# Distribuciones estadísticas

## Análisis estadístico utilizando R


UNQ UNTreF CONICET

Pablo Etchemendy

Ignacio Spiousas

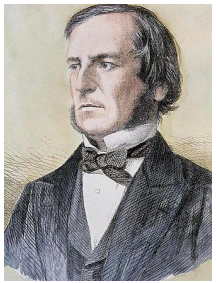
**Agosto 2021**

- Binomial
- Uniforme
- Normal (aka Gaussiana)


## Experimento de Bernoulli

- Cualquier experimento que tenga dos resultados posibles:
  - Cara o ceca
  - Verdadero o Falso
  - Éxito o fracaso
  - Si o no
  - Etc.

- También conocido como variable Booleana



## Distribución binomial

- Describe la frecuencia esperada para el resultado de N experimentos de Bernoulli independientes
  - Tirar N monedas iguales y contar la cantidad de “caras”
  - Permite describir monedas balanceadas y desbalanceadas

Se representa mediante una

**función matemática**

$$f(k) = \Pr(X = k)$$

que describe la **probabilidad de que la variable aleatoria tome un dado valor k**

$$f(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

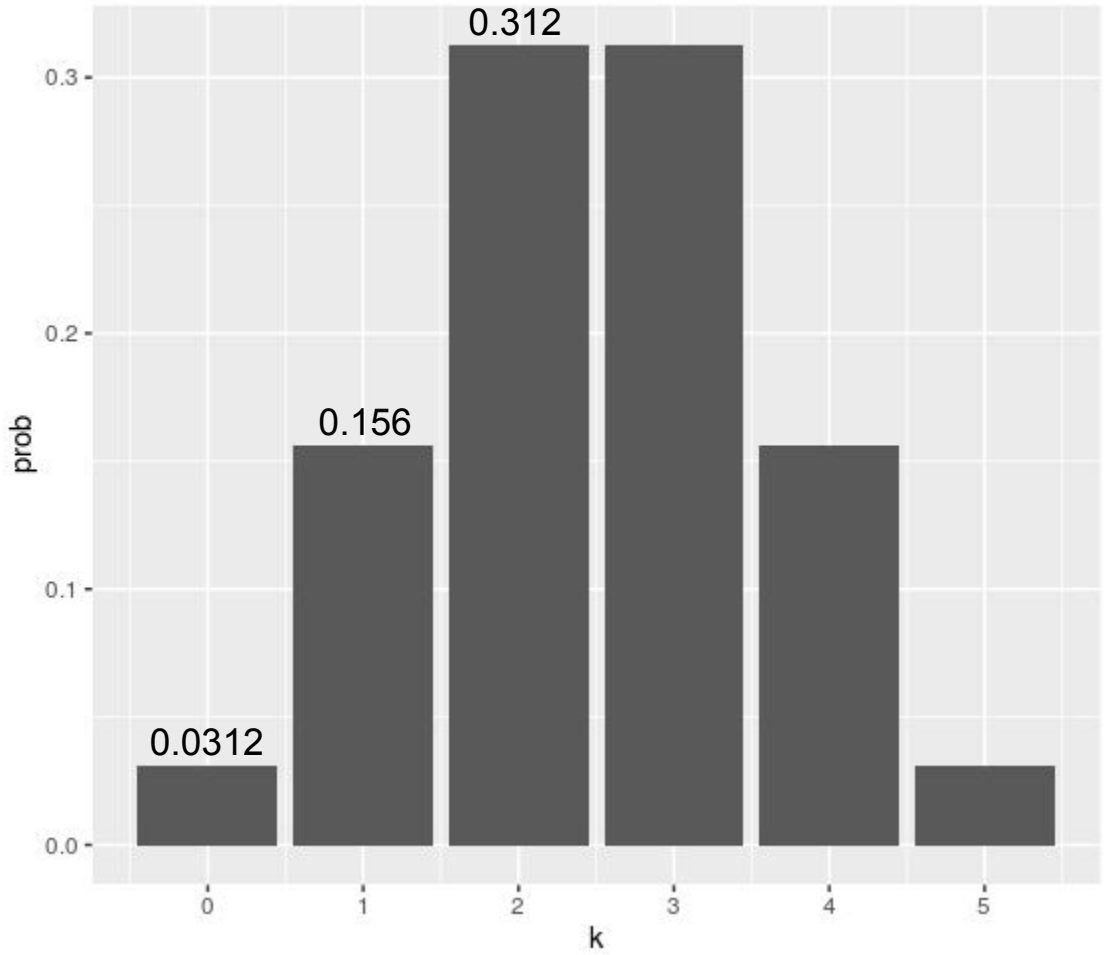
p y n son dos parámetros de la distribución

