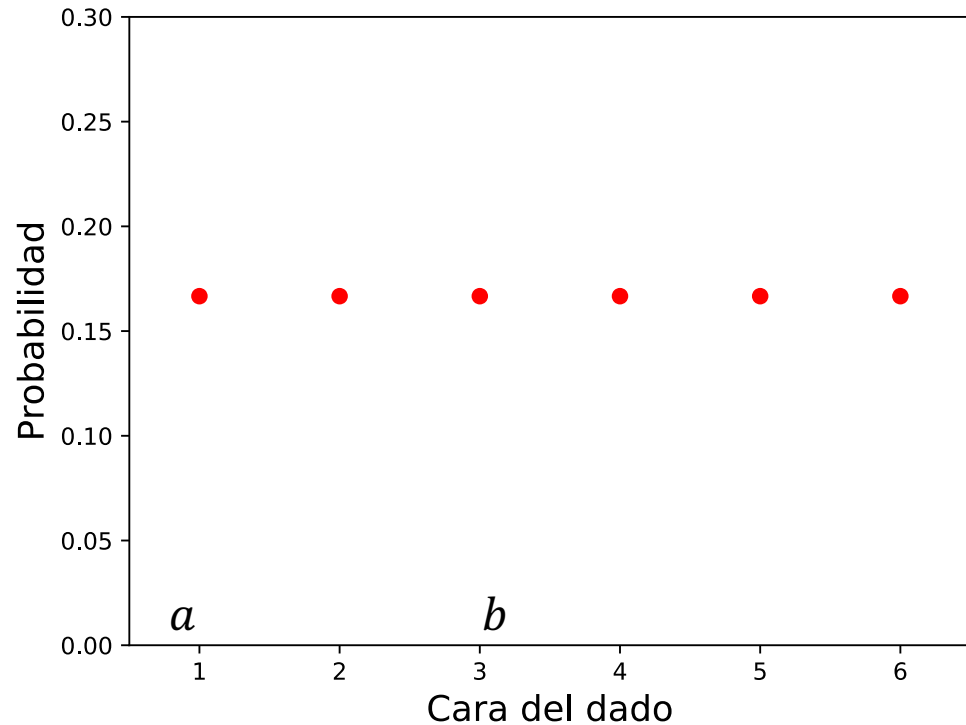
N: número total de resultados (ej: mediciones)

n_k : número de veces que se obtuvo x_k

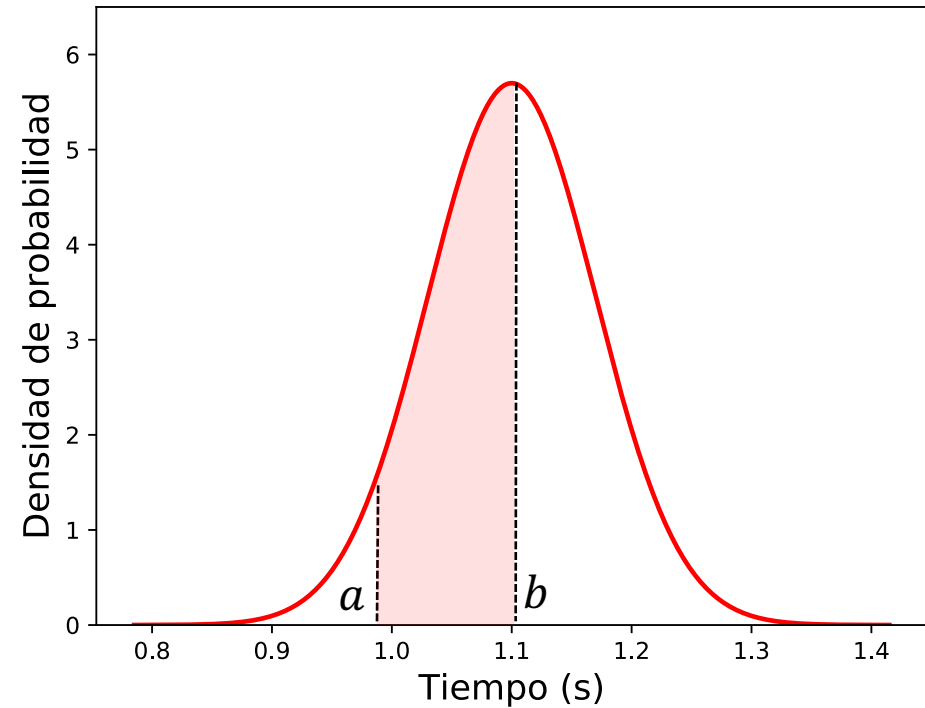
Cálculo de probabilidad en un cierto intervalo

¿Cuál es la probabilidad de que una medición esté comprendida entre a y b ?

¿Tiene sentido preguntarse cuál es la probabilidad de medir un cierto *número real*?



$$Prob(a \leq X \leq b) = \sum_{k=a}^b P_k$$



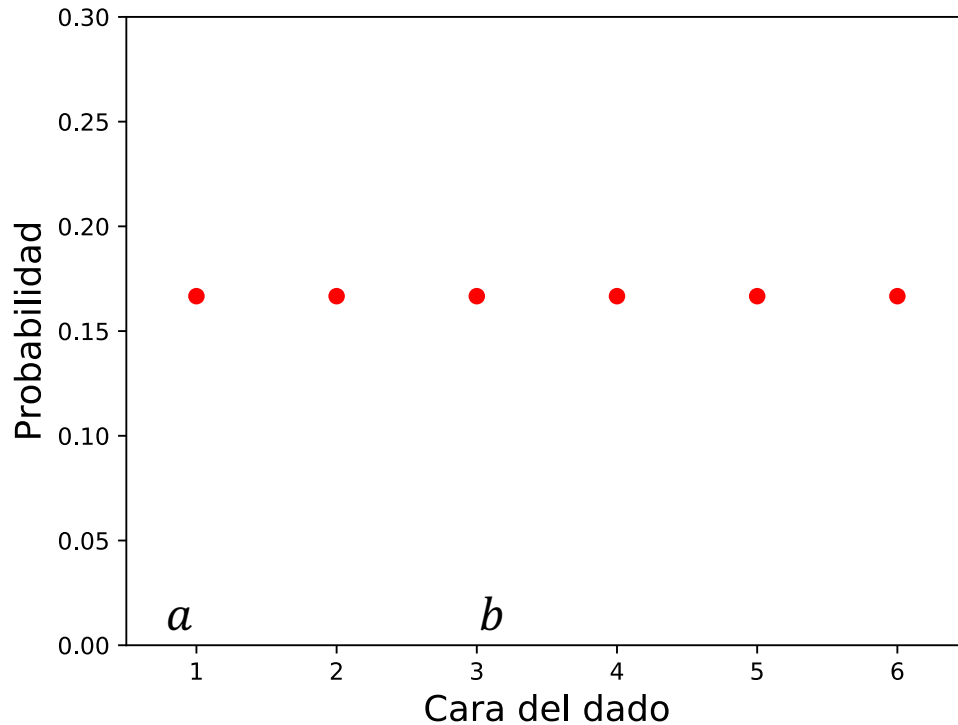
$$Prob(a \leq X \leq b) = \int_a^b f(x)dx$$

Cálculo de probabilidad en un cierto intervalo

¿Cuál es la probabilidad de que una medición esté comprendida entre a y b ?

¿Tiene sentido preguntarse cuál es la probabilidad de medir un cierto *número real*?

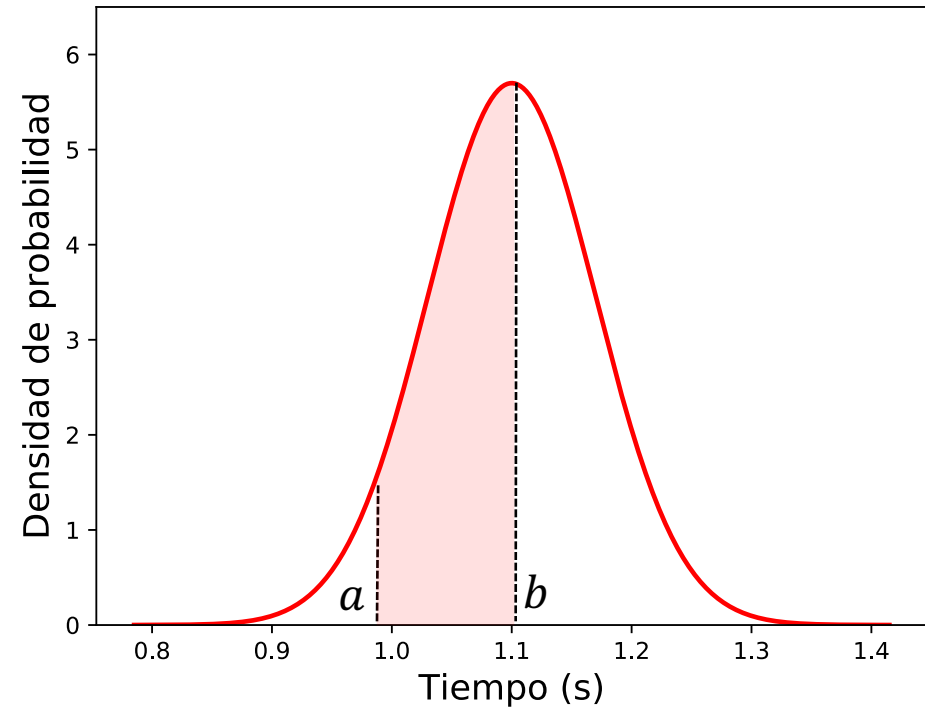
Ejemplo: ¿Cuál es la probabilidad que al tirar el dado salga 3 o menor que 3?



$$Prob(a \leq X \leq b) = \sum_{k=a}^b P_k$$

ej: $Prob(1 \leq X \leq 3) = \sum_{k=1}^3 P_k = 1/6 + 1/6 + 1/6 = 0,50$

Ejemplo: ¿Cuál es la probabilidad que al medir el período del faro el cronómetro indique entre 1,0 y 1,1 s?

