

# Introducción al análisis cuantitativo de datos lingüísticos

## Bloque 1.2: Probabilidad y distribuciones

Ezequiel Koile (MPI-SSH)  
Carolina Gattei (IFIBA – CONICET)

# El concepto de probabilidad

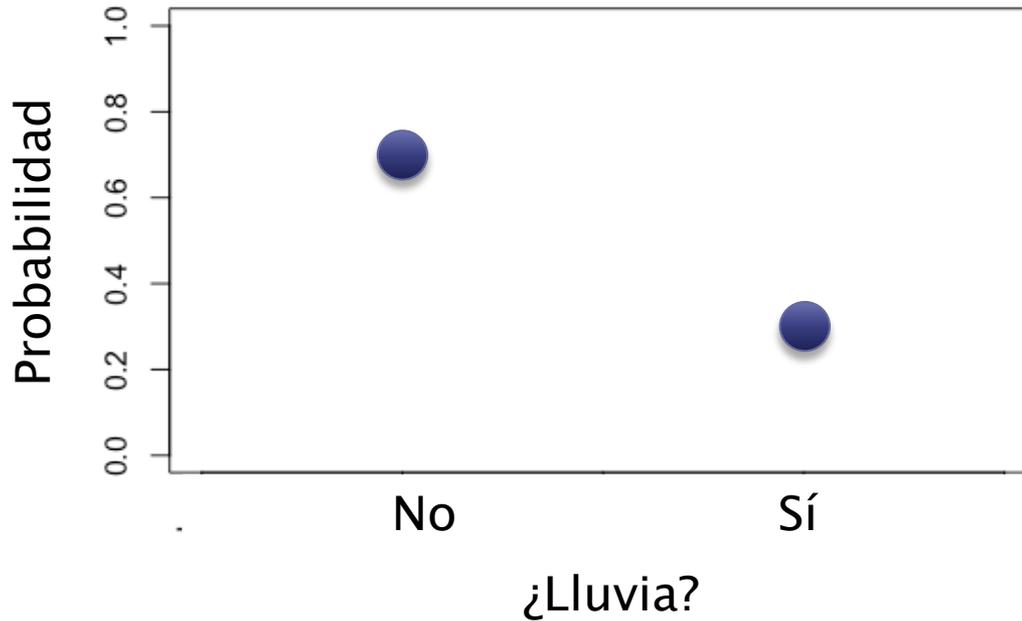
- ▶ Nos interesa saber cuán probable es un evento
- ▶ En lugar de porcentajes, utilizaremos números entre 0 y 1:
  - 65%      -->      0.65
  - 50%      -->      0.5
  - 12.3%    -->      0.123
  - 100%     -->      1
  - 0%        -->      0
- ▶ Las probabilidades se refieren a los eventos que pertenecen a un *espacio muestral*: el conjunto de todas las cosas que pueden pasar, de acuerdo con nuestro modelo.
- ▶ En general, veremos cosas como:
  - $\text{Pr}(\text{RESULTADO} = \text{VALOR}) = \text{NÚMERO ENTRE 0 Y 1}$

# Ejemplo 1:

## ¿Va a llover hoy en Buenos Aires?

- ▶ El pronóstico dice que hay un 30% de probabilidad de lluvias para hoy:
  - $\Pr(\text{lluvia} = \text{sí}) = 0.3$
- ▶ ¿Cuál será la probabilidad de que no llueva hoy?  
Claramente...
  - $\Pr(\text{lluvia} = \text{no}) = 0.7$
- ▶ Como regla general, la suma de las probabilidades de todos los eventos del espacio muestral será 1 (algo va a pasar):
  - $\Pr(\text{lluvia} = \text{sí}) + \Pr(\text{lluvia} = \text{no}) = 1$