- p: probabilidad, para cada moneda individual, de obtener cara
- n: cantidad de monedas que forman una realización

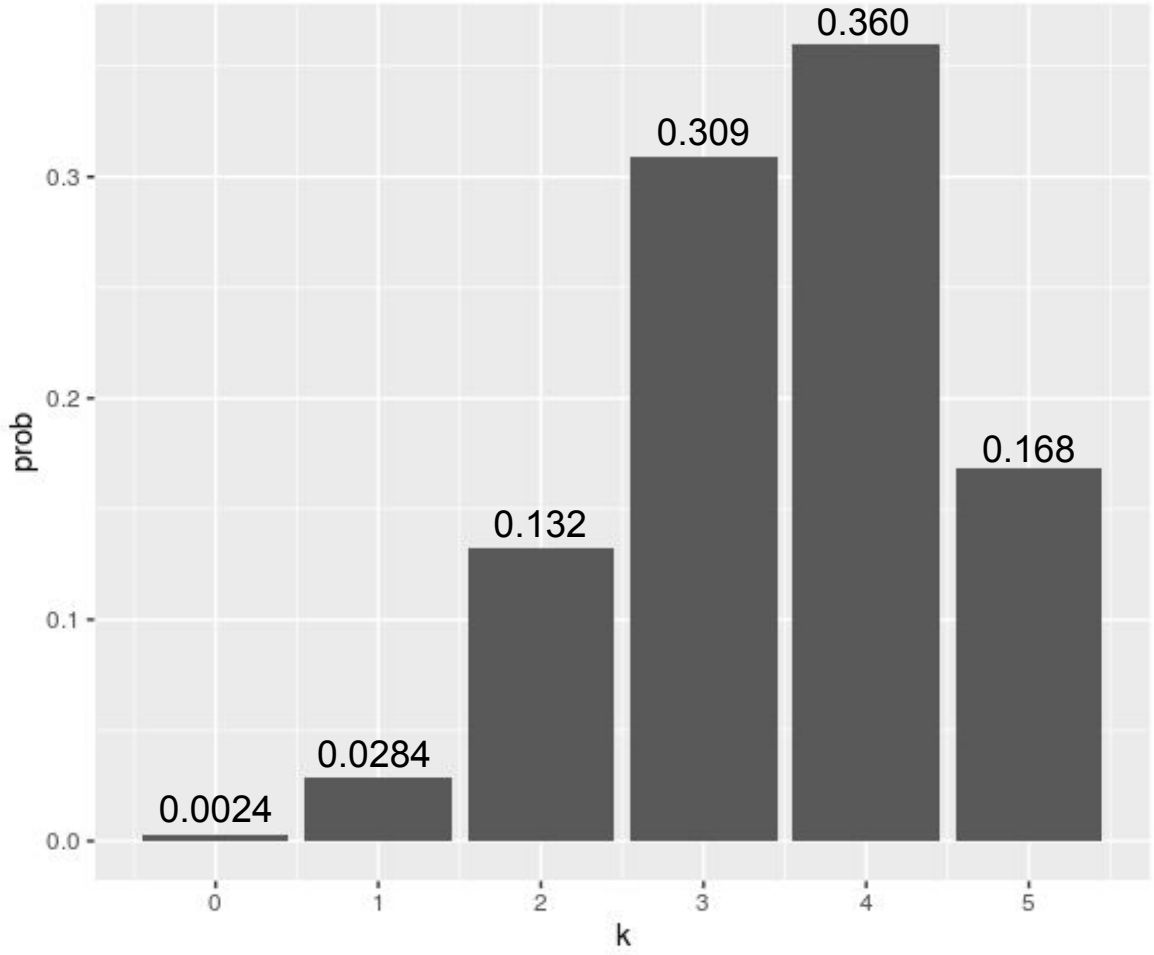
$$f(k; n, p)$$

Hago explícitos los parámetros de la distribución

n = 5
p = 0.5



n = 5
p = 0.7



La distribución binomial describe una variable aleatoria discreta

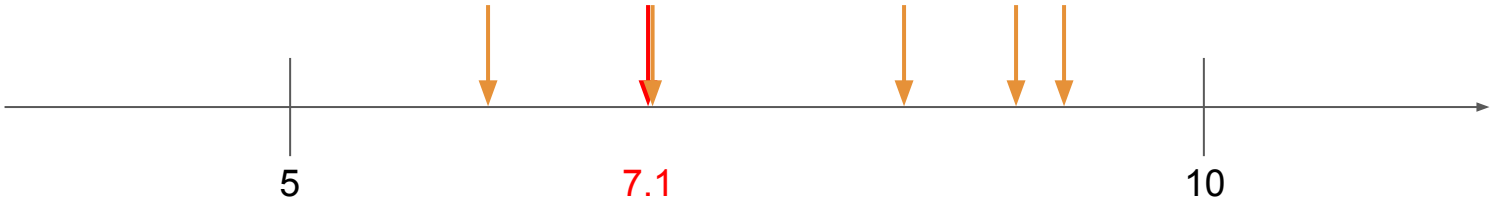
Veamos ahora el caso de una variable aleatoria continua...

## Distribución uniforme

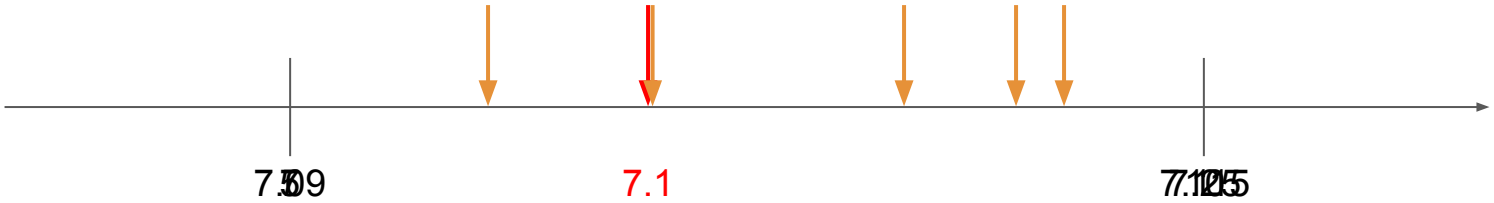
- Describe una variable continua cuyos valores posibles son equiprobables dentro de un cierto intervalo finito
  - Por ejemplo, consideremos el intervalo entre 5 y 10.