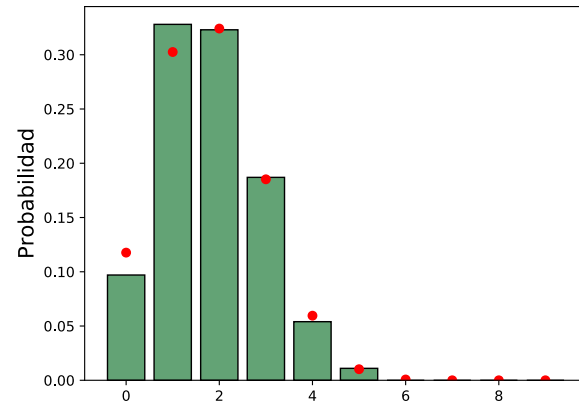


$$Prob(a \leq X \leq b) = \int_a^b f(x)dx$$

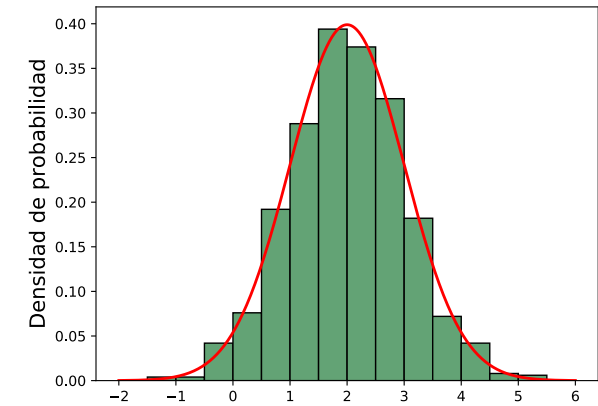
ej: $Prob(1,0 \leq X \leq 1,1) = \int_{1,0}^{1,1} f(x)dx \approx 0,46$

Ejemplos de distribuciones de probabilidad

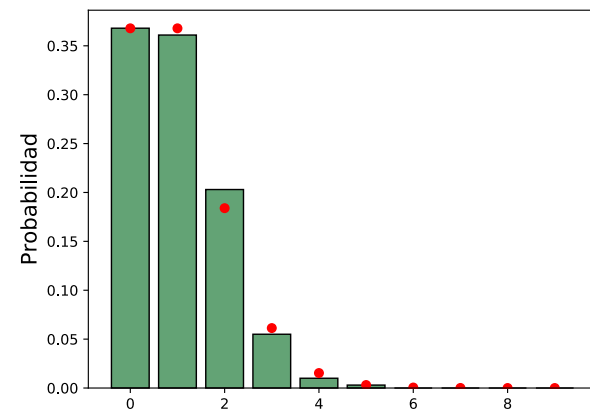
Binomial



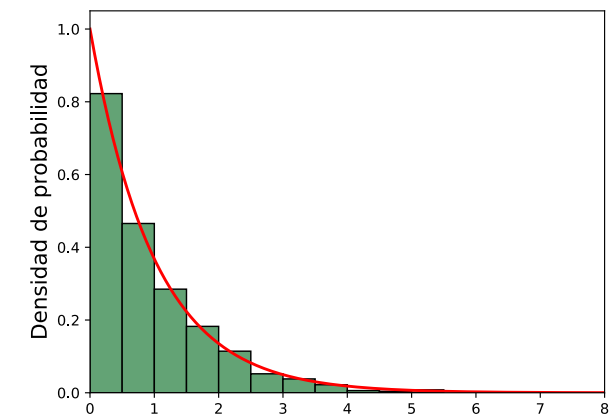
Gauss



Poisson

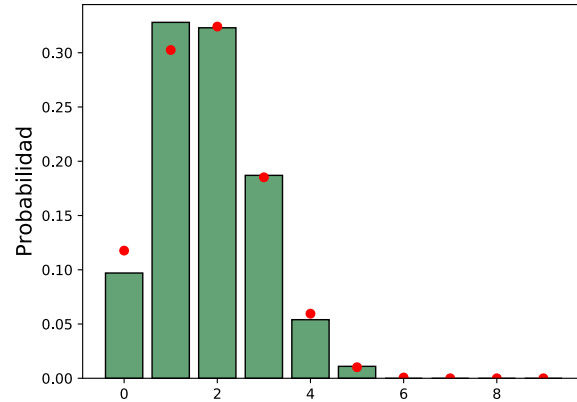


Exponencial

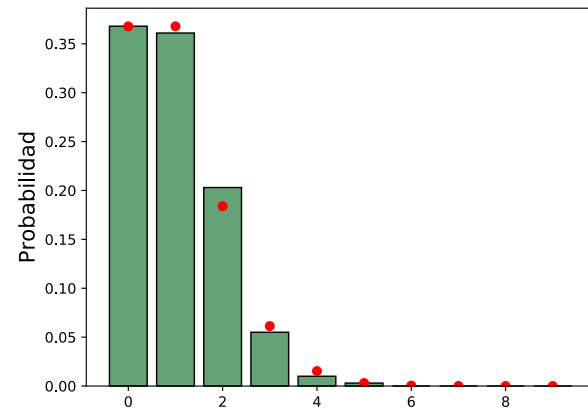


Ejemplos de distribuciones de probabilidad

Binomial

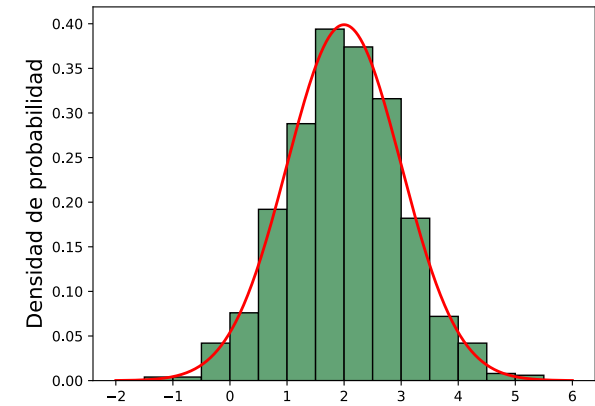


Poisson

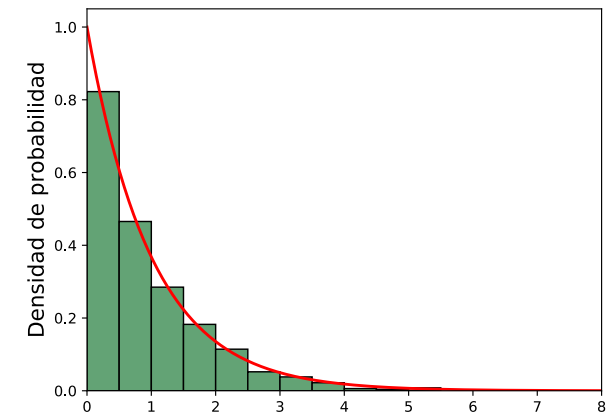


¡La de Gauss no es la única
distribución de probabilidad!

