Introducción a la estadística de distribuciones

Docentes del Laboratorio

Gustavo Grinblat

Jorge Alliende

Federico Petrovich

Laboratorio de Mecánica y Termodinámica

Cátedra: Prof. Ana Amador

Función de distribución y densidad de probabilidad

- Se toma una muestra de tamaño N → $x_1, x_2, x_3, \dots, x_N$

Función de distribución y densidad de probabilidad

- Se toma una muestra de tamaño N → $x_1, x_2, x_3, \dots, x_N$

	x_{10}	x_4	x_3	x_1	x_2	x_6	x_8	x_5	x_9	x_7
N = 10 $\longrightarrow$	8,08	8,17	8,22	8,26	8,27	8,31	8,33	8,37	8,43	8,51

Función de distribución y densidad de probabilidad

- Se toma una muestra de tamaño N → $x_1, x_2, x_3, \dots, x_N$

	x_{10}	x_4	x_3	x_1	x_2	x_6	x_8	x_5	x_9	x_7
N = 10 →	8,08	8,17	8,22	8,26	8,27	8,31	8,33	8,37	8,43	8,51

- Dividimos el intervalo (x_{min}, x_{max}) en *m* sub-intervalos iguales de ancho a.

Función de distribución y densidad de probabilidad

- Se toma una muestra de tamaño N → $x_1, x_2, x_3, \dots, x_N$

	x_{10}	x_4	x_3	x_1	x_2	x_6	x_8	x_5	x_9	x_7
N = 10 ⟶	8,08	8,17	8,22	8,26	8,27	8,31	8,33	8,37	8,43	8,51

m = 5; a = 0,1 | | | | | |

x_{min}= 8,05 8,15 8,25 8,35 8,45 8,55 = x_{max}

- Dividimos el intervalo (x_{min}, x_{max}) en m sub-intervalos iguales de ancho a.

Función de distribución y densidad de probabilidad

- Se toma una muestra de tamaño N → $x_1, x_2, x_3, \ldots, x_N$

	x_{10}	x_4	x_3	x_1	x_2	x_6	x_8	x_5	x_9	x_7
N = 10 ⟶	8,08	8,17	8,22	8,26	8,27	8,31	8,33	8,37	8,43	8,51

m = 5; a = 0,1

x_{min} = 8,05 | 8,15 | 8,25 | 8,35 | 8,45 | 8,55 = x_{max}

- Dividimos el intervalo (x_{min}, x_{max}) en m sub-intervalos iguales de ancho a.

- Denotamos como n_j al número de elementos contenidos en el j-ésimo intervalo.

Función de distribución y densidad de probabilidad

- Se toma una muestra de tamaño N → $x_1, x_2, x_3, \ldots, x_N$

	x_{10}	x_4	x_3	x_1	x_2	x_6	x_8	x_5	x_9	x_7
$N = 10 \longrightarrow$	8,08	8,17	8,22	8,26	8,27	8,31	8,33	8,37	8,43	8,51

$m = 5$; $a = 0,1$

$x_{min} = 8,05$ | 8,15 | 8,25 | 8,35 | 8,45 | $8,55 = x_{max}$

- Dividimos el intervalo (x_{min}, x_{max}) en m sub-intervalos iguales de ancho a.

- Denotamos como n_j al número de elementos contenidos en el j-ésimo intervalo.

$$n_1 = 1 \qquad n_2 = 2 \qquad n_3 = 4 \qquad n_4 = 2 \qquad n_5 = 1$$

Función de distribución y densidad de probabilidad

	x_{10}	x_4	x_3	x_1	x_2	x_6	x_8	x_5	x_9	x_7
$N = 10 \longrightarrow$	8,08	8,17	8,22	8,26	8,27	8,31	8,33	8,37	8,43	8,51

$m = 5$; $a = 0,1$

x_{min} = 8,05 | 8,15 | 8,25 | 8,35 | 8,45 | 8,55 = x_{max}

$n_1 = 1$ $n_2 = 2$ $n_3 = 4$ $n_4 = 2$ $n_5 = 1$

Función de distribución y densidad de probabilidad

$$\begin{array}{lcccccccccc} & x_{10} & x_4 & x_3 & x_1 & x_2 & x_6 & x_8 & x_5 & x_9 & x_7 \\ N = 10 \longrightarrow & 8{,}08 & 8{,}17 & 8{,}22 & 8{,}26 & 8{,}27 & 8{,}31 & 8{,}33 & 8{,}37 & 8{,}43 & 8{,}51 \end{array}$$

$m = 5$; $a = 0{,}1$

$x_{min} = 8{,}05$ | 8,15 | 8,25 | 8,35 | 8,45 | $8{,}55 = x_{max}$

$n_1 = 1$ $n_2 = 2$ $n_3 = 4$ $n_4 = 2$ $n_5 = 1$

- <u>Función de distribución</u>: $f_j = \dfrac{n_j}{N}$

Función de distribución y densidad de probabilidad

	x_{10}	x_4	x_3	x_1	x_2	x_6	x_8	x_5	x_9	x_7
N = 10 ⟶	8,08	8,17	8,22	8,26	8,27	8,31	8,33	8,37	8,43	8,51

m = 5; a = 0,1

x_{min}= 8,05 | 8,15 | 8,25 | 8,35 | 8,45 | 8,55 = x_{max}

$n_1 = 1$ $n_2 = 2$ $n_3 = 4$ $n_4 = 2$ $n_5 = 1$

- Función de distribución: $f_j = \frac{n_j}{N}$ $(\sum_{j=1}^{m} f_j = 1 \rightarrow$ la función está normalizada)

  f se denomina también frecuencia de ocurrencia.

Función de distribución y densidad de probabilidad

	x_{10}	x_4	x_3	x_1	x_2	x_6	x_8	x_5	x_9	x_7
N = 10 ⟶	8,08	8,17	8,22	8,26	8,27	8,31	8,33	8,37	8,43	8,51

m = 5; a = 0,1

x_{min} = 8,05 | 8,15 | 8,25 | 8,35 | 8,45 | 8,55 = x_{max}

$n_1 = 1$ $n_2 = 2$ $n_3 = 4$ $n_4 = 2$ $n_5 = 1$

- Función de distribución: $f_j = \frac{n_j}{N}$ $\left(\sum_{j=1}^{m} f_j = 1 \rightarrow \text{la función está normalizada}\right)$

  f se denomina también frecuencia de ocurrencia.

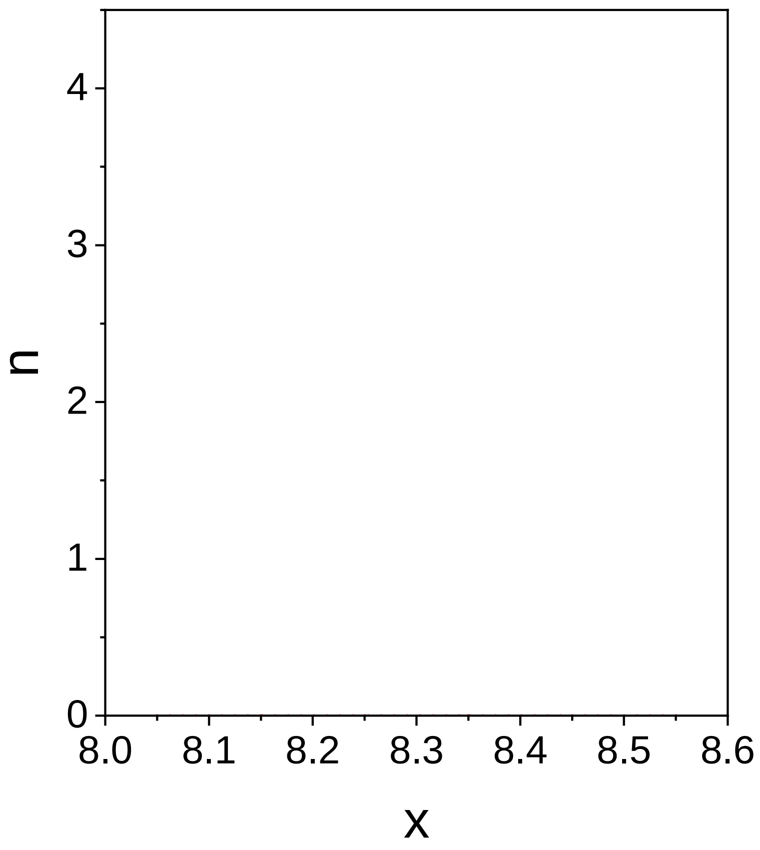
- Densidad de probabilidad: $\mathrm{d}_j = \frac{f_j}{a}$

Elaboración de un histograma

Se grafica f, n ó d vs. x, utilizando columnas centradas en $x_{min} + a\left(j - \frac{1}{2}\right)$

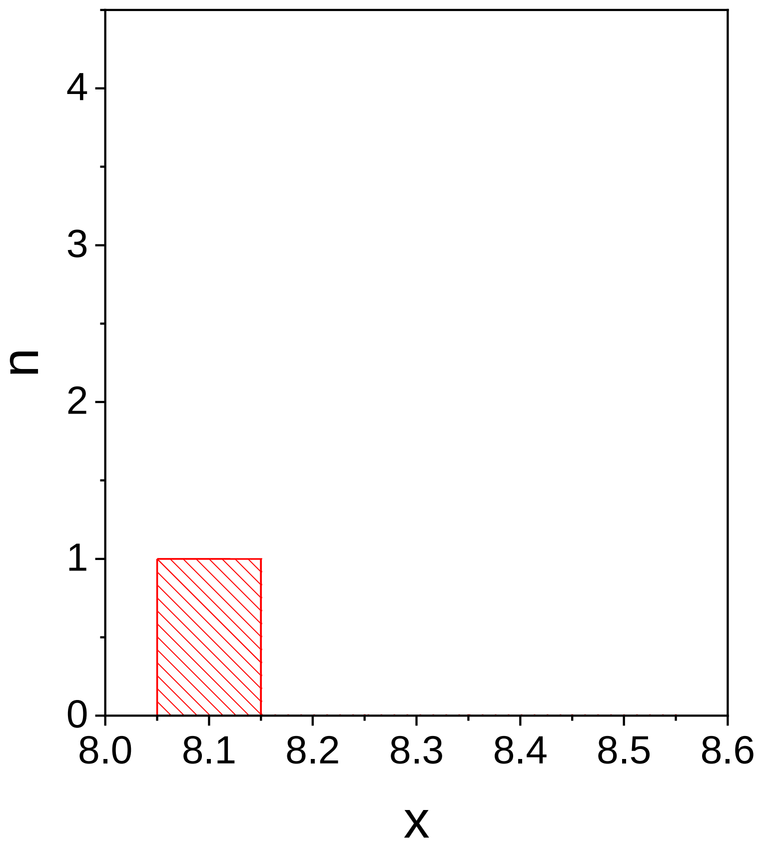
Elaboración de un histograma

Se grafica f, n ó d vs. x, utilizando columnas centradas en $x_{min} + a\left(j - \frac{1}{2}\right)$