  - Las chances de obtener cualquier valor fuera del intervalo definido por 5 y 10 es 0.
  - ¿Pero cuál es la chance de obtener un cierto valor dentro de ese intervalo?
    - Por ejemplo, ¿cuál es la chance de obtener un 6?


5
10

¿Cuál es la chance de volver a obtener el mismo número
si tiro el dardo infinitas veces?



¿Cuál es la chance de volver a obtener el mismo número
si tiro el dardo infinitas veces?



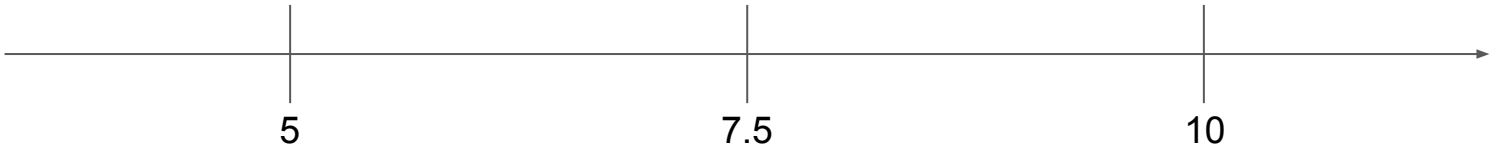
Recordemos que la frecuencia relativa nos servía como estimación de la probabilidad

$$\Pr(X = 7.1) = f(7.1) = \frac{1}{N} \longrightarrow 0 \text{ (si } N \to \infty)$$

¿Dónde está mi probabilidad?