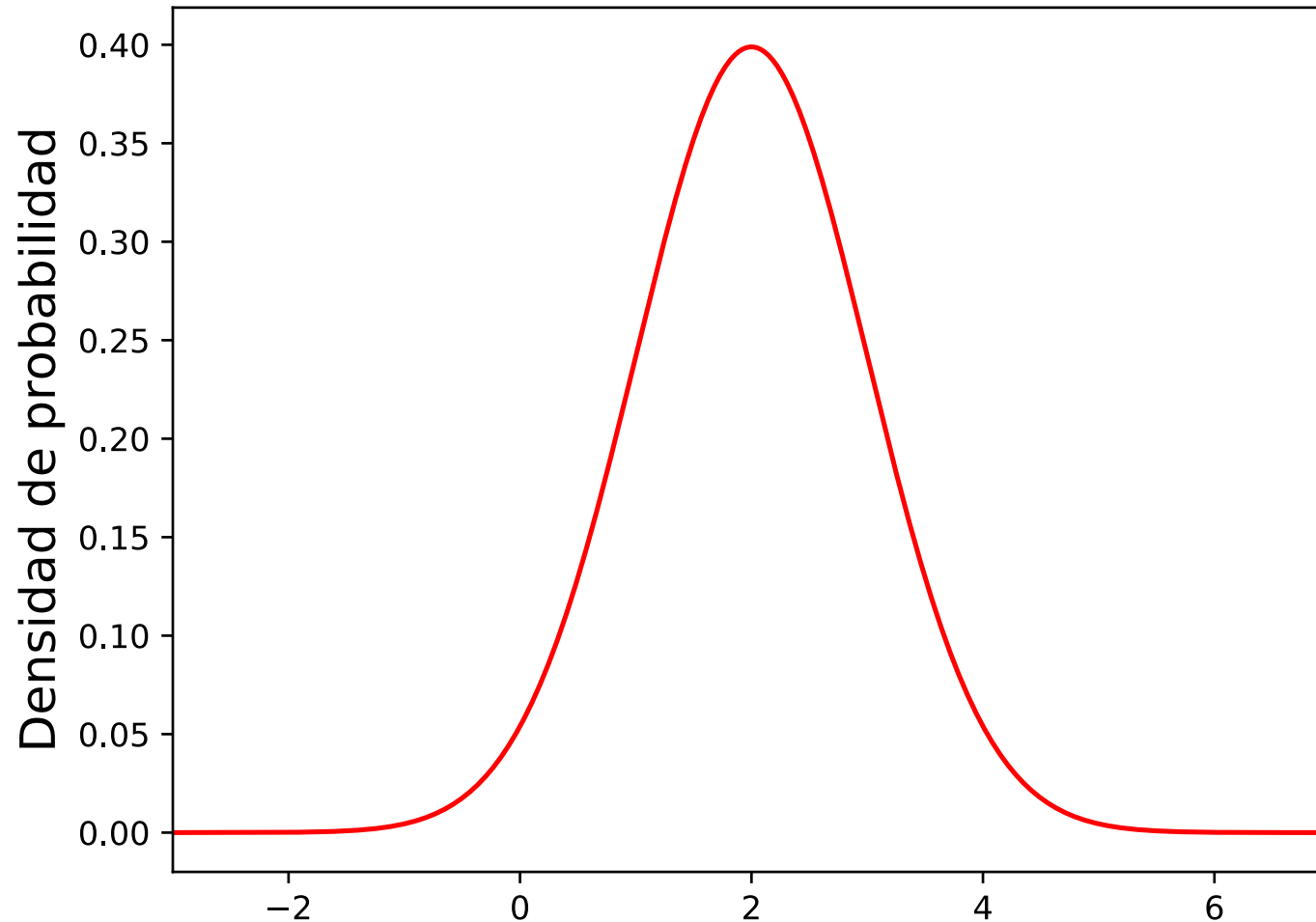
Gauss



Exponencial



Distribución de Gauss

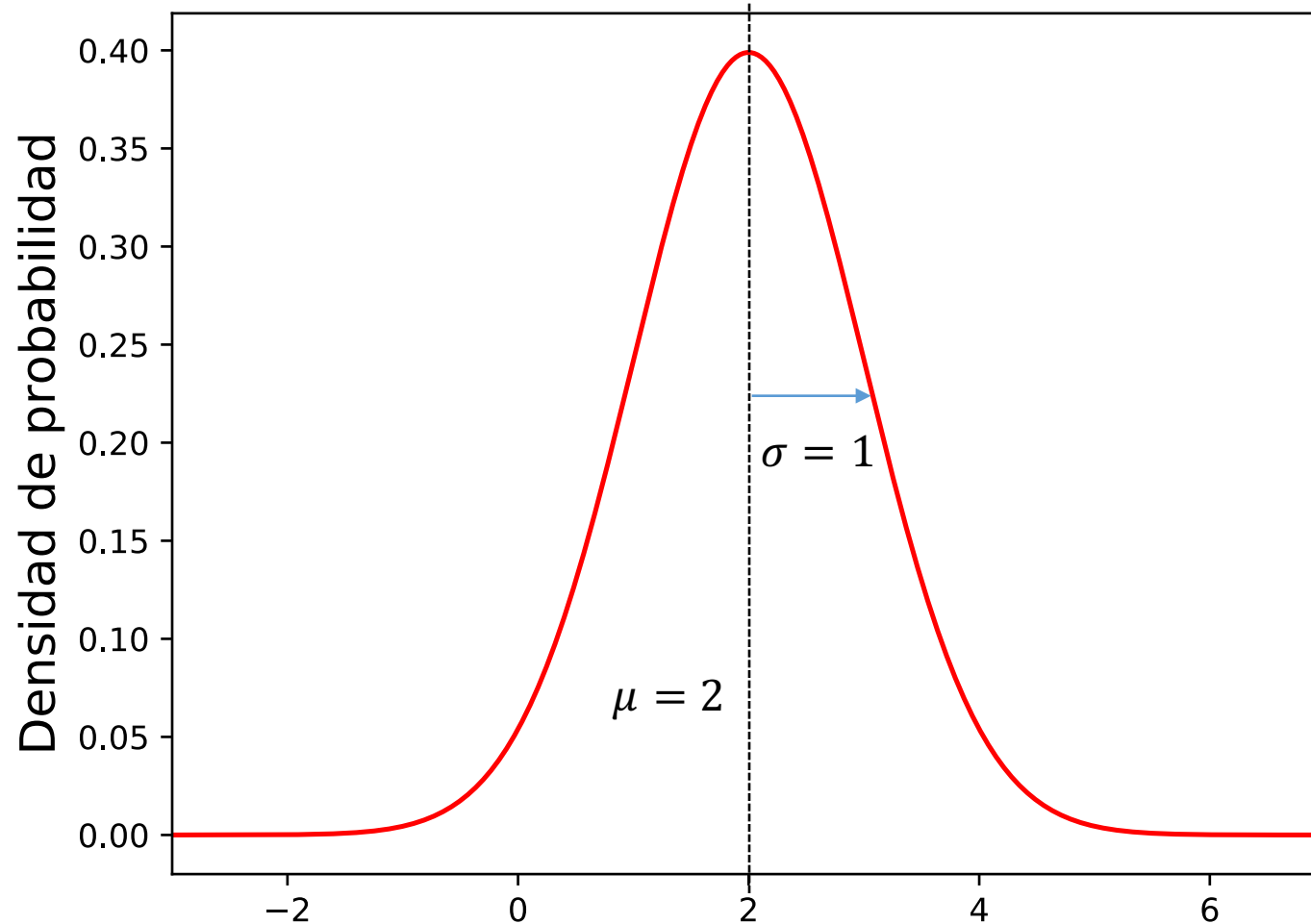


La distribución de Gauss es una buena *aproximación* para muchísimos casos*

$$G_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

* Ver Teorema central del límite

Distribución de Gauss

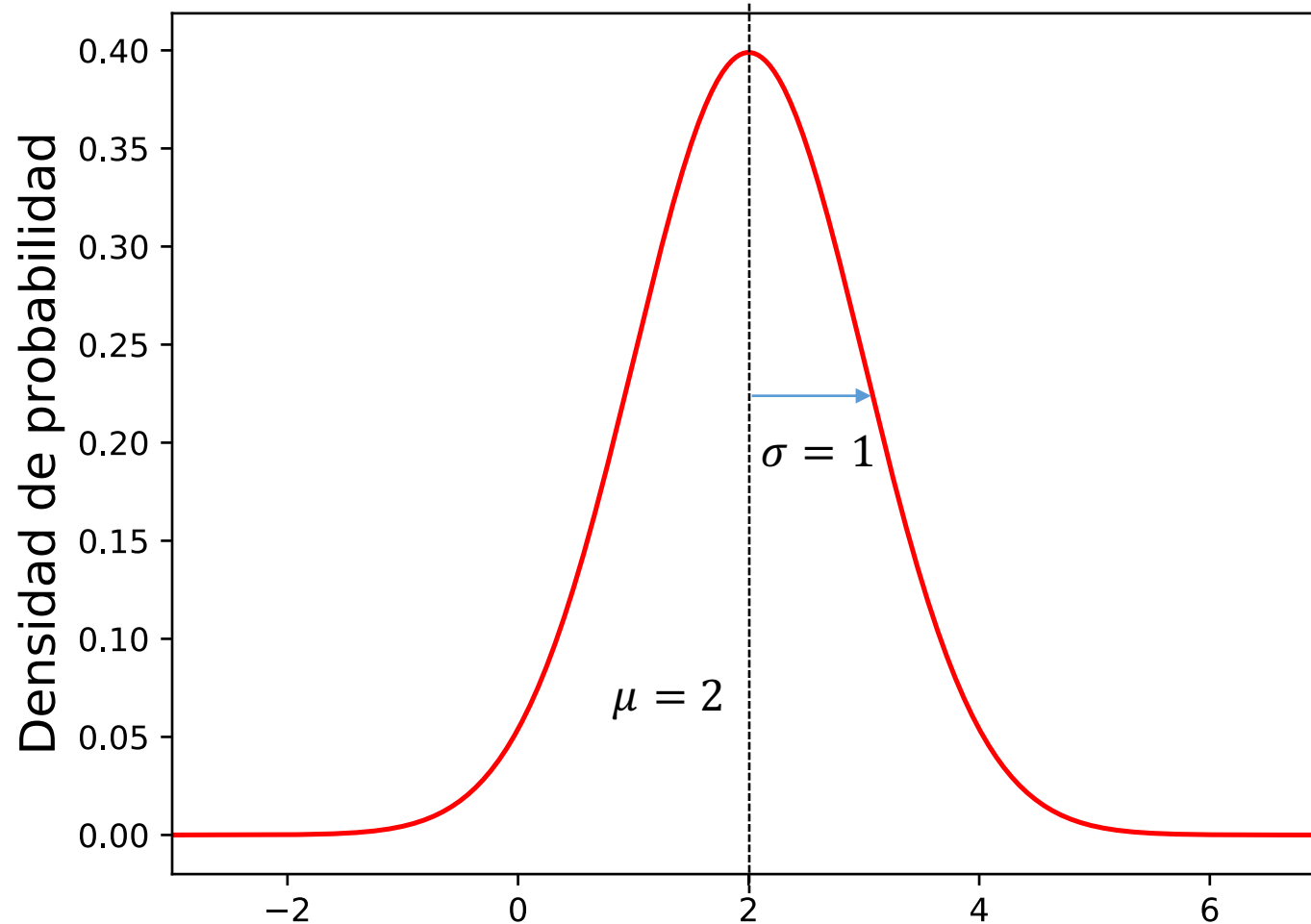


La distribución de Gauss es una buena *aproximación* para muchísimos casos*

$$G_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

* Ver Teorema central del límite

Distribución de Gauss



La distribución de Gauss es una buena *aproximación* para muchísimos casos*

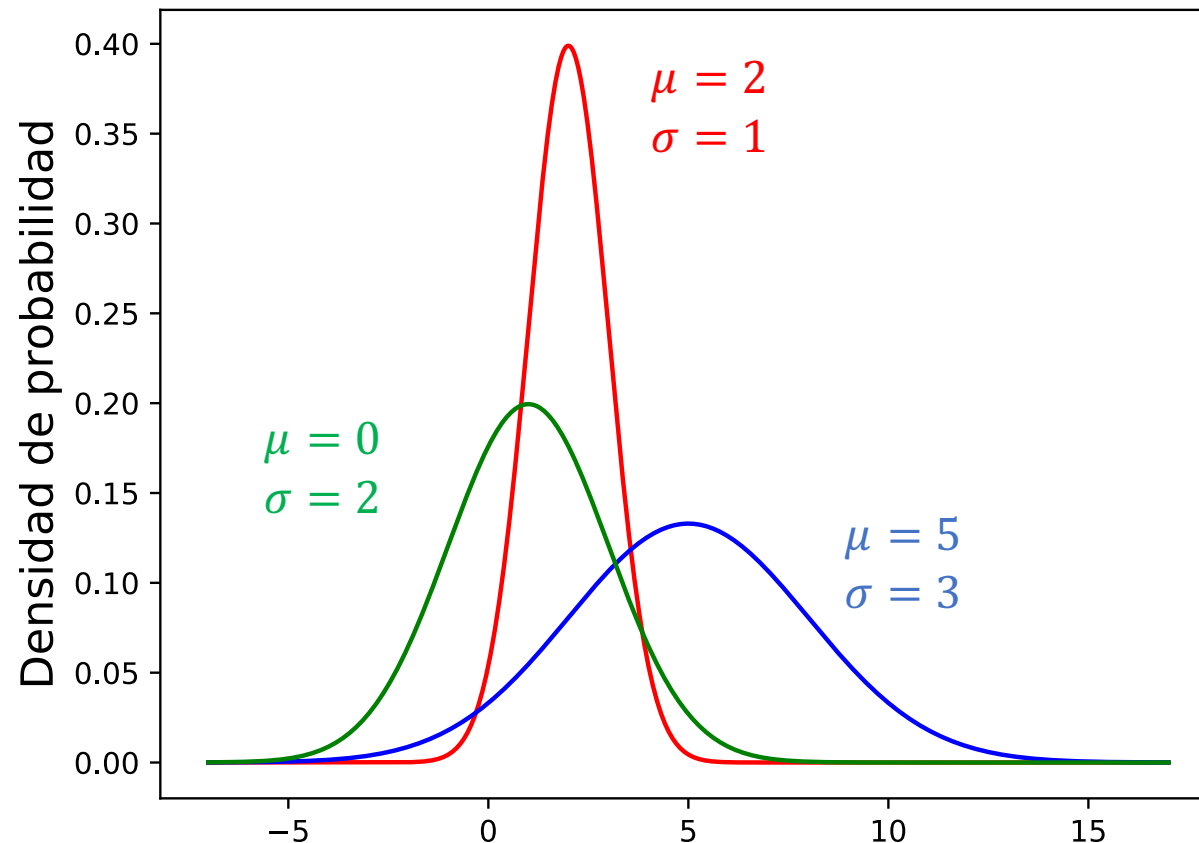
$$G_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$\int_{-\infty}^{+\infty} G_{\mu,\sigma}(x) dx = 1$$