# Ejemplo 1: ¿Va a llover hoy en Buenos Aires?



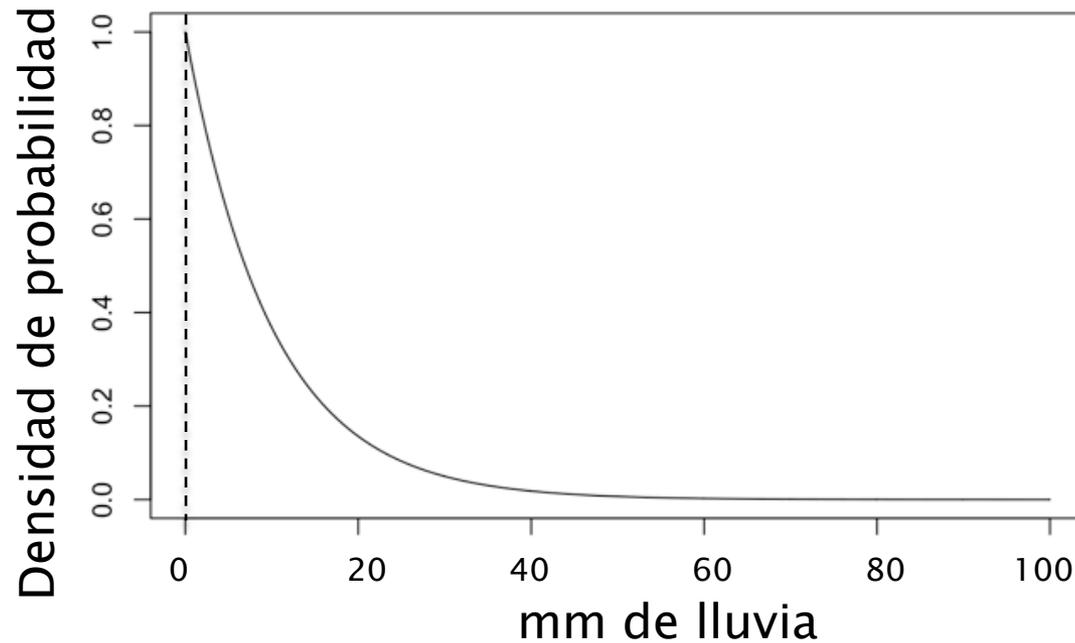
# Ejemplo 2:

## ¿Cuánto va a llover hoy en Buenos Aires?

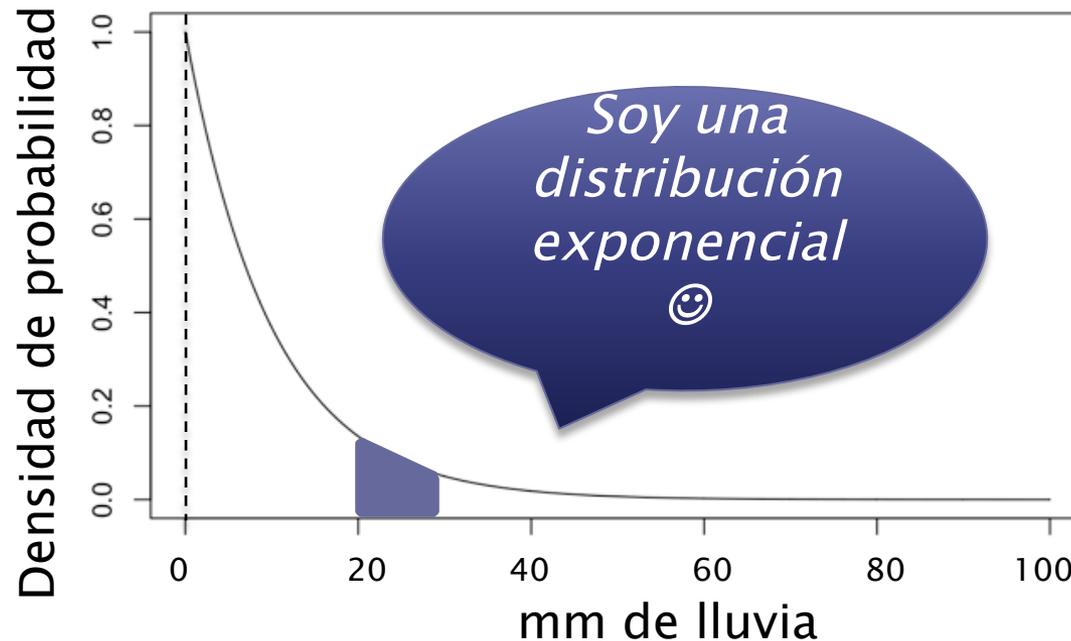
- ▶ Queremos saber ahora cuántos milímetros van a llover hoy en Buenos Aires
  - $\Pr(\text{mm de lluvia} = 20^*) = 0.10$
- ▶ Cualquier número real no negativo es un valor posible para esta pregunta, pero algunos serán más probables que otros
- ▶ Si quisiéramos describir todos los valores posibles, necesitaremos una función en el espacio de los números reales. Esta es una *distribución de probabilidad*:
  - $\Pr(\text{mm de lluvia} = \text{tantos}) = f(\text{tantos})$

(\*entre 20 y 30mm)

## Ejemplo 2: ¿Cuánto va a llover hoy en Buenos Aires?



# Ejemplo 2: ¿Cuánto va a llover hoy en Buenos Aires?



# Probabilidad conjunta (más de una variable)

- $\Pr(\text{lluvia} = \text{sí}) = 0.3$
- $\Pr(\text{día} = \text{martes}) = 1/7 = 0.1428\dots$

# Probabilidad conjunta (más de una variable)

- $\Pr(\text{llueve}) = 0.3$
- $\Pr(\text{es martes}) = 1/7 = 0.1428\dots$
- $\Pr(\text{llueve Y es martes}) = \Pr(\text{llueve}) \times \Pr(\text{es martes})$   
 $= 0.3 \times 0.1428\dots$   
 $= 0.04285\dots$

¿Siempre podemos hacer esto?

# Este razonamiento funciona siempre, ¿no?