Voy a redefinir el problema…

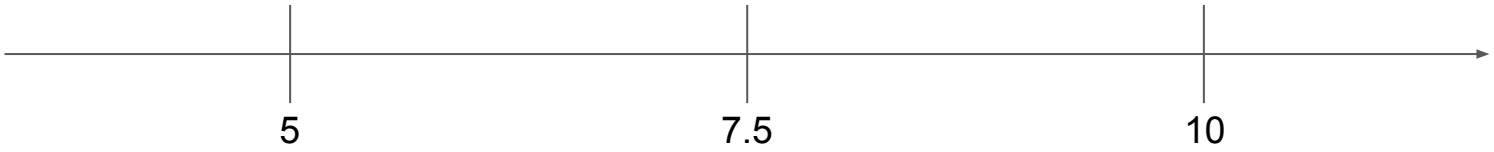
Voy a preguntarme por la probabilidad de obtener un valor
en un sub-intervalo



Tengo dos sub-intervalos → tengo dos probabilidades

$$\text{Prob}(5 \leq x < 7.5) = 0.5$$

$$\text{Prob}(7.5 \leq x \leq 10) = 0.5$$



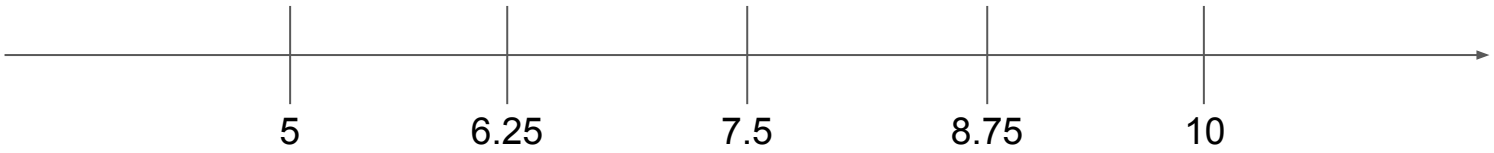
Si divido en cuatro sub-intervalos

$$\text{Prob}(5 \le x < 6.25) = 0.25$$

$$\text{Prob}(6.25 \le x < 7.5) = 0.25$$

$$\text{Prob}(7.5 \le x < 8.75) = 0.25$$

$$\text{Prob}(8.75 \le x \le 10) = 0.25$$



Si divido en N sub-intervalos

$$\text{Prob}(5 \le x < 5+5/N) = 1/N$$

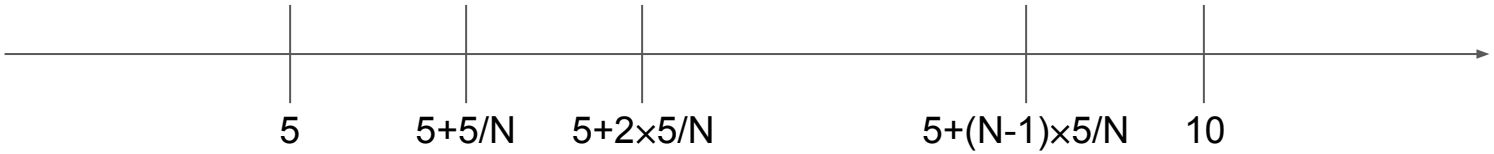
...

$$\text{Prob}(5+(N-1)\times 5/N \le x \le 10) = 1/N$$

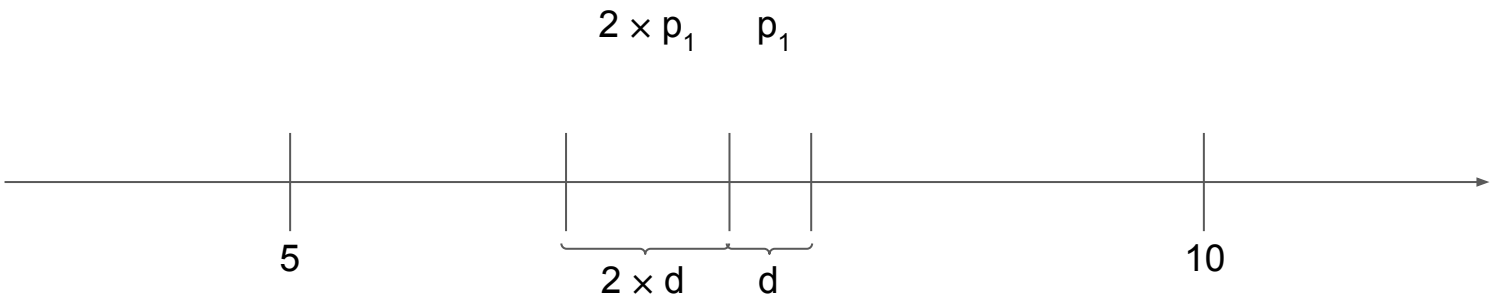
Lo que ocurre aquí es que podemos dividir la probabilidad de obtener un valor cualquiera (1 por definición) en N partes iguales (ya que la variable es uniforme)