* Ver Teorema central del límite

Distribución de Gauss: algunas propiedades

- Está centrada en $x = \mu$
- Es simétrica alrededor de $x = \mu$
- Tiende exponencialmente a 0 para $|x - \mu| \gg \sigma$
- El parámetro σ da una medida de su ancho



$$G_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

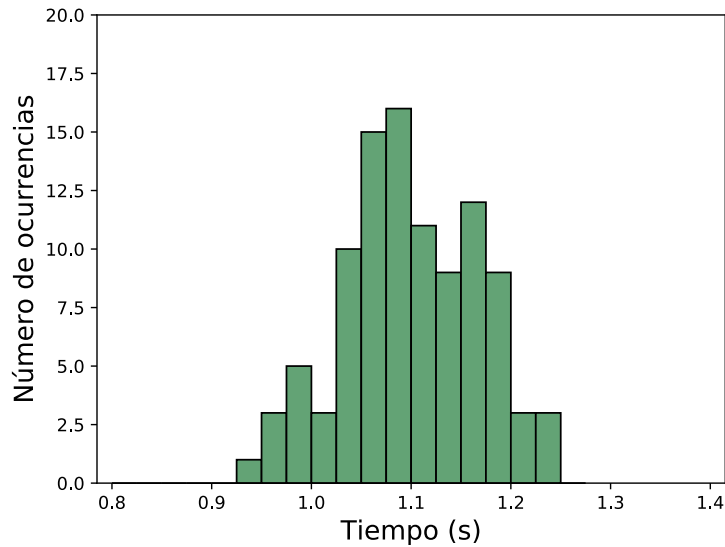
$$\int_{-\infty}^{+\infty} G_{\mu,\sigma}(x) dx = 1$$

Estadística

por ejemplo μ y σ de la gaussiana

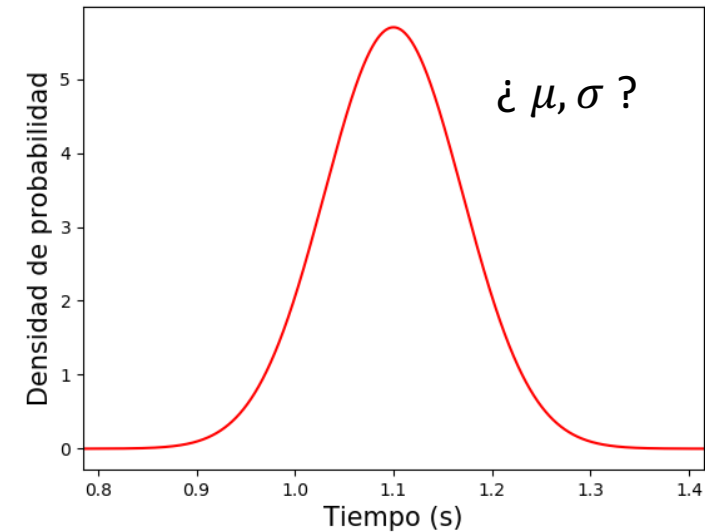
Objetivo: *estimar* los parámetros de la distribución de probabilidad de la variable aleatoria a partir de los datos.

Datos obtenidos



¿De qué distribución de probabilidad provienen mis datos?

¿

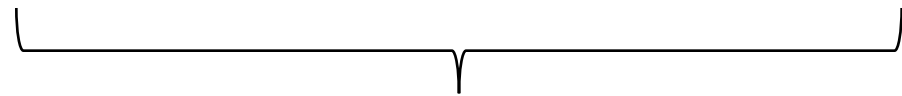


?