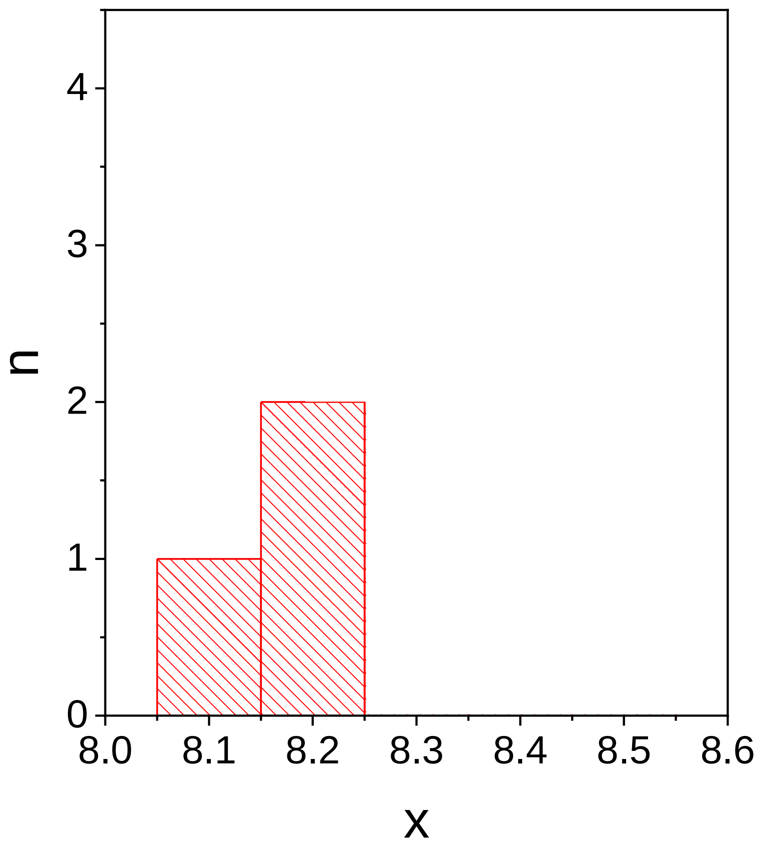
Elaboración de un histograma

Se grafica f, n ó d vs. x, utilizando columnas centradas en $x_{min} + a\left(j - \frac{1}{2}\right)$



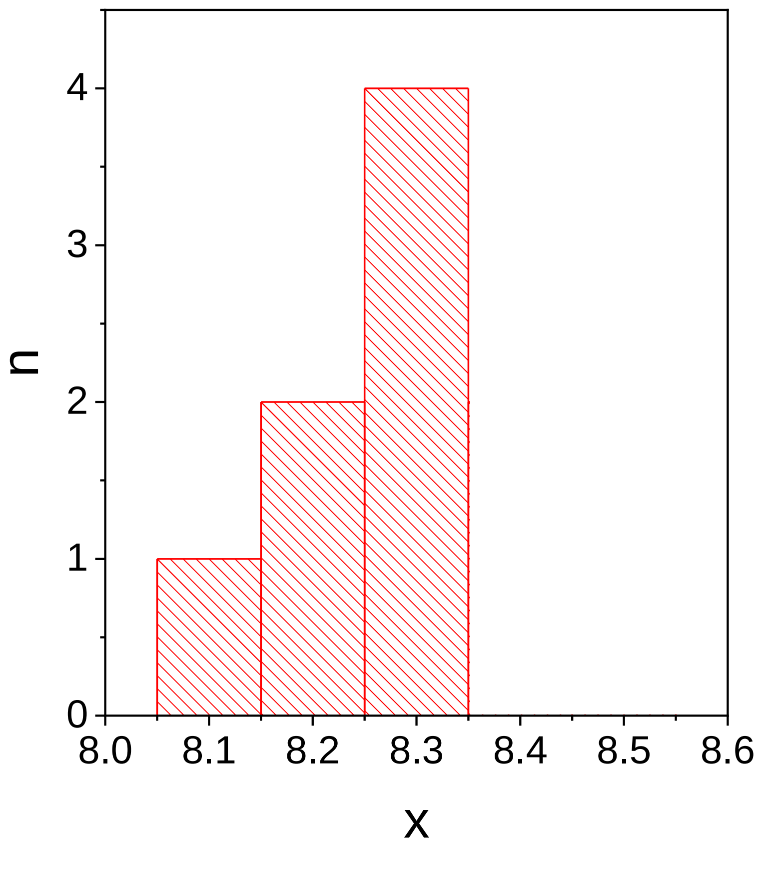
Elaboración de un histograma

Se grafica f, n ó d vs. x, utilizando columnas centradas en $x_{min} + a\left(j - \frac{1}{2}\right)$



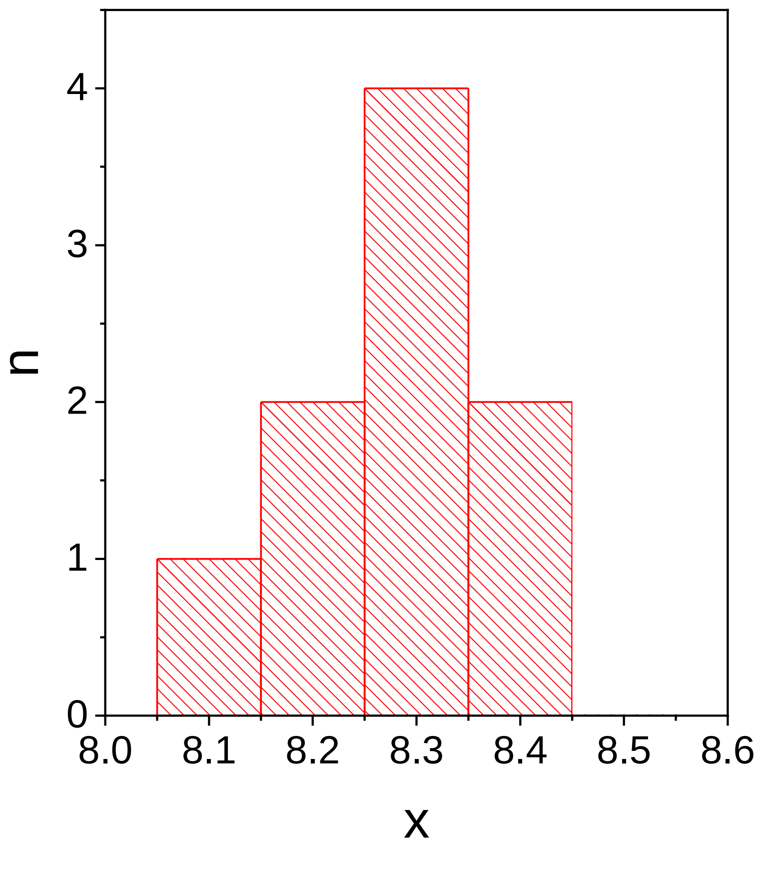
Elaboración de un histograma

Se grafica f, n ó d vs. x, utilizando columnas centradas en $x_{min} + a\left(j - \frac{1}{2}\right)$



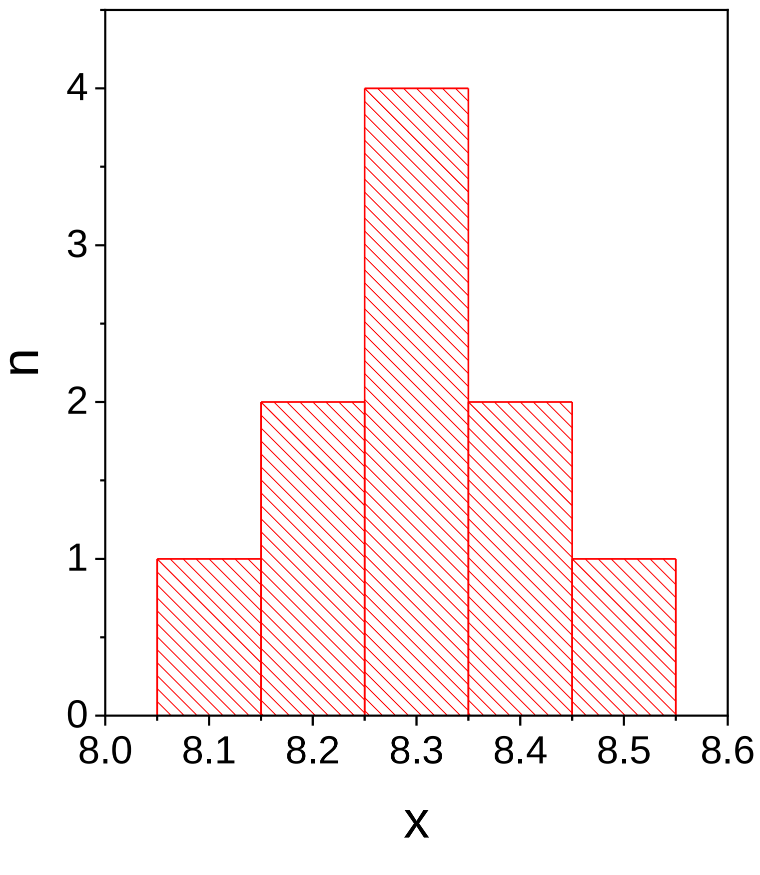
Elaboración de un histograma

Se grafica f, n ó d vs. x, utilizando columnas centradas en $x_{min} + a\left(j - \frac{1}{2}\right)$



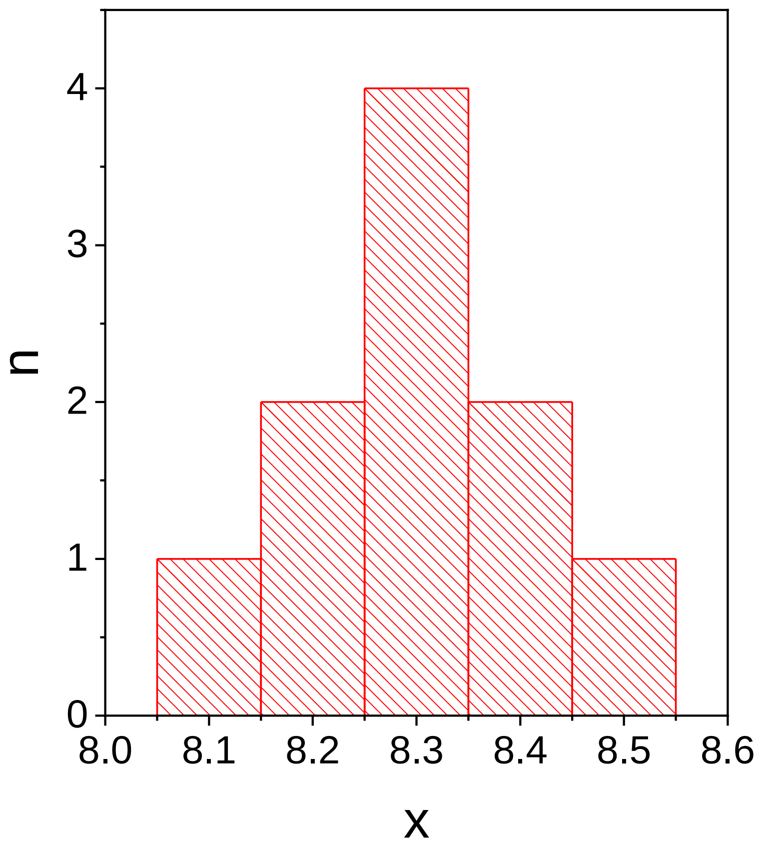
Elaboración de un histograma

Se grafica f, n ó d vs. x, utilizando columnas centradas en $x_{min} + a\left(j - \frac{1}{2}\right)$