Para cada intervalo:

- Su probabilidad es 1/N
- Su tamaño es (10-5)/N = 5/N

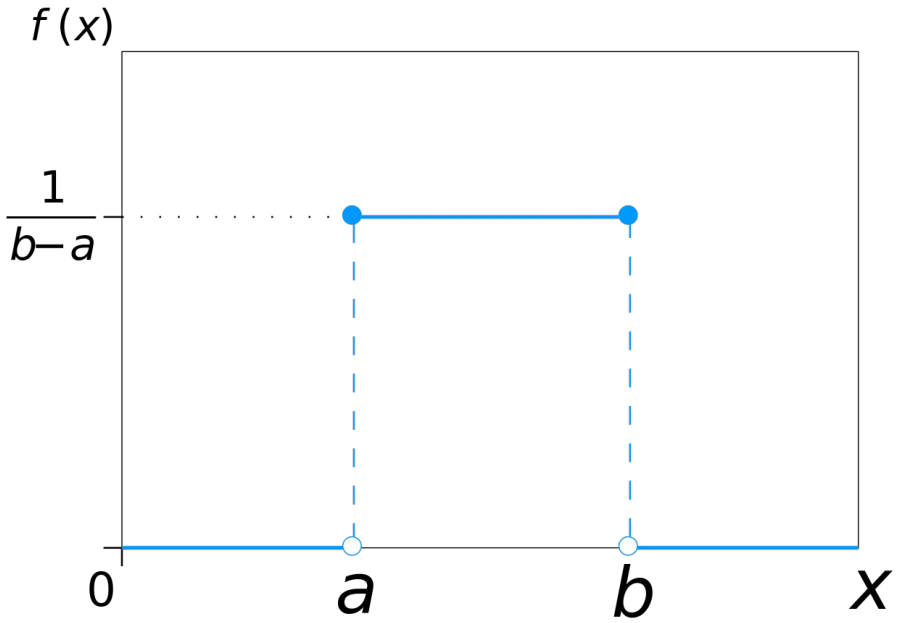


La probabilidad de obtener un número en un cierto intervalo es proporcional a la longitud del intervalo.



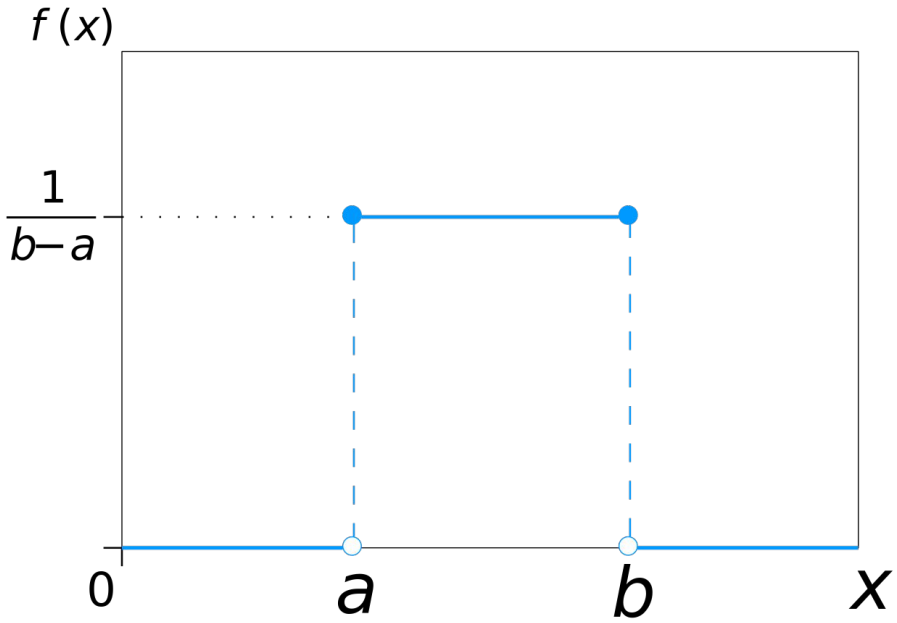
Para variables continuas, se define una función de **densidad de probabilidad**