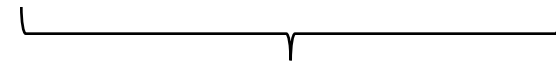
Estimación de los parámetros de la distribución de Gauss

$$\bar{X}_n = \frac{1}{n} \sum_{i=1} x_i \quad \longrightarrow \quad E(X) = \mu$$

$$S_n^2 = \frac{1}{n} \sum_i (x_i - \bar{X}_n)^2 \quad \longrightarrow \quad VAR(X) = \sigma^2$$



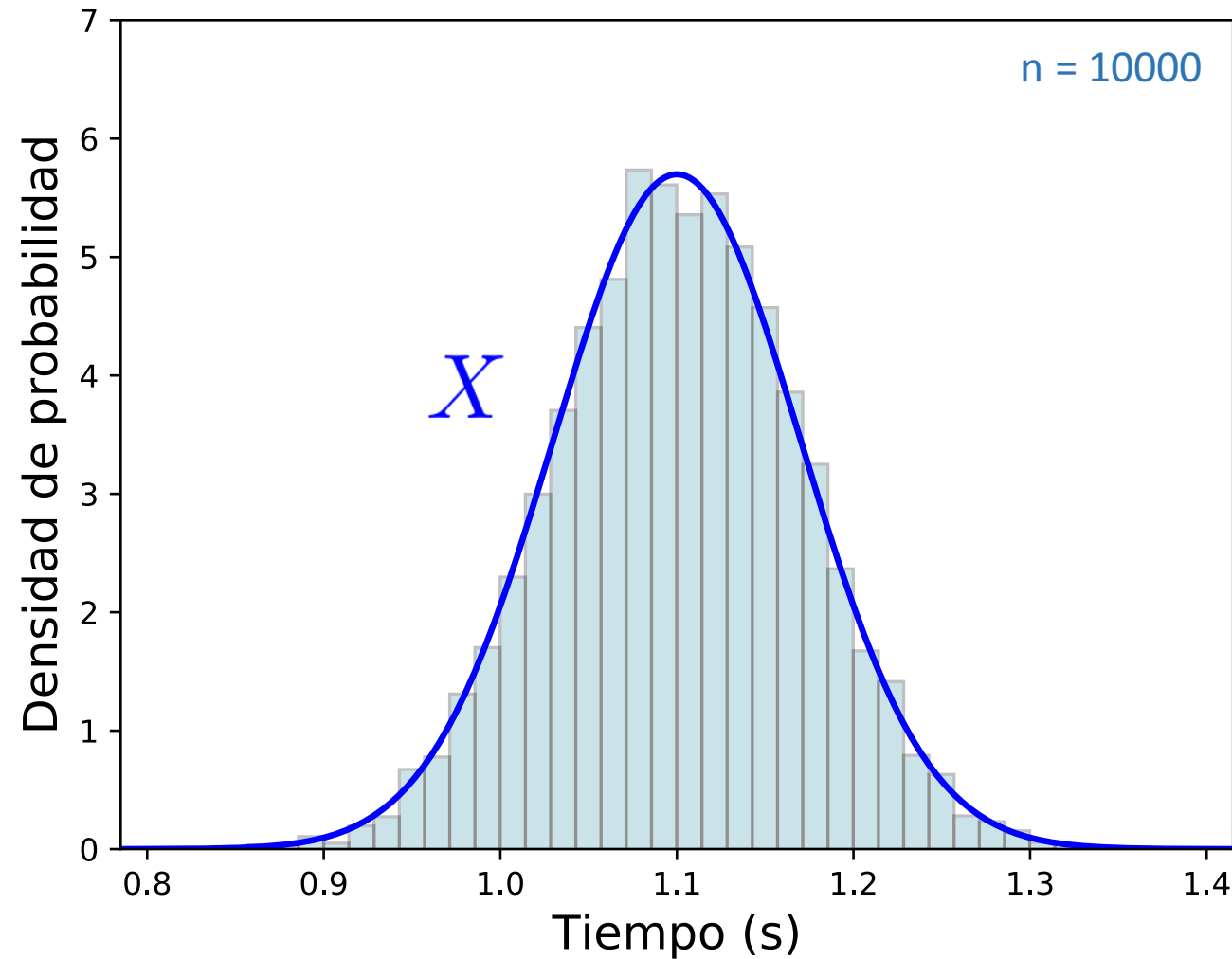
Son *estimadores*: funciones de los datos medidos x_i



Parámetros desconocidos de la distribución de Gauss

n: cantidad de mediciones de la variable aleatoria X realizadas

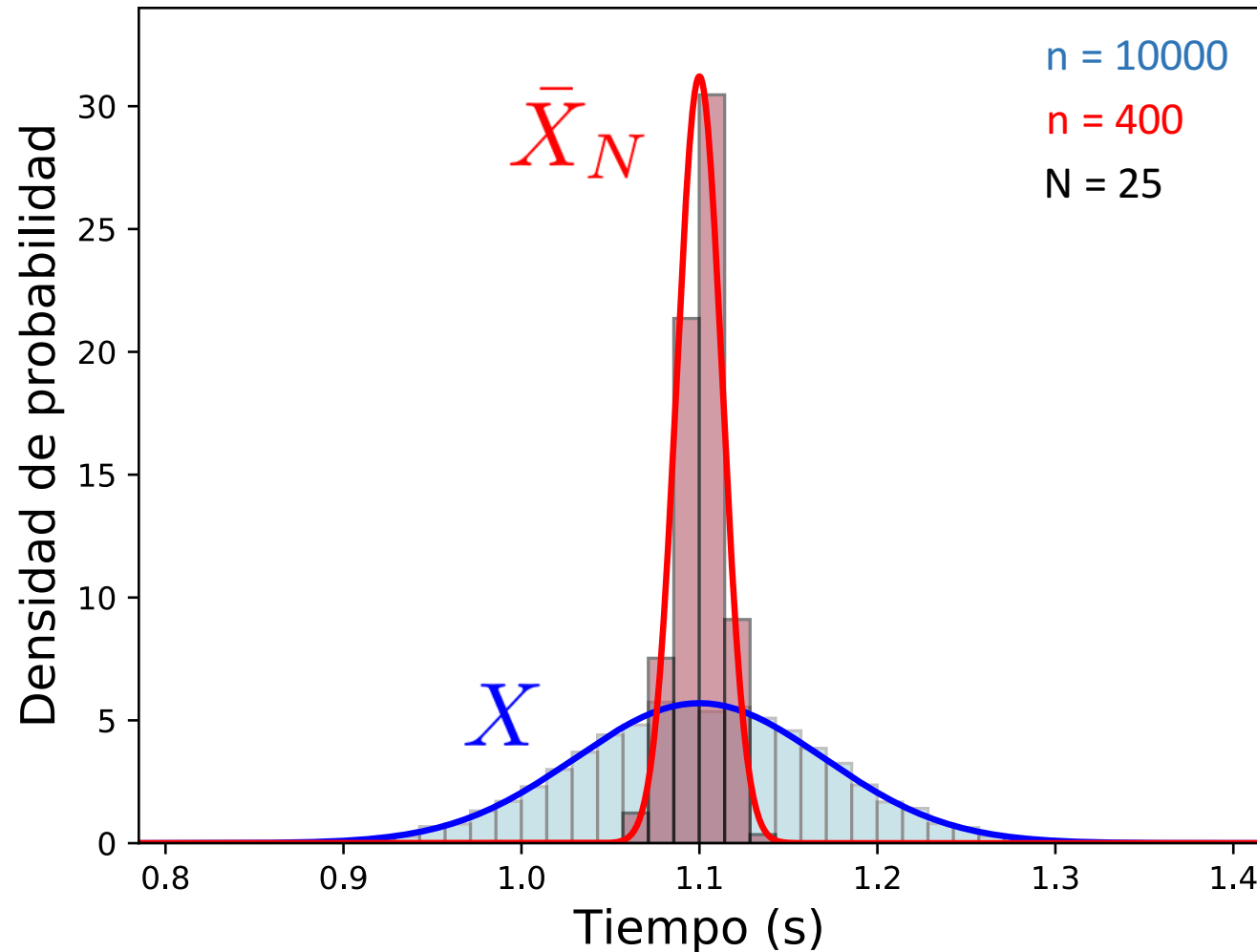
Distribución de la variable aleatoria “una medición”



$$\langle X \rangle = 1,100 \text{ s}$$

$$\sigma_X = 0,070 \text{ s}$$

Distribución de la variable aleatoria “promedio de N mediciones”



$$\langle X \rangle = 1,100 \text{ s} \quad \text{promedio de } X$$

$$\sigma_X = 0,070 \text{ s} \quad \text{desv. estándar de } X$$

$$\langle \bar{X}_N \rangle = 1,100 \text{ s} \quad \text{promedio de } \bar{X}_N$$

$$\sigma_{\bar{X}_N} = 0,014 \text{ s} \quad \text{desv. estándar de } \bar{X}_N$$