Elaboración de un histograma

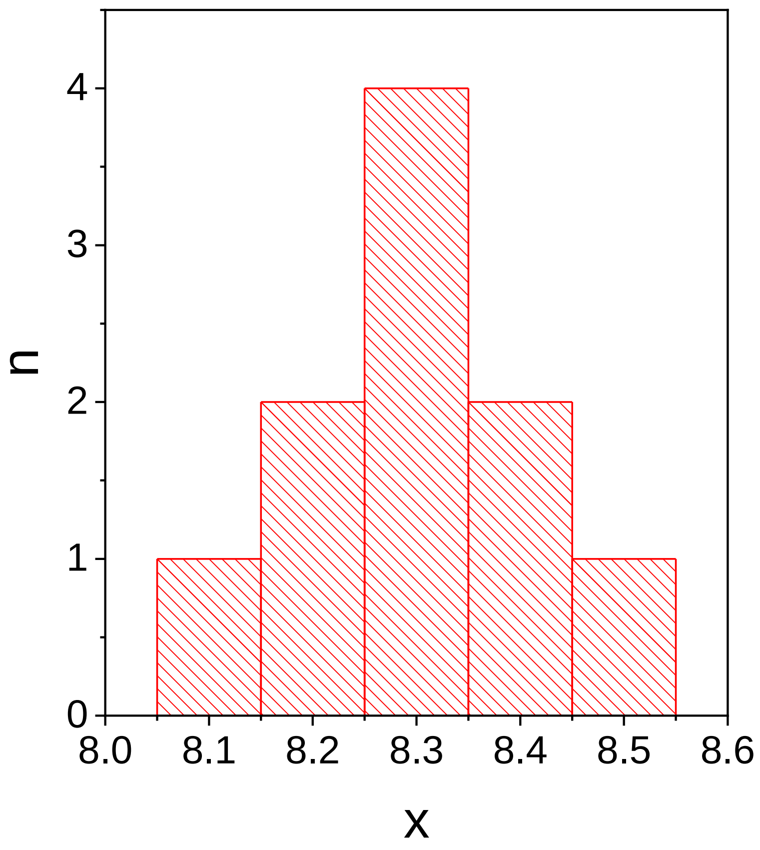
Se grafica f, n ó d vs. x, utilizando columnas centradas en $x_{min} + a\left(j - \frac{1}{2}\right)$



- <u>Media</u>: Es el valor medio: $\bar{\mathrm{x}} = \langle x \rangle = \frac{1}{N} \sum_{i=1}^{N} x_i$
- <u>Moda</u>: Valor de x donde está la máxima frecuencia
- <u>Mediana</u>: Valor de x que divide a la primera mitad de los valores, de la segunda mitad.

Elaboración de un histograma

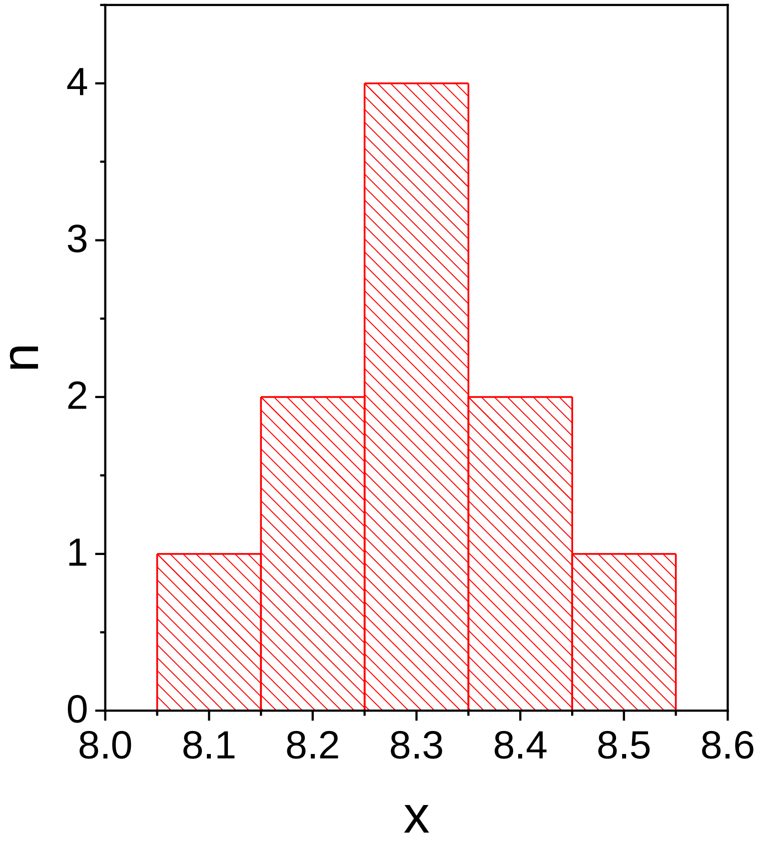
Se grafica f, n ó d vs. x, utilizando columnas centradas en $x_{min} + a\left(j - \frac{1}{2}\right)$



- Varianza: $\sigma_x^2 = \frac{1}{N'} \sum_{i=1}^{N} (x_i - \bar{x})^2$

Elaboración de un histograma

Se grafica f, n ó d vs. x, utilizando columnas centradas en $x_{min} + a\left(j - \frac{1}{2}\right)$

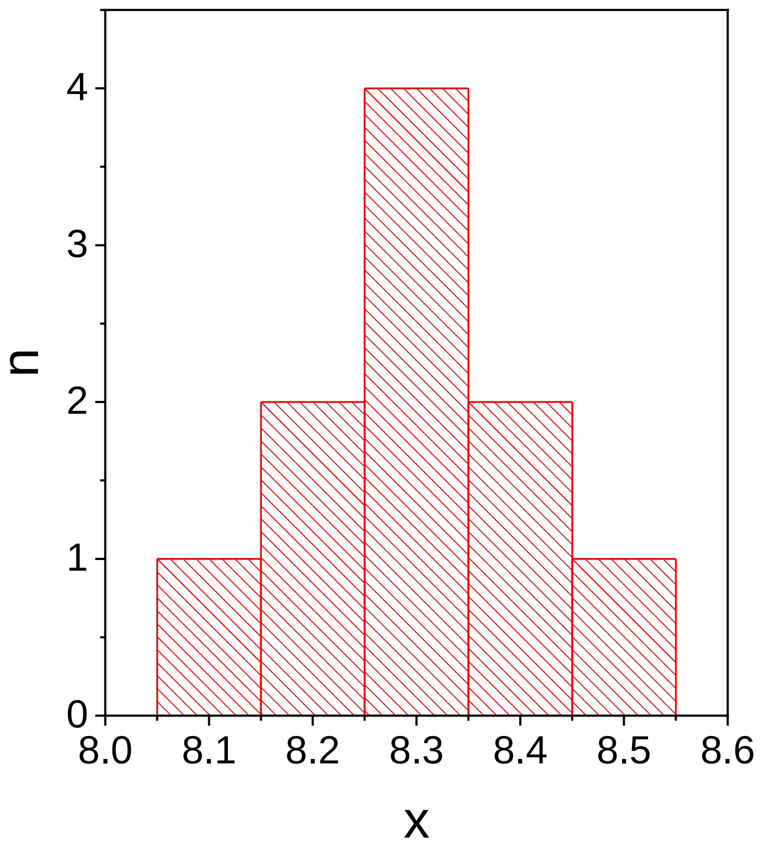


- Varianza: $\sigma_x^2 = \frac{1}{N'}\sum_{i=1}^{N}(x_i - \bar{x})^2$

$N' = N$ si se trata de la población total
$N' = N - 1$ si se trata de una muestra de la población

Elaboración de un histograma

Se grafica f, n ó d vs. x, utilizando columnas centradas en $x_{min} + a\left(j - \frac{1}{2}\right)$



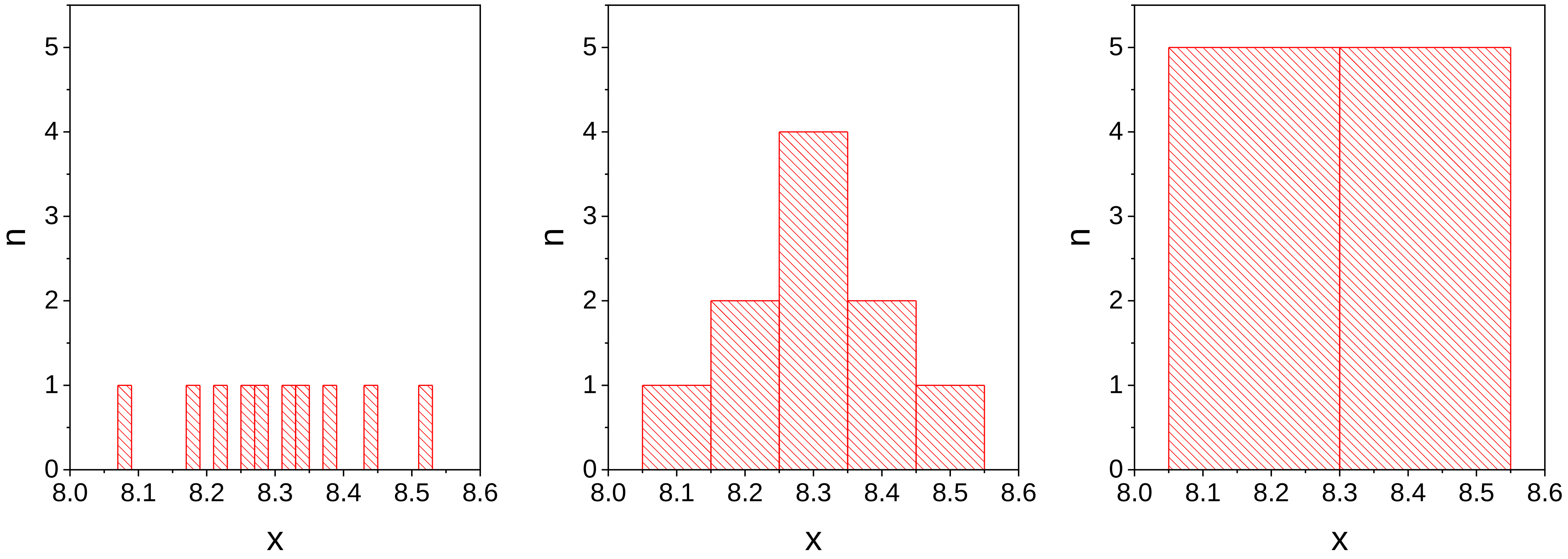
- <u>Varianza</u>: $\sigma_x^2 = \frac{1}{N'}\sum_{i=1}^{N}(x_i - \bar{x})^2$