En este caso, la densidad será uniforme, por lo tanto, la densidad es constante en el intervalo [a, b]



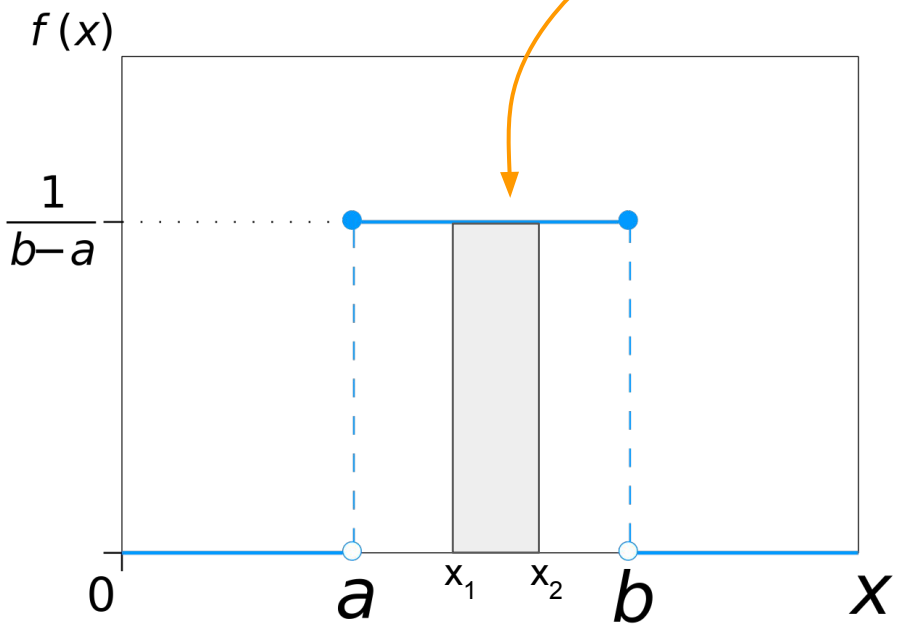
En toda función **densidad de probabilidad**, el área bajo la curva debe valer 1

Esto significa que la probabilidad de obtener un valor cualquiera en el intervalo [a, b] es 100%



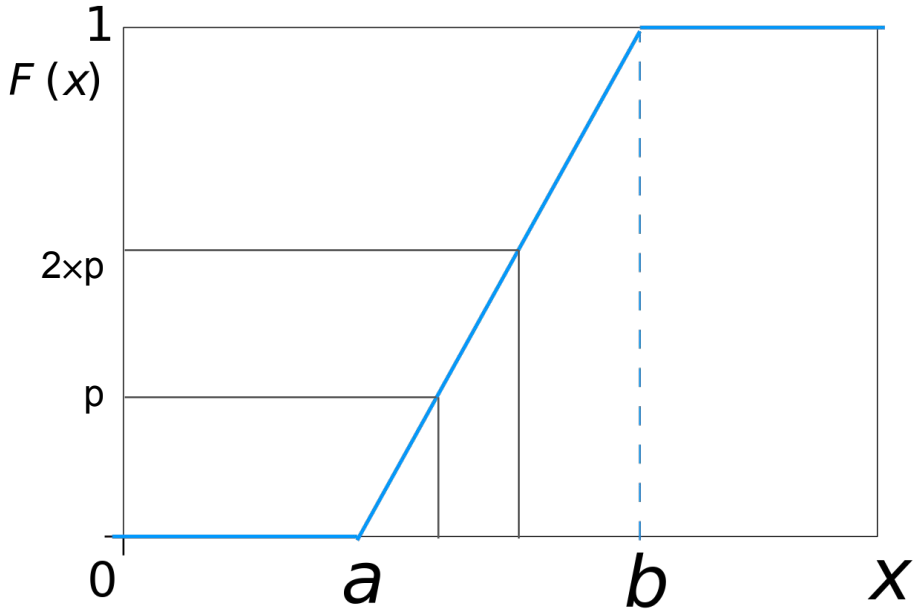
La probabilidad de obtener un valor entre $x_1$ y $x_2$ es

$$\Pr(x_1 \leq X \leq x_2) = \frac{x_2 - x_1}{b - a}$$



Se define la función de **probabilidad acumulada** como

$$g(x) = \Pr(a \leq X \leq x)$$



Distribución Gaussiana (aka normal)