$$\sigma_{\bar{X}_N} = \frac{\sigma_X}{\sqrt{N}}^*$$

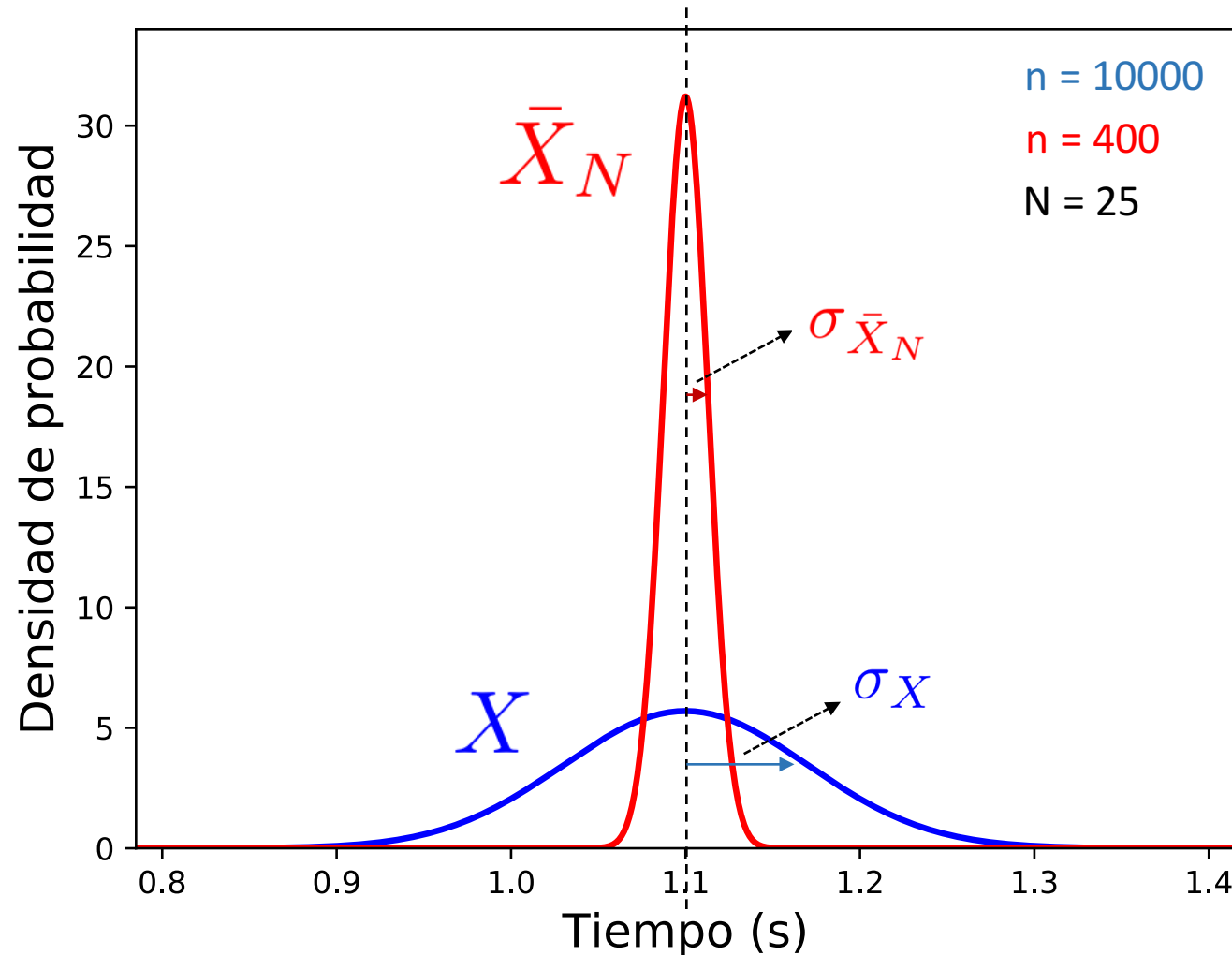
X : variable aleatoria “una medición”

$\bar{X}_N$: variable aleatoria “promedio de N mediciones”

n : número de cuentas en el histograma

* La demostración de este resultado general está en el apéndice

Distribución de la variable aleatoria “promedio de N mediciones”



$$\langle X \rangle = 1,100 \text{ s} \quad \text{promedio de } X$$

$$\sigma_X = 0,070 \text{ s} \quad \text{desv. estándar de } X$$

$$\langle \bar{X}_N \rangle = 1,100 \text{ s} \quad \text{promedio de } \bar{X}_N$$

$$\sigma_{\bar{X}_N} = 0,014 \text{ s} \quad \text{desv. estándar de } \bar{X}_N$$

$$\sigma_{\bar{X}_N} = \frac{\sigma_X}{\sqrt{N}}^*$$

X : variable aleatoria “una medición”

$\bar{X}_N$: variable aleatoria “promedio de N mediciones”

n : número de cuentas en el histograma

* La demostración de este resultado general está en el apéndice

Error estadístico y error instrumental

Llamaremos *error estadístico* a la desviación estándar de la media de N mediciones

$$e_{stat} \equiv \sigma_{\bar{X}_N}$$

- El error estadístico disminuye a medida que se realizan más mediciones con dependencia $\propto \frac{1}{\sqrt{N}}$
- El error instrumental *no depende* de la cantidad de mediciones realizadas

El error total de una medición es la suma en cuadratura* de todas las contribuciones de error *independientes*. Si volvemos al ejemplo de la medición del período del faro considerando por ahora solamente el error instrumental (por ej. la precisión para medir tiempo del cronómetro) y el error estadístico (la variabilidad en la medición por parte del experimentador)

$$e_{tot}^2 = e_{stat}^2 + e_{inst}^2$$

¿Cuántas veces es necesario medir para que e_{stat}^2 sea despreciable respecto a e_{inst}^2 ?

* Este resultado es un resultado de la estadística y proviene de suponer que las contribuciones son independientes entre sí. Ver por ejemplo Baird D C, *Experimentación*, Prentice-Hall Hispanoamérica (1991), pp 46-50

Resultado de una medición e intervalo de confianza

¿Cuál es la probabilidad de que una medición (o un promedio de N mediciones) esté entre $\mu - \sigma$ y $\mu + \sigma$?

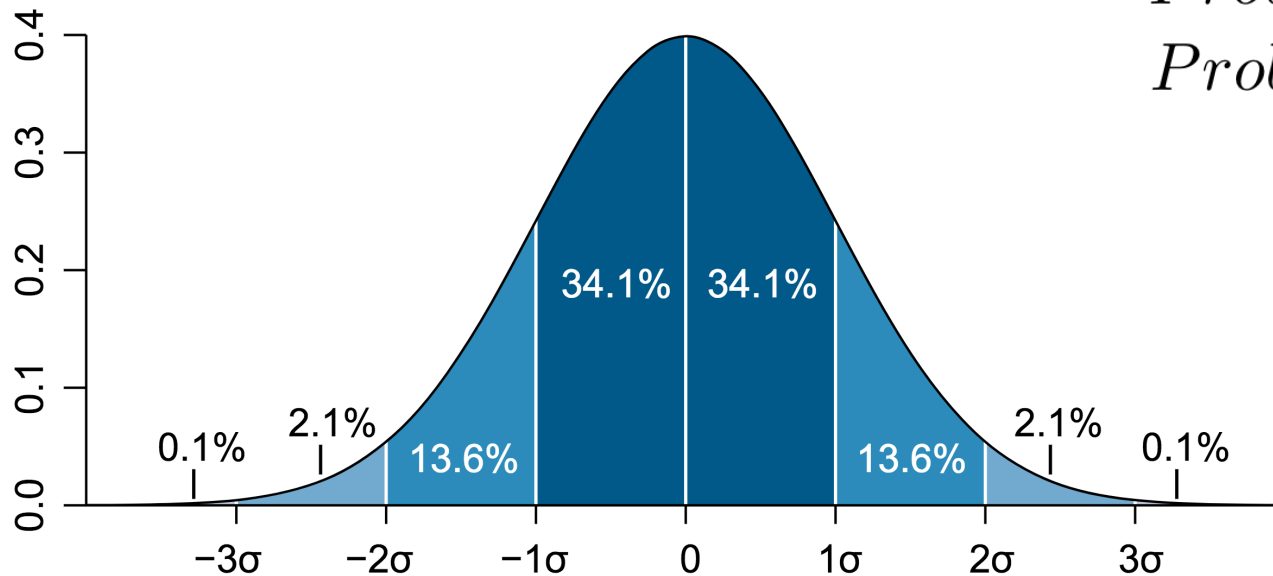
¿Y entre $\mu - t\sigma$ y $\mu + t\sigma$?

