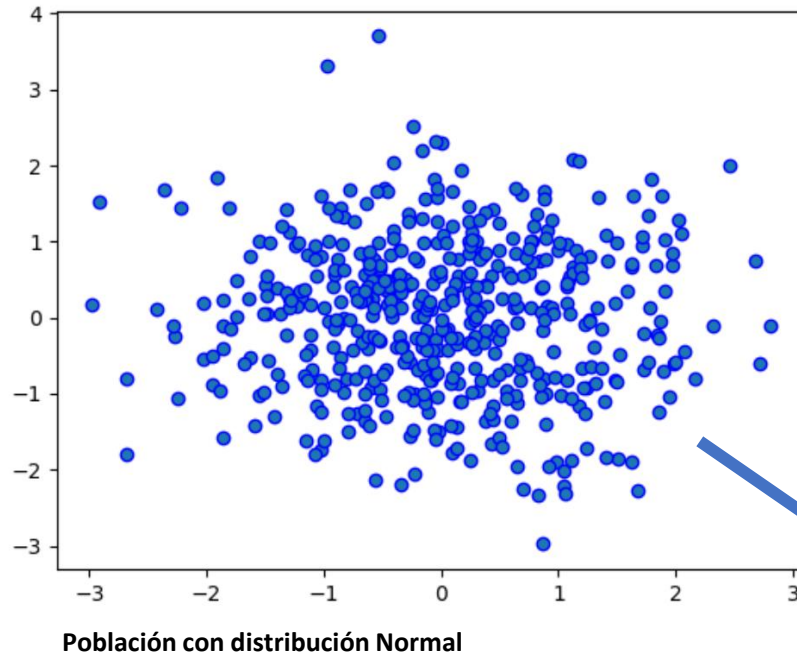




Ajuste de distribuciones

Rodrigo Maranzana

De los datos a los parámetros

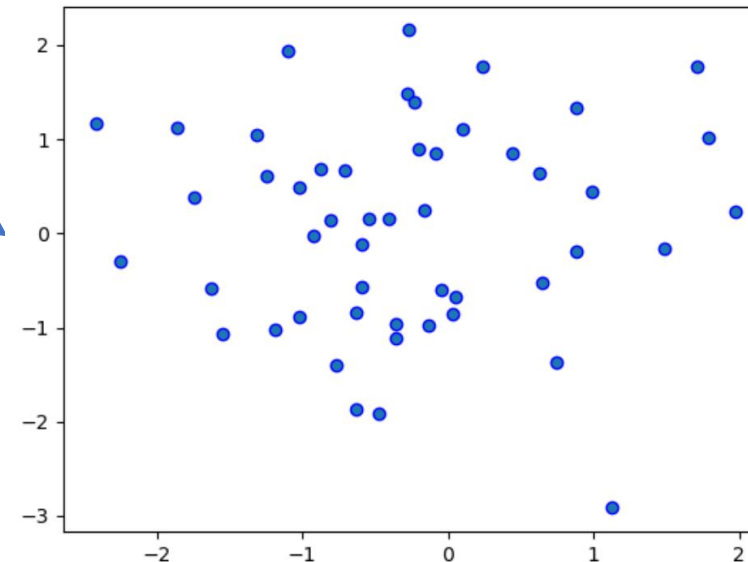


Supuestos:

- * Muestras aleatorias.
- * Conozco la distribución de los datos.

¿Cómo conozco sus parámetros μ y σ ?

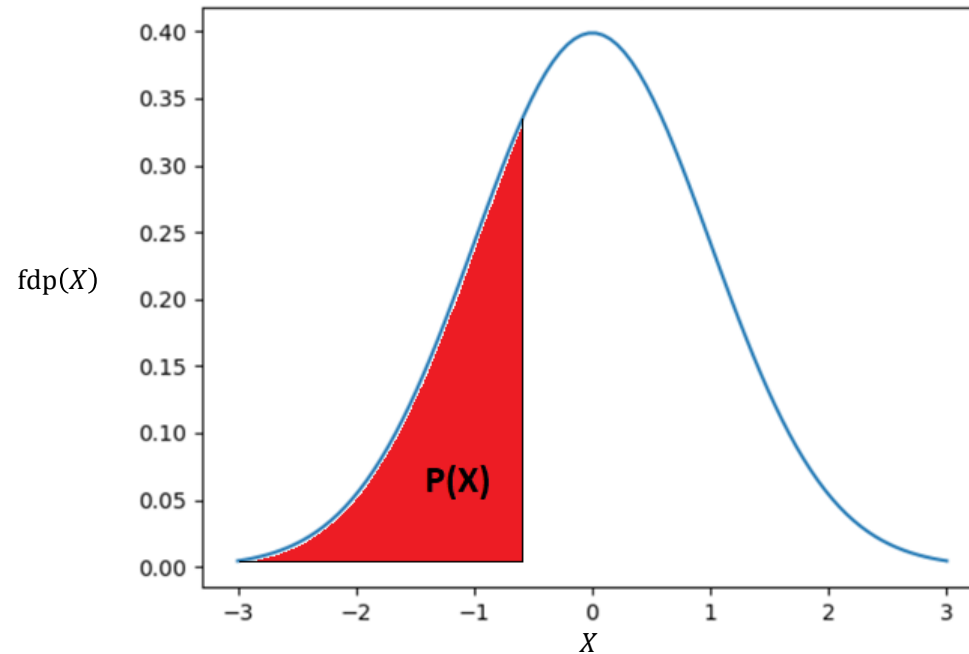
Muestra



Probabilidad

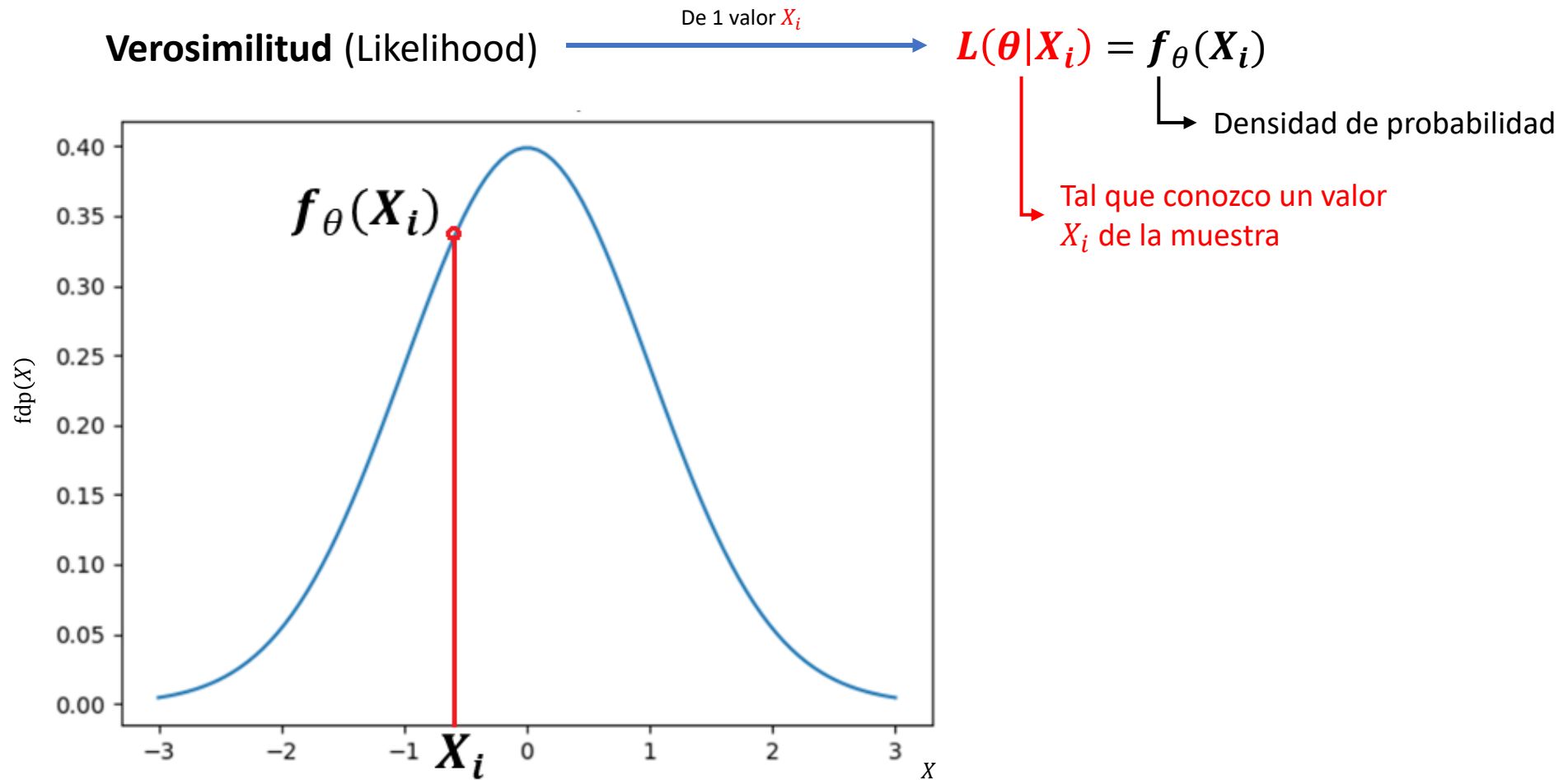
Recordemos: probabilidad de X $\rightarrow P(X) = P(X|\theta)$

Tal que conozco sus parámetros θ



*fdp: función de densidad de probabilidad.

Verosimilitud (likelihood)



Sample Likelihood

Verosimilitud (Likelihood)