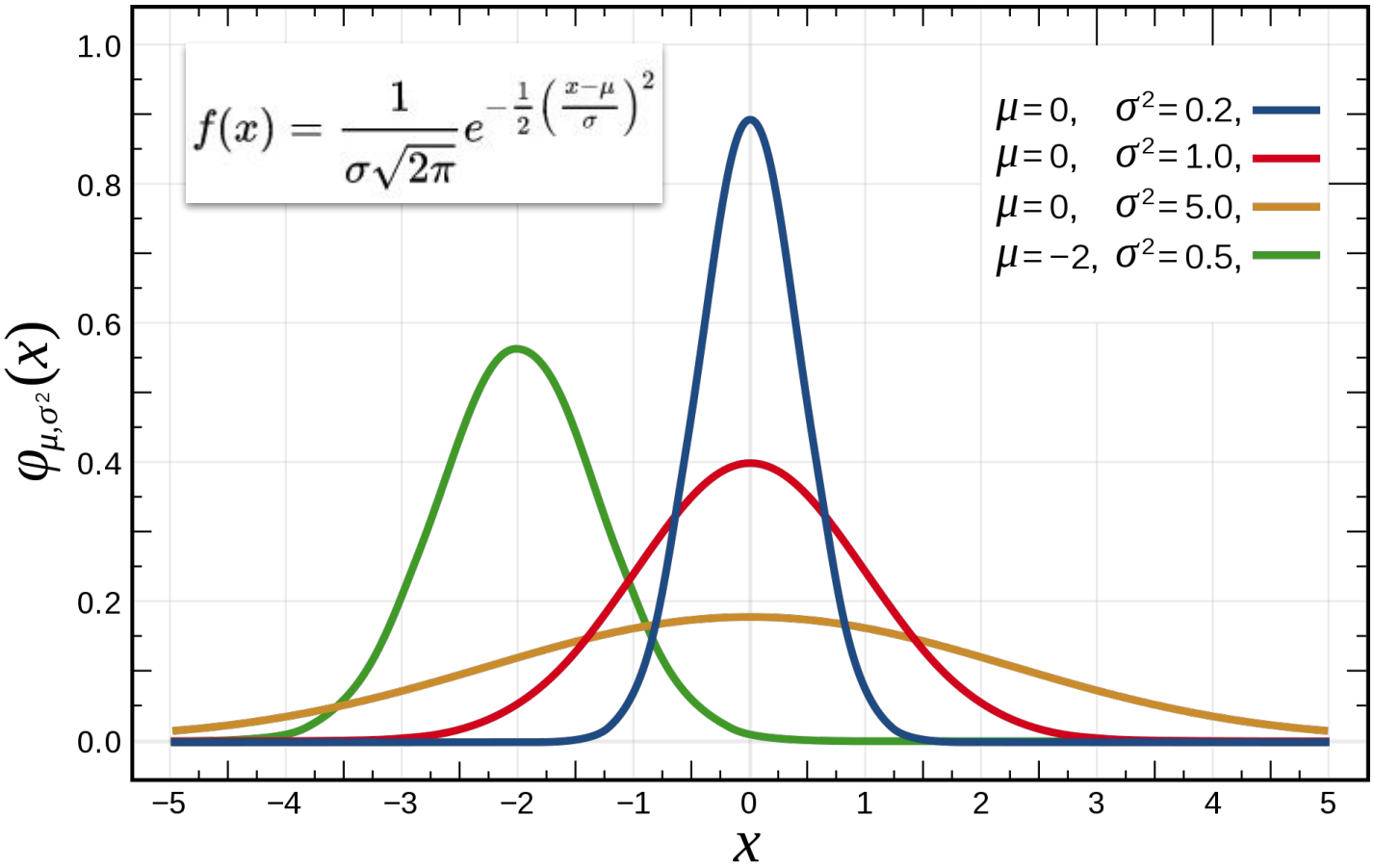
## Distribución Gaussiana (aka normal)

- Describe una variable continua
- Valores posibles: todos los números reales $[-\infty, \infty]$
- La mayoría de la probabilidad está concentrada alrededor de un valor central (primer parámetro)...
- ...y disminuye para valores más lejanos.
- Qué se considera “lejano” depende del segundo parámetro: la dispersión alrededor del valor central.
- Distribución simétrica