$$Prob(\mu - \sigma \leq X \leq \mu + \sigma) \approx 0,6827$$

$$Prob(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 0,9545$$

$$Prob(\mu - 3\sigma \leq X \leq \mu + 3\sigma) \approx 0,9974$$

Nivel de confianza



La notación

resultado = valor $\pm$ error

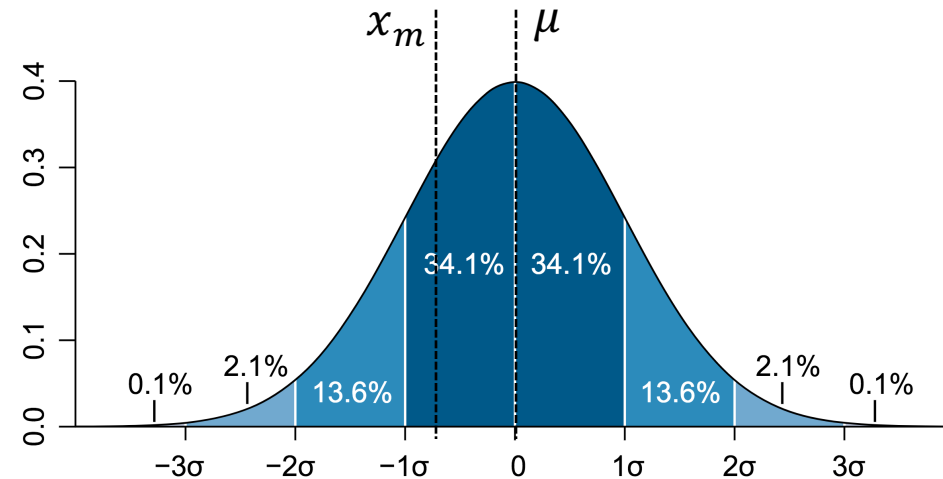
Significa *error* = σ

(nivel de confianza del **68,3%**)

Resultado de una medición e intervalo de confianza

resultado = valor $\pm$ error

ej: T = (1,10 $\pm$ 0,07) s



Dos interpretaciones:

$$Prob(\mu - \sigma \leq X \leq \mu + \sigma) \approx 0,6827 \rightarrow$$

La probabilidad de que la *próxima medición* se encuentre entre $\mu - \sigma$ y $\mu + \sigma$ es del 68,3%

=

$$Prob(X - \sigma \leq \mu \leq X + \sigma) \approx 0,6827 \rightarrow$$

La probabilidad de que el *valor real* μ se encuentre entre $x_m - \sigma$ y $x_m + \sigma$ es del 68,3%

x_m : valor medido

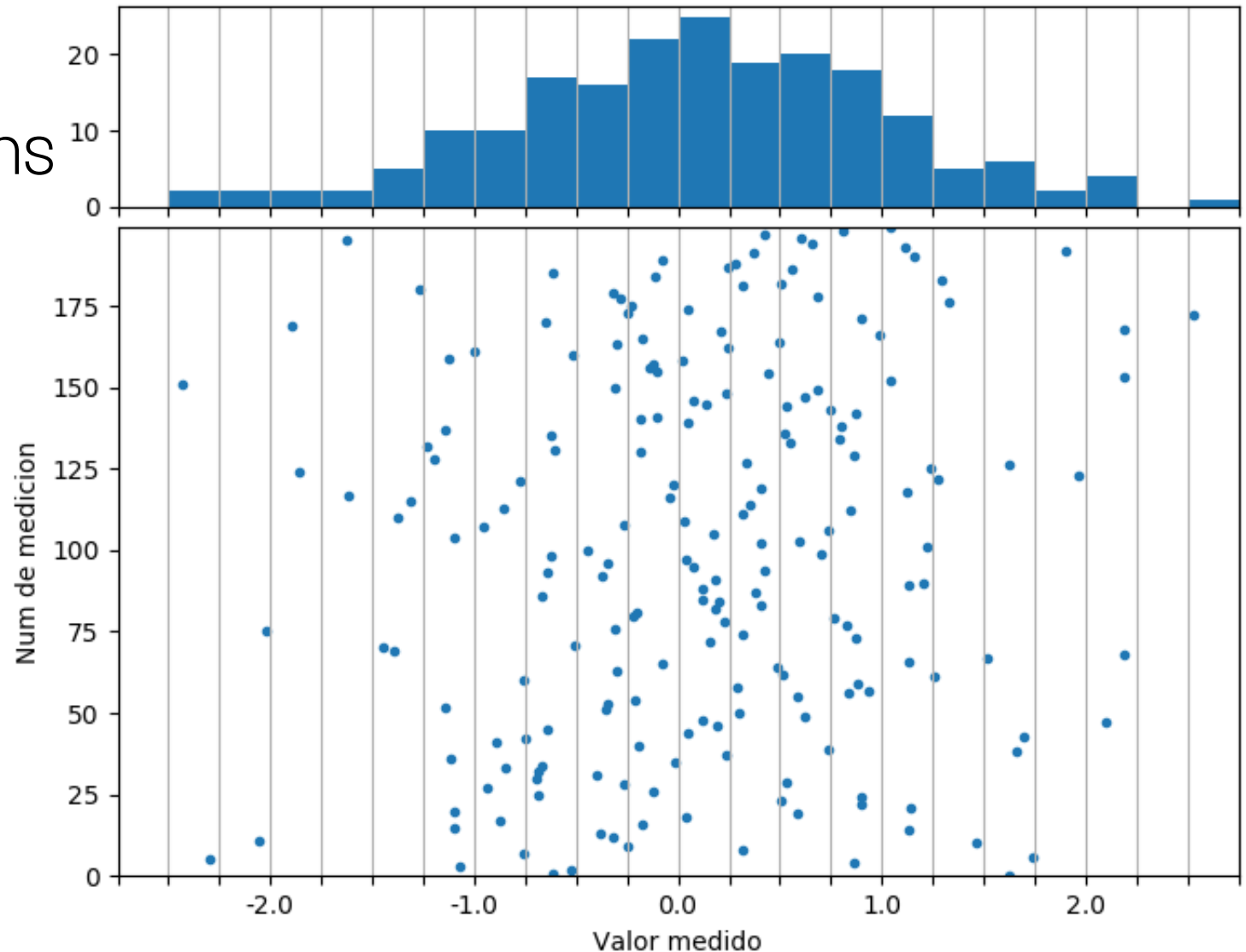
Histogramas

→ **Relevar variable aleatoria**

→ bin size

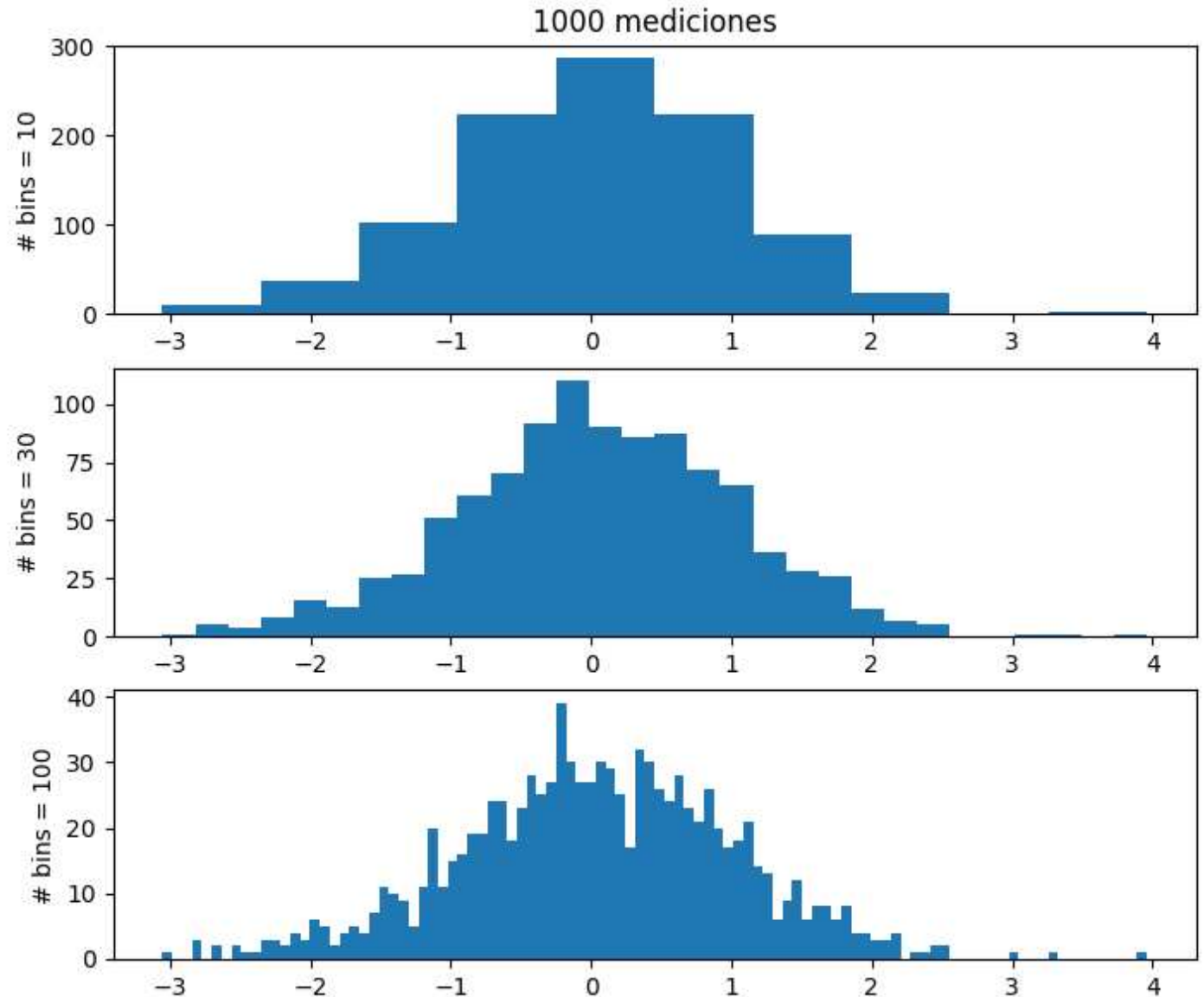
→ Límites

→ Numero de bins



Histogramas

$$\text{bin size} = 3.49 \cdot \sigma \cdot N^{-1/3} = 3.49 \cdot \frac{\sigma}{\sqrt[3]{N}}$$