$N' = N$ si se trata de la población total
$N' = N - 1$ si se trata de una muestra de la población

- <u>Desviación estándar</u>: σ_x

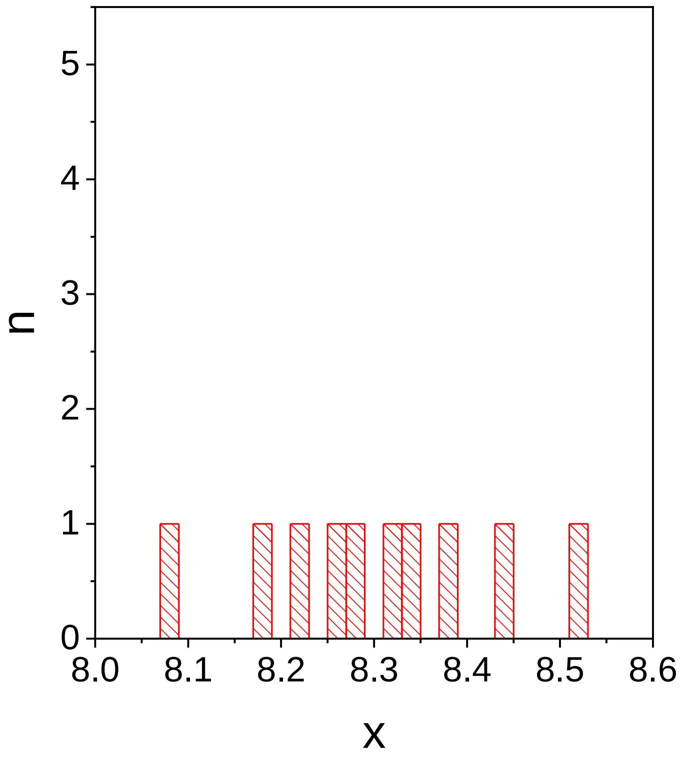
Ancho óptimo del sub-intervalo

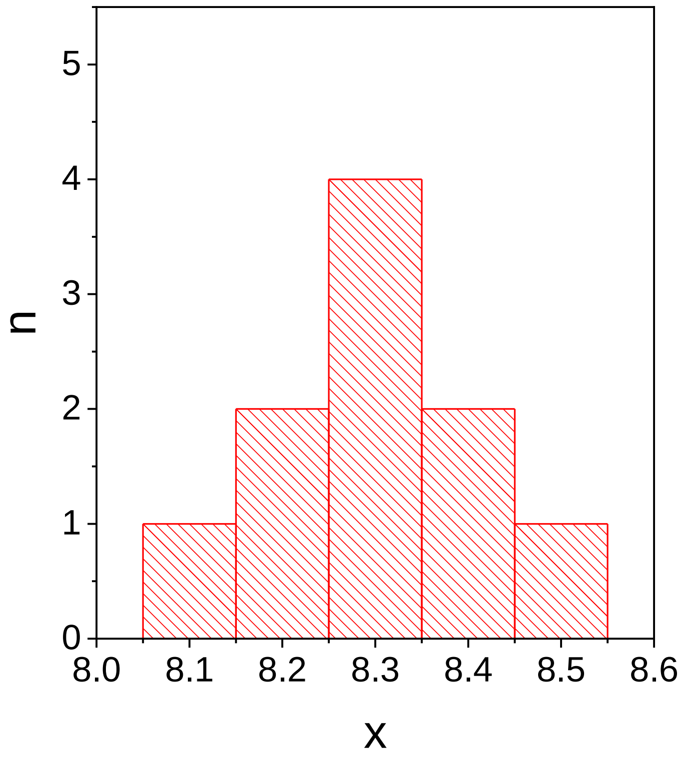
Un ancho demasiado pequeño o demasiado grande impide observar el patrón subyacente.

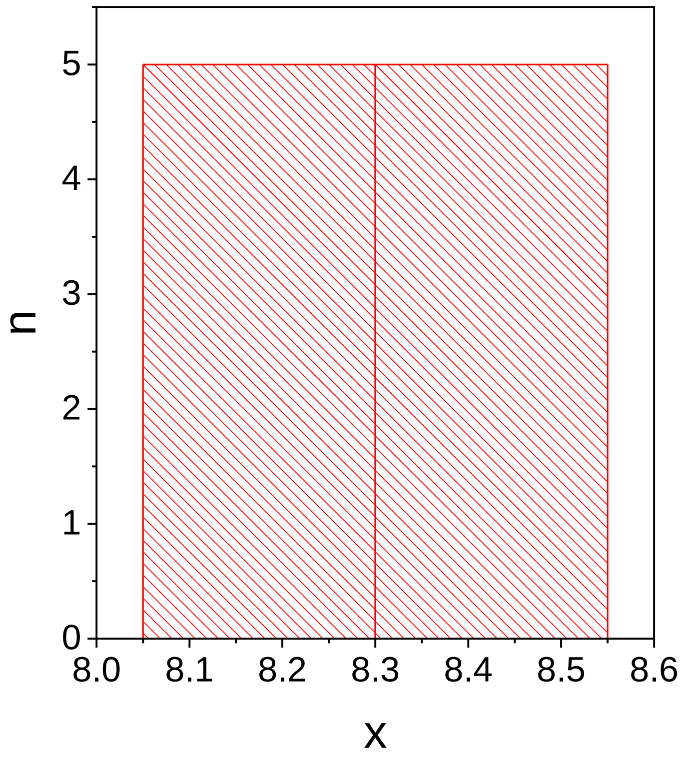


Ancho óptimo del sub-intervalo

Un ancho demasiado pequeño o demasiado grande impide observar el patrón subyacente.







Ancho de columna a considerar[1]: $a = 3.5\sigma_x N^{-\frac{1}{3}}$

[1]David W. Scott, Biometrika, Vol. 66, No. 3 (Dec., 1979), pp. 605-610

Evolución de una distribución según el valor de N

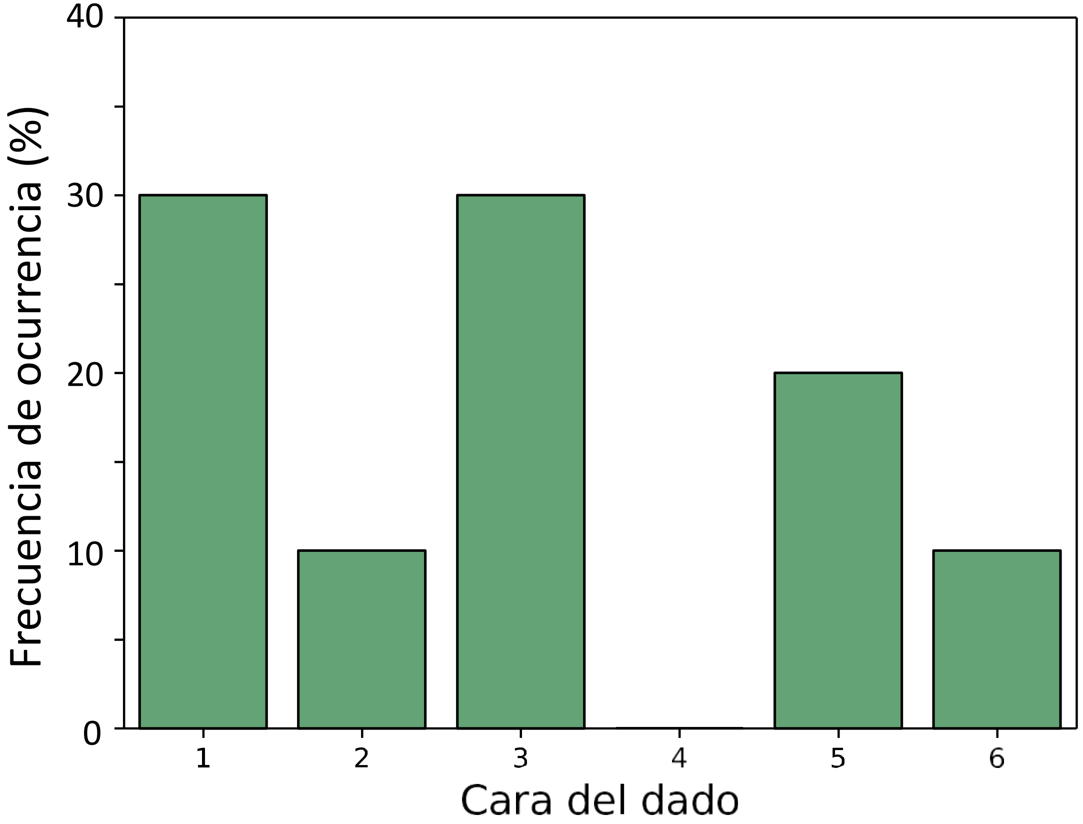
Lanzamos un dado
N = 10 veces

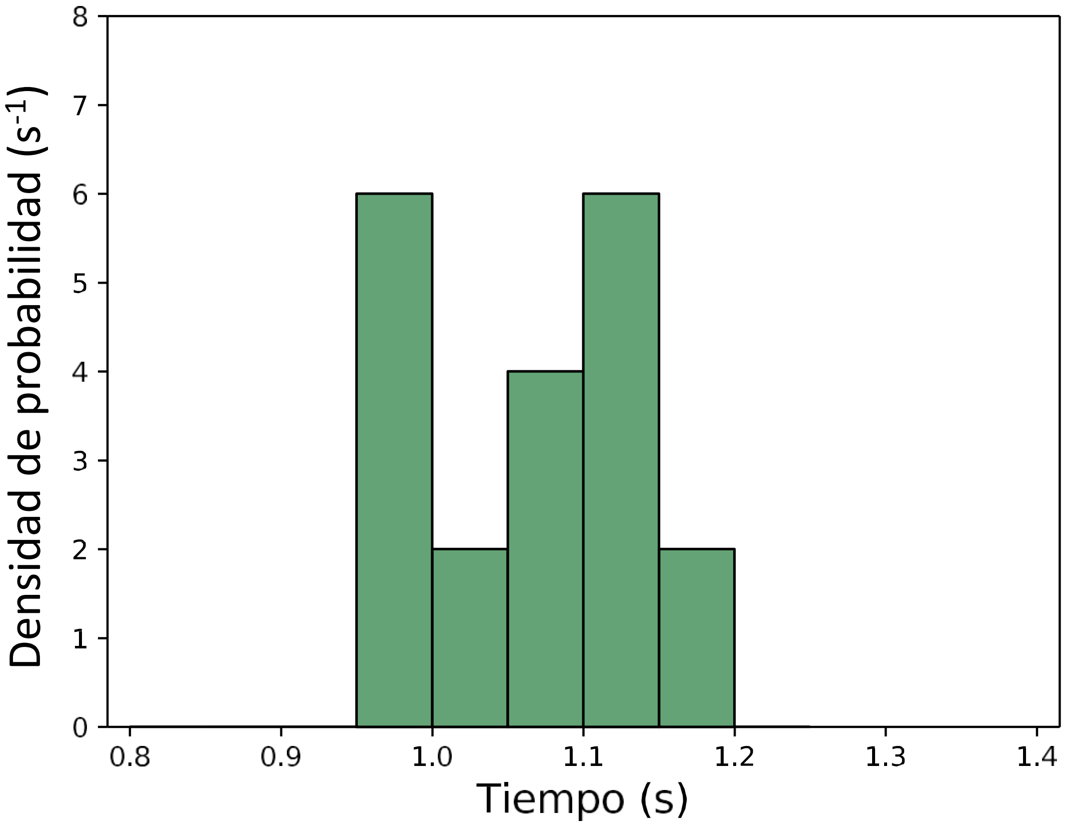
Medimos el periodo de un faro
N = 10 veces

Evolución de una distribución según el valor de N

Lanzamos un dado
N = 10 veces

Medimos el periodo de un faro
N = 10 veces

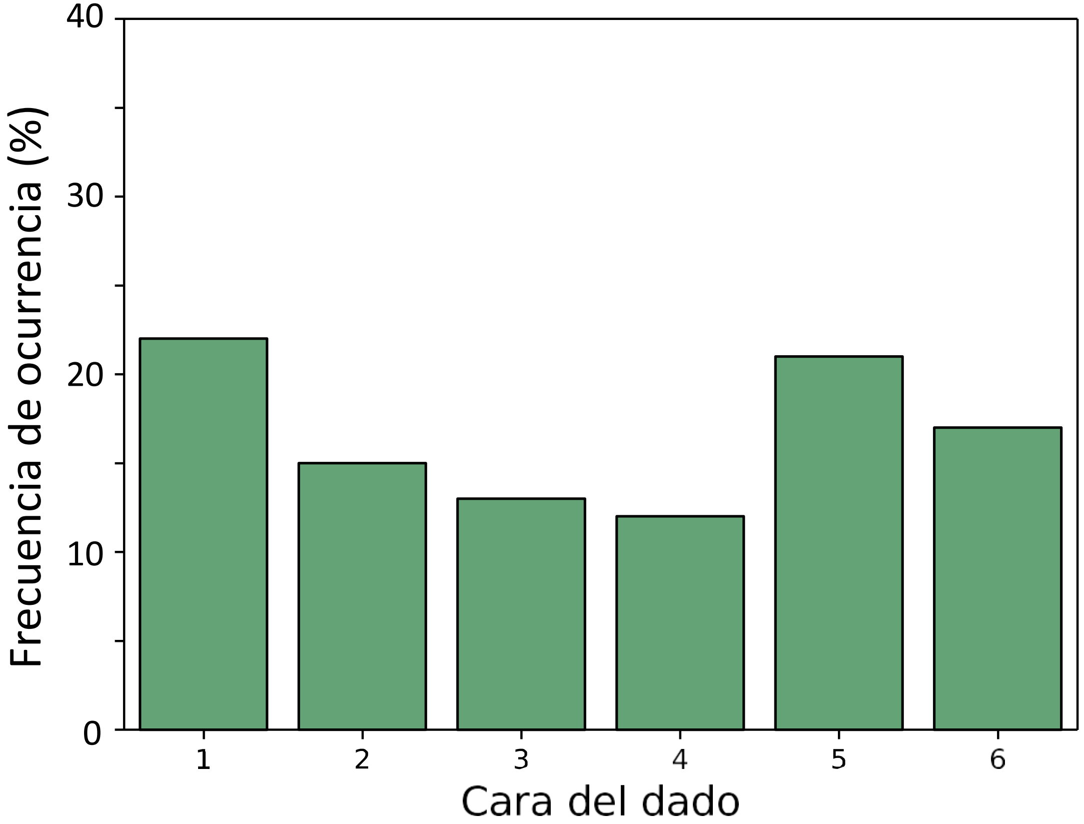


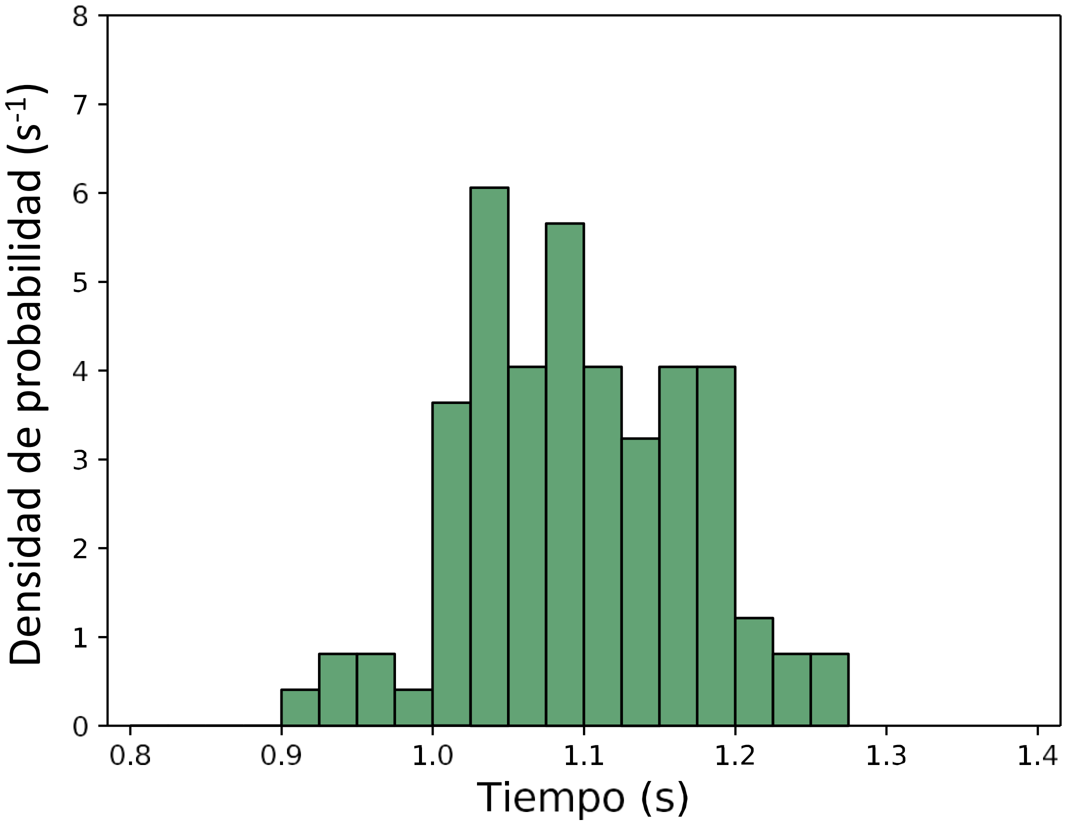


Evolución de una distribución según el valor de N

Lanzamos un dado
N = 100 veces

Medimos el periodo de un faro
N = 100 veces

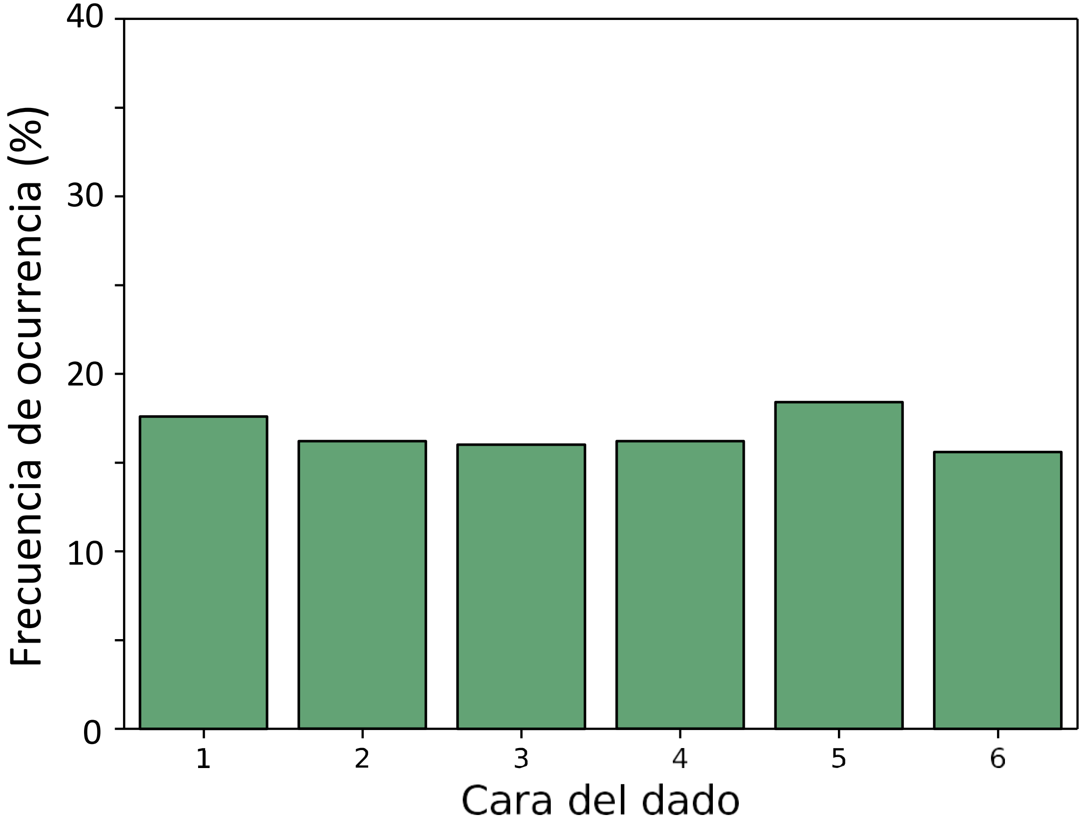


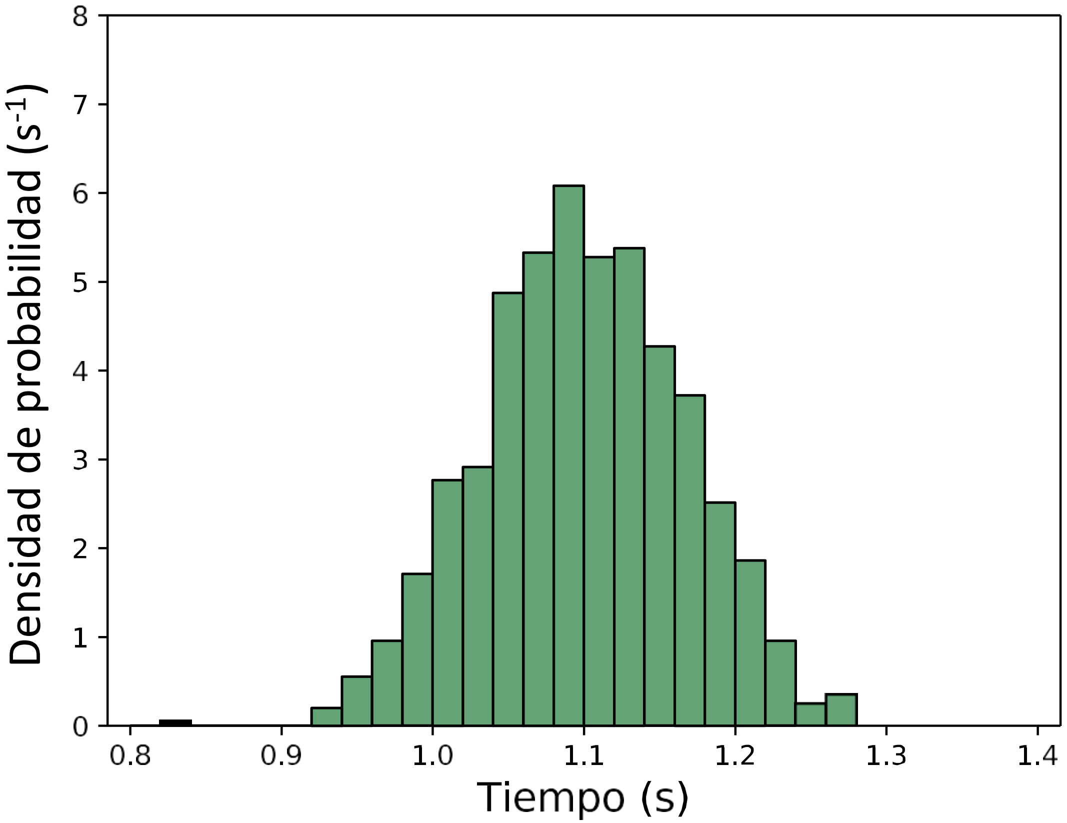


Evolución de una distribución según el valor de N

Lanzamos un dado
N = 1000 veces

Medimos el periodo de un faro
N = 1000 veces

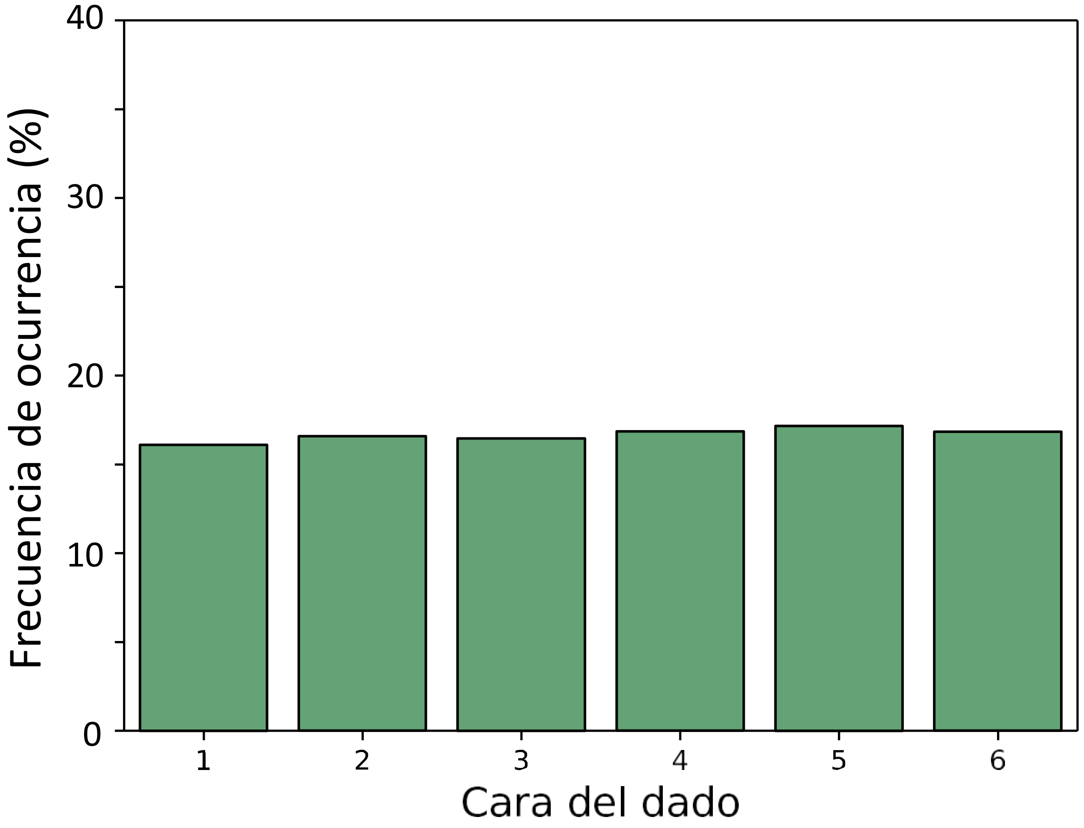


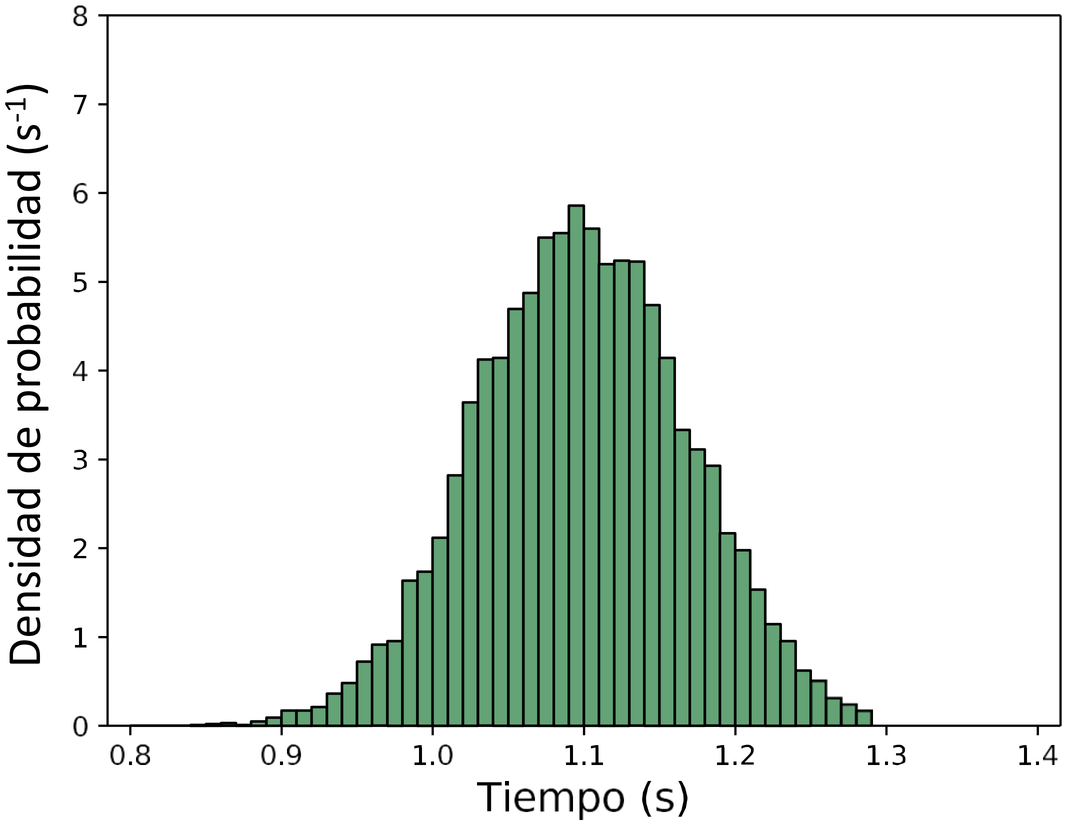


Evolución de una distribución según el valor de N

Lanzamos un dado
N = 10000 veces

Medimos el periodo de un faro
N = 10000 veces

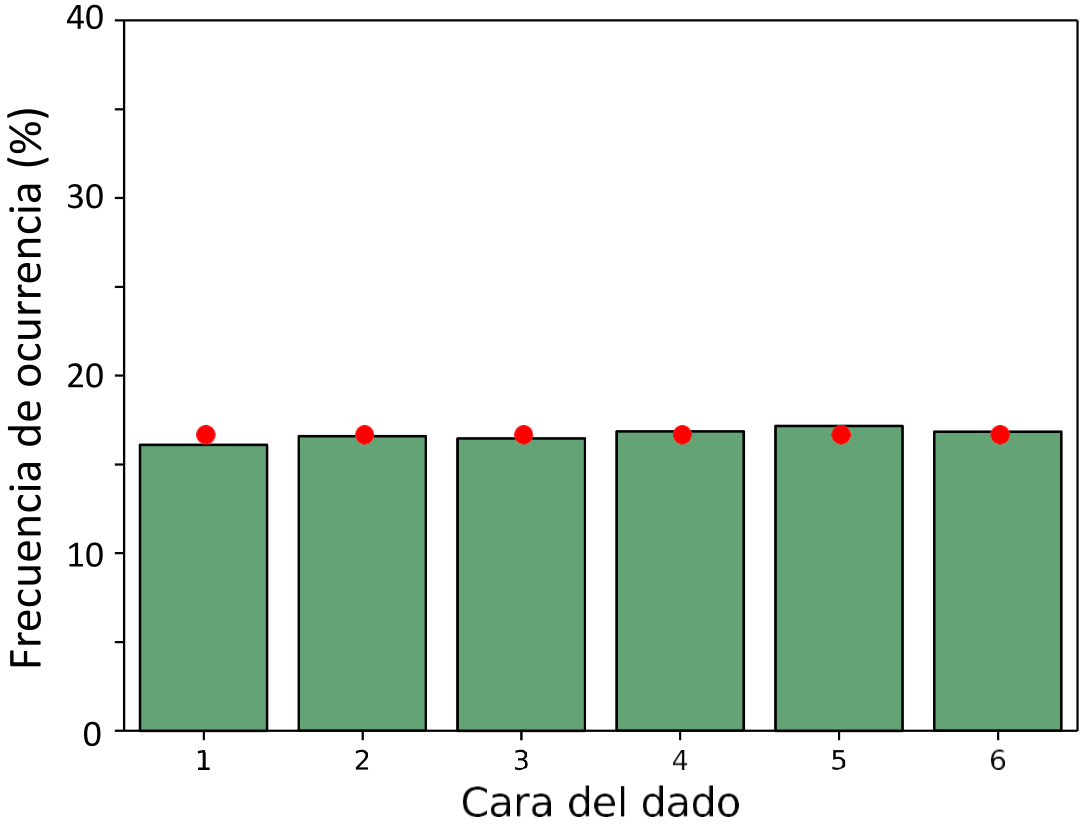


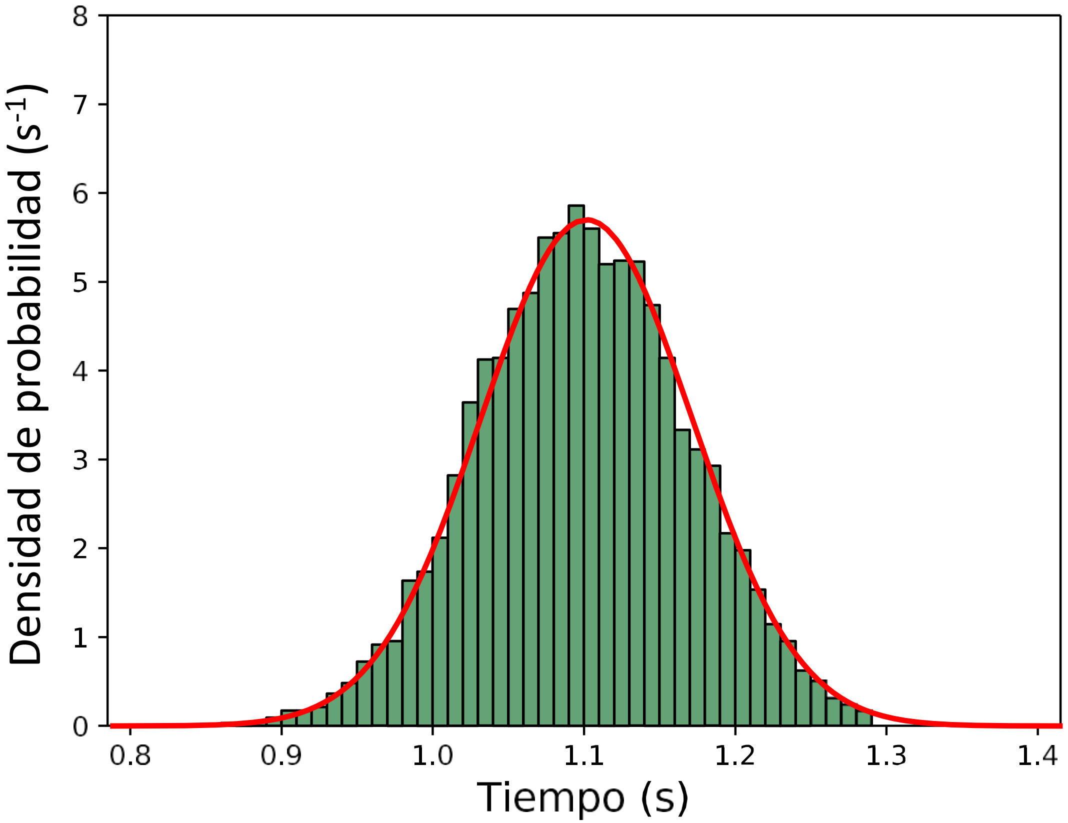


Distribuciones de probabilidad

Todas las posibilidades tienen igual probabilidad

Distribución normal o gaussiana

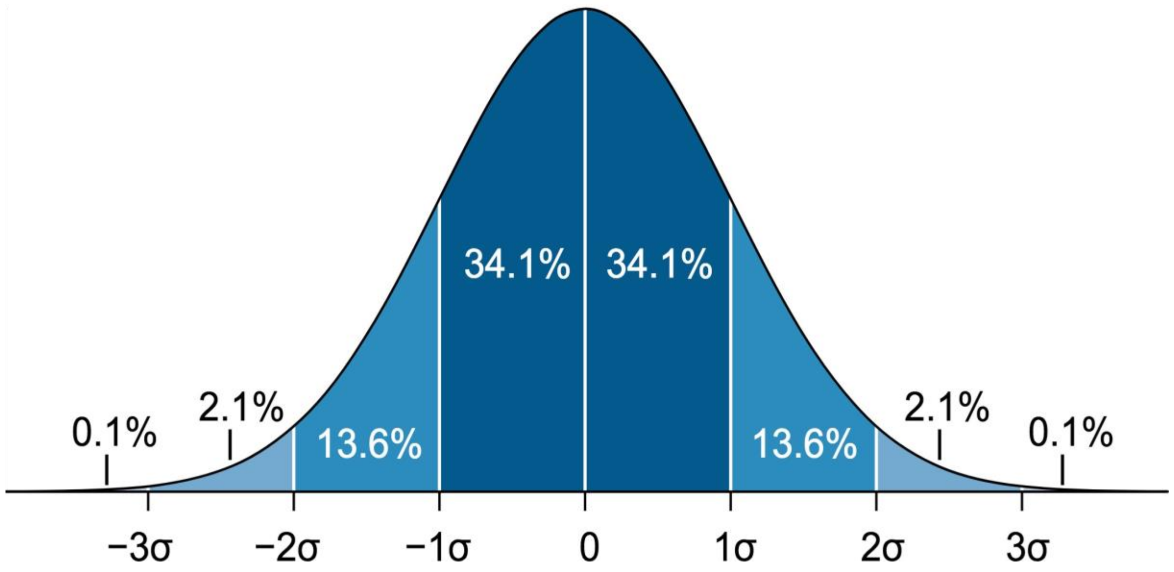




Distribución gaussiana

Función gaussiana

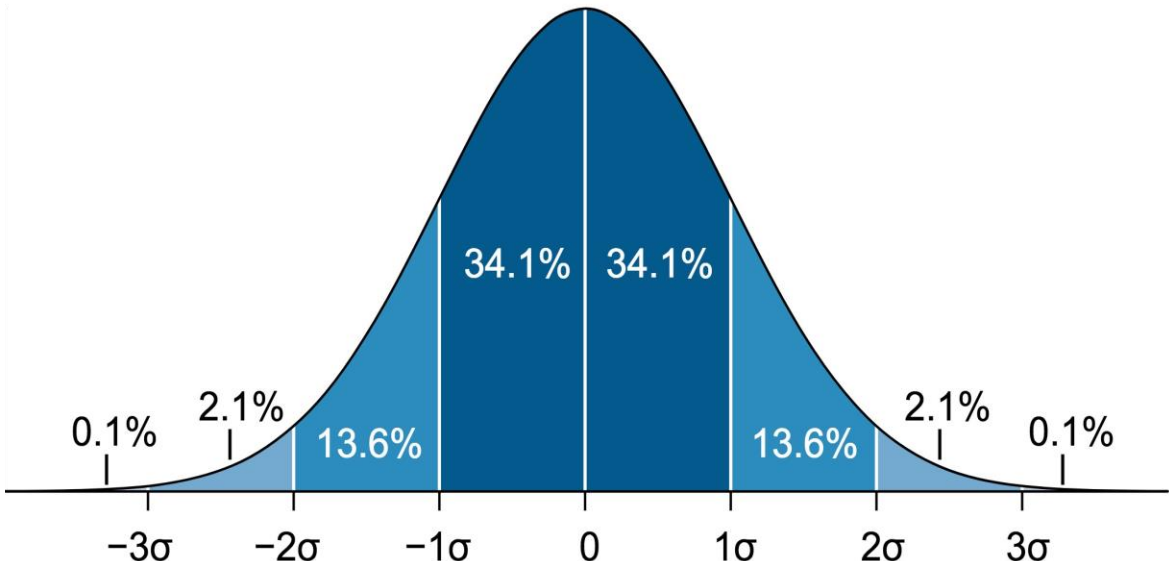
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{(x-x_c)^2}{2\sigma^2}\right)$$



Distribución gaussiana

<u>Función gaussiana</u>

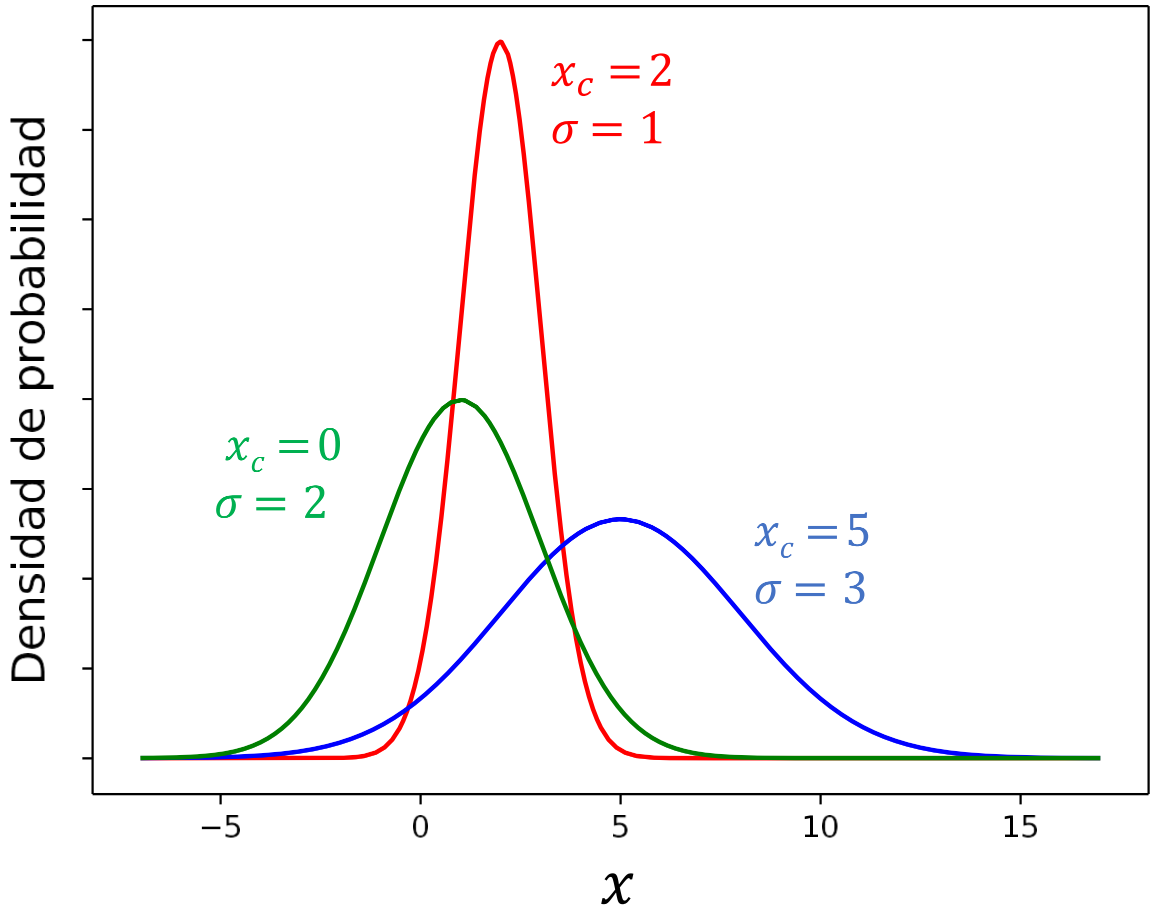
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{(x - x_c)^2}{2\sigma^2}\right)$$



- x_c: centro de la distribución
- σ : Medida del ancho de la distribución
- 2σ contiene el 68.3% de los valores
- 4σ contiene el 95.5% de los valores
- 6σ contiene el 99.7% de los valores

Distribución gaussiana

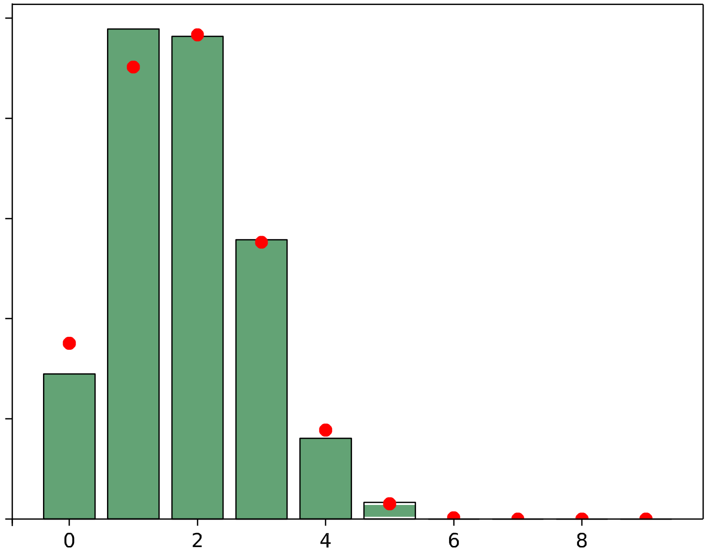
Variando x_c y σ



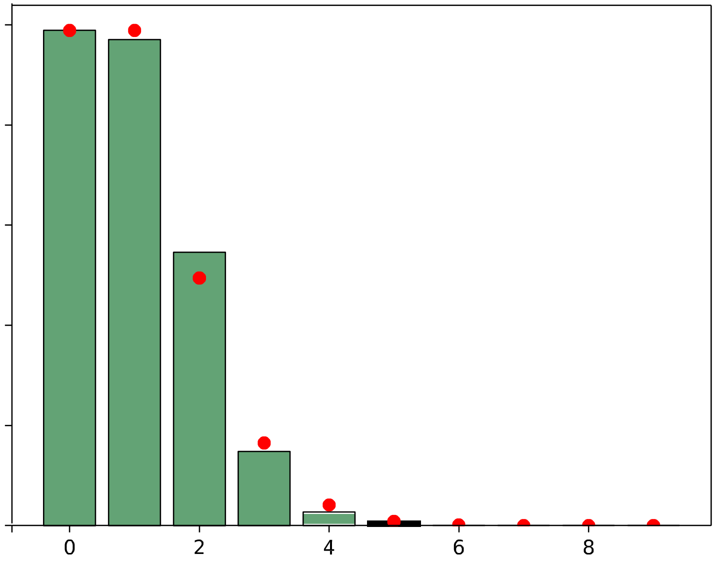
- x_c: centro de la distribución
- σ : Medida del ancho de la distribución
- 2σ contiene el 68.3% de los valores
- 4σ contiene el 95.5% de los valores
- 6σ contiene el 99.7% de los valores

Existen otras distribuciones

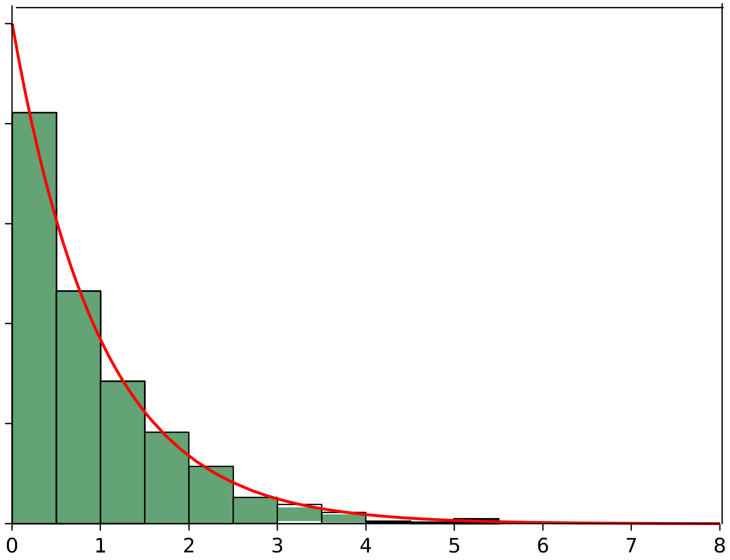
Distribución binomial



Distribución de Poisson



Distribución exponencial



Error de una magnitud según cuántas veces se mide

- N veces $\rightarrow \sigma_{est} = \sigma_{\bar{x}} = \sqrt{\frac{1}{N(N-1)}\sum_{i=1}^{N}(x_i - \overline{\mathrm{x}})^2} = \frac{\sigma_x}{\sqrt{N}}$ ($\sigma_{\bar{x}}$: el error estándar del promedio)

Error de una magnitud según cuántas veces se mide

- N veces $\rightarrow \sigma_{est} = \sigma_{\bar{x}} = \sqrt{\frac{1}{N(N-1)}\sum_{i=1}^{N}(x_i - \bar{\mathrm{x}})^2} = \frac{\sigma_x}{\sqrt{N}}$ ($\sigma_{\bar{x}}$: el error estándar del promedio)

Si reportamos un resultado como $\bar{x} \pm \sigma_{\bar{x}} \rightarrow$ el <u>nivel de confianza</u> es del 68%

Error de una magnitud según cuántas veces se mide

- N veces $\longrightarrow \sigma_{est} = \sigma_{\bar{x}} = \sqrt{\frac{1}{N(N-1)}\sum_{i=1}^{N}(x_i - \bar{\mathrm{x}})^2} = \frac{\sigma_x}{\sqrt{N}}$ ($\sigma_{\bar{x}}$: el error estándar del promedio)

Si reportamos un resultado como $\bar{x} \pm \sigma_{\bar{x}}$ $\longrightarrow$ el <u>nivel de confianza</u> es del 68%

- El error total se calcula como $\Delta x = \sqrt{\sigma_{ap}^2 + \sigma_{exac}^2 + \cdots + \sigma_{est}^2}$

Error de una magnitud según cuántas veces se mide

- N veces $\longrightarrow \sigma_{est} = \sigma_{\bar{x}} = \sqrt{\frac{1}{N(N-1)}\sum_{i=1}^{N}(x_i - \bar{\mathrm{x}})^2} = \frac{\sigma_x}{\sqrt{N}}$ ($\sigma_{\bar{x}}$: el error estándar del promedio)

Si reportamos un resultado como $\bar{x} \pm \sigma_{\bar{x}}$ $\longrightarrow$ el <u>nivel de confianza</u> es del 68%

- El error total se calcula como $\Delta x = \sqrt{\sigma_{ap}^2 + \sigma_{exac}^2 + \cdots + \sigma_{est}^2}$

- Si a mayor *N*, menor σ_{est} ¿Cuántas veces tiene sentido medir?

Error de una magnitud según cuántas veces se mide

- N veces $\rightarrow \sigma_{est} = \sigma_{\bar{x}} = \sqrt{\frac{1}{N(N-1)}\sum_{i=1}^{N}(x_i - \bar{\mathrm{x}})^2} = \frac{\sigma_x}{\sqrt{N}}$ ($\sigma_{\bar{x}}$: el error estándar del promedio)

Si reportamos un resultado como $\bar{x} \pm \sigma_{\bar{x}}$ $\rightarrow$ el <u>nivel de confianza</u> es del 68%

- El error total se calcula como $\Delta x = \sqrt{\sigma_{ap}^2 + \sigma_{exac}^2 + \cdots + \sigma_{est}^2}$

- Si a mayor *N*, menor σ_{est} ¿Cuántas veces tiene sentido medir?

Un criterio sería medir tantas veces como para conseguir al menos $\sigma_{est} \approx \sigma_{instrumento}$

Discrepancia y promedios pesados

Discrepancia

- Medición 1: $X_1 = \overline{X_1} \pm \Delta X_1$
- Medición 2: $X_2 = \overline{X_2} \pm \Delta X_2$

Discrepancia y promedios pesados

Discrepancia

- Medición 1: $X_1 = \overline{X_1} \pm \Delta X_1$
- Medición 2: $X_2 = \overline{X_2} \pm \Delta X_2$

$\Rightarrow$ Definimos $\Delta X = \sqrt{\Delta X_1^2 + \Delta X_2^2}$

Discrepancia y promedios pesados

Discrepancia

- Medición 1: $X_1 = \overline{X_1} \pm \Delta X_1$
- Medición 2: $X_2 = \overline{X_2} \pm \Delta X_2$

$\longrightarrow$ Definimos $\Delta X = \sqrt{\Delta X_1^2 + \Delta X_2^2}$

$|\overline{X_2} - \overline{X_1}| \geq \Delta X \rightarrow$ Mediciones distintas entre sí con un límite de confianza del 68%

$|\overline{X_2} - \overline{X_1}| \geq 2\Delta X \rightarrow$ Mediciones distintas entre sí con un límite de confianza del 95%

Discrepancia y promedios pesados

Discrepancia

- Medición 1: $X_1 = \overline{X_1} \pm \Delta X_1$
- Medición 2: $X_2 = \overline{X_2} \pm \Delta X_2$

Definimos $\Delta X = \sqrt{\Delta X_1^2 + \Delta X_2^2}$

$|\overline{X_2} - \overline{X_1}| \geq \Delta X \rightarrow$ Mediciones distintas entre sí con un límite de confianza del 68%

$|\overline{X_2} - \overline{X_1}| \geq 2\Delta X \rightarrow$ Mediciones distintas entre sí con un límite de confianza del 95%

Promedio pesados entre L mediciones independientes con su error

$$\boxed{\langle x \rangle_p = \sigma^2_{\langle x \rangle_p} \sum_{k=1}^{L} \frac{x_k}{\sigma_k^2}}$$

Discrepancia y promedios pesados

Discrepancia

- Medición 1: $X_1 = \overline{X_1} \pm \Delta X_1$
- Medición 2: $X_2 = \overline{X_2} \pm \Delta X_2$

Definimos $\Delta X = \sqrt{\Delta X_1^2 + \Delta X_2^2}$

$|\overline{X_2} - \overline{X_1}| \geq \Delta X \rightarrow$ Mediciones distintas entre sí con un límite de confianza del 68%

$|\overline{X_2} - \overline{X_1}| \geq 2\Delta X \rightarrow$ Mediciones distintas entre sí con un límite de confianza del 95%

Promedio pesados entre L mediciones independientes con su error

$$\boxed{\langle x \rangle_p = \sigma_{\langle x \rangle_p}^2 \sum_{k=1}^{L} \frac{x_k}{\sigma_k^2}}$$

$$\boxed{\frac{1}{\sigma_{\langle x \rangle_p}^2} = \sum_{k=1}^{L} \frac{1}{\sigma_k^2}}$$

$\langle x \rangle_p$: promedio pesado de x

$\sigma_{\langle x \rangle_p}$: Error del promedio pesado