**Parámetros**

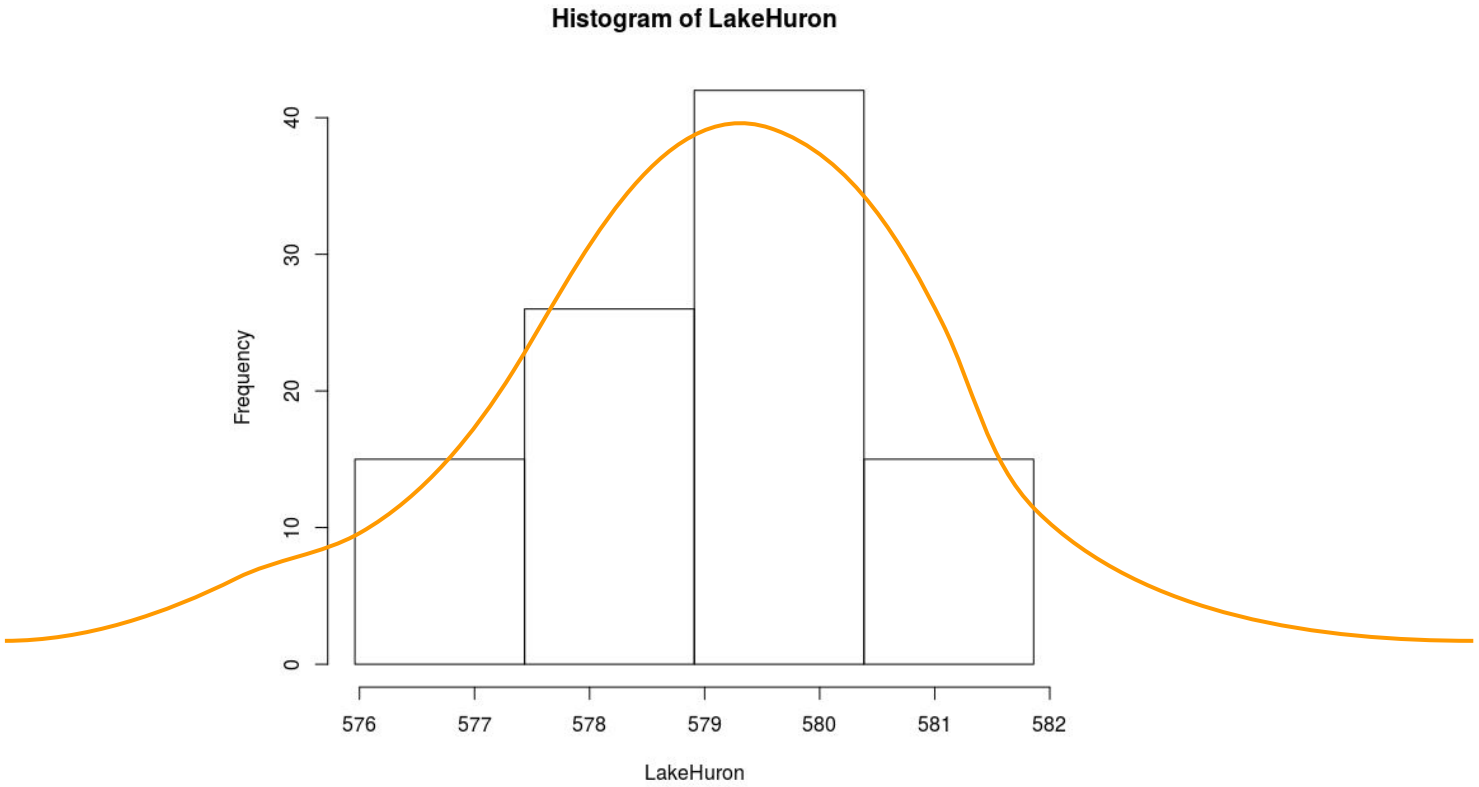
$$X \sim N(\mu, \sigma^2)$$

- μ: parámetro central (aka, la media)
- σ: parámetro de dispersión (aka, desvío estándar)
- $\sigma^2$: ídem σ, y se lo conoce como varianza




https://en.wikipedia.org/wiki/Normal_distribution#/media/File:Normal_Distribution_PDF.svg

**¿Y para qué sirve?**


Histogram of LakeHuron
Frequency
0
10
20
30
40
576
577
578
579
580
581
582
LakeHuron