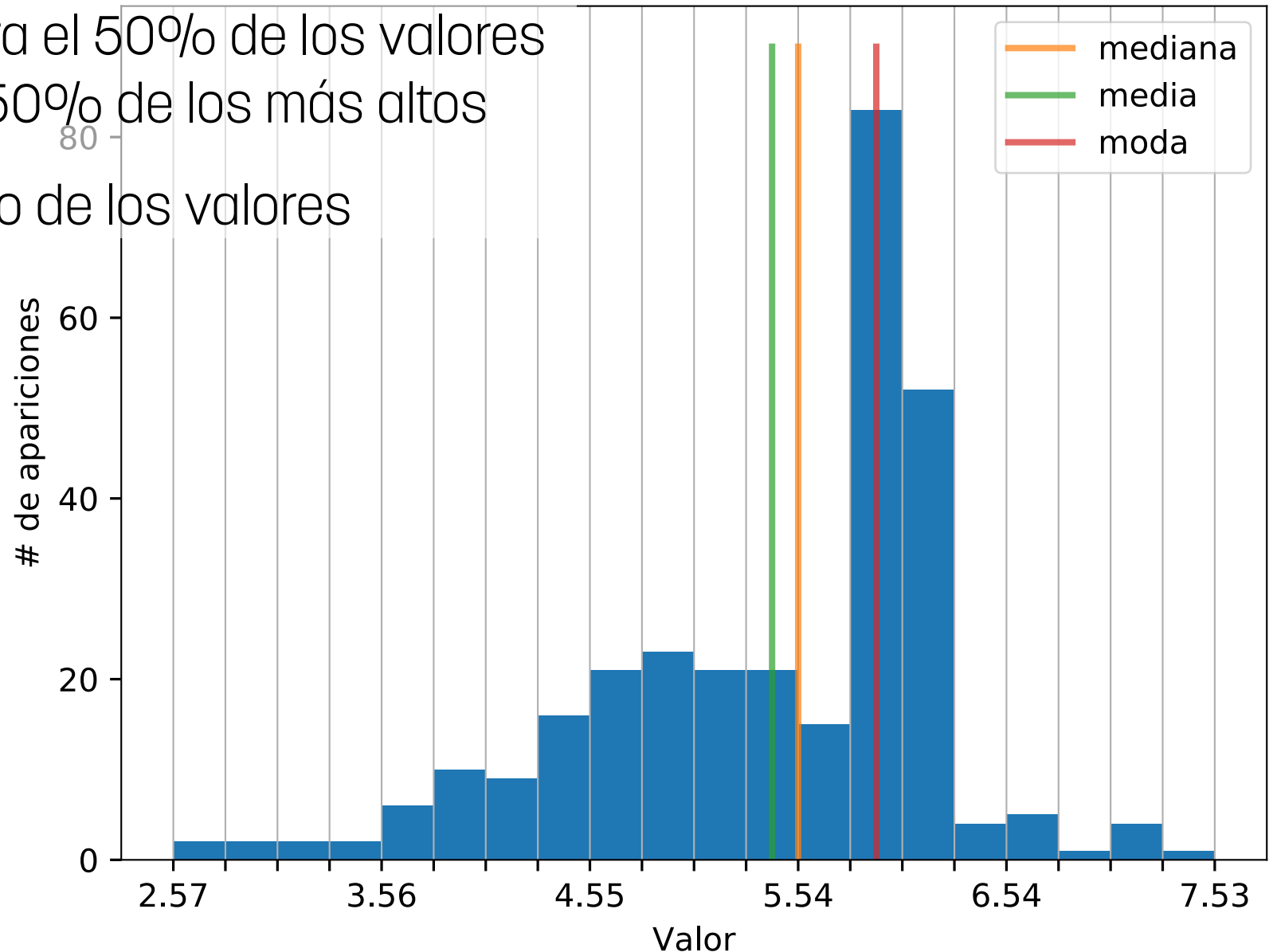
*On optimal and Data
based histograms,*
D. W. Scott,
Biometrika,
1979

Estadística

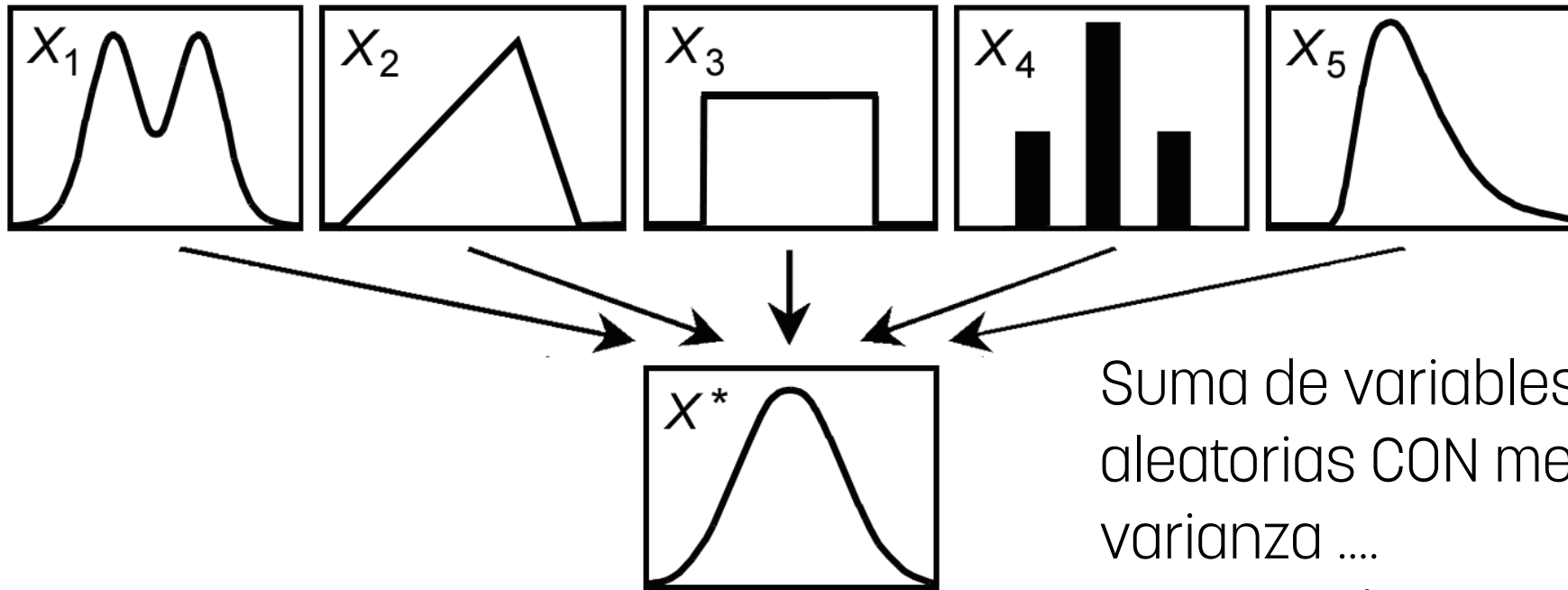
Moda: Valor más frecuente

Mediana: Separa el 50% de los valores más bajos del 50% de los más altos

Media: Promedio de los valores



Teorema central del límite



$$X^* = X_1 + X_2 + X_3 + X_4 + X_5$$

$$G_{\sigma, \mu}(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Sumario

→ **Desviación Estándar (de la muestra)**

→ Permite predecir probabilidad de hallar valores al medir

→ SD o σ

→ **Error Estándar**

→ $SD/\sqrt{N}$

→ Incerteza del valor medio μ o x_0

→ SE o σ_{x_0}

→ **Valor medio**

→ Estimación del valor real que se trata de medir

→ μ o x_0 o $\langle x \rangle$ o $\bar{x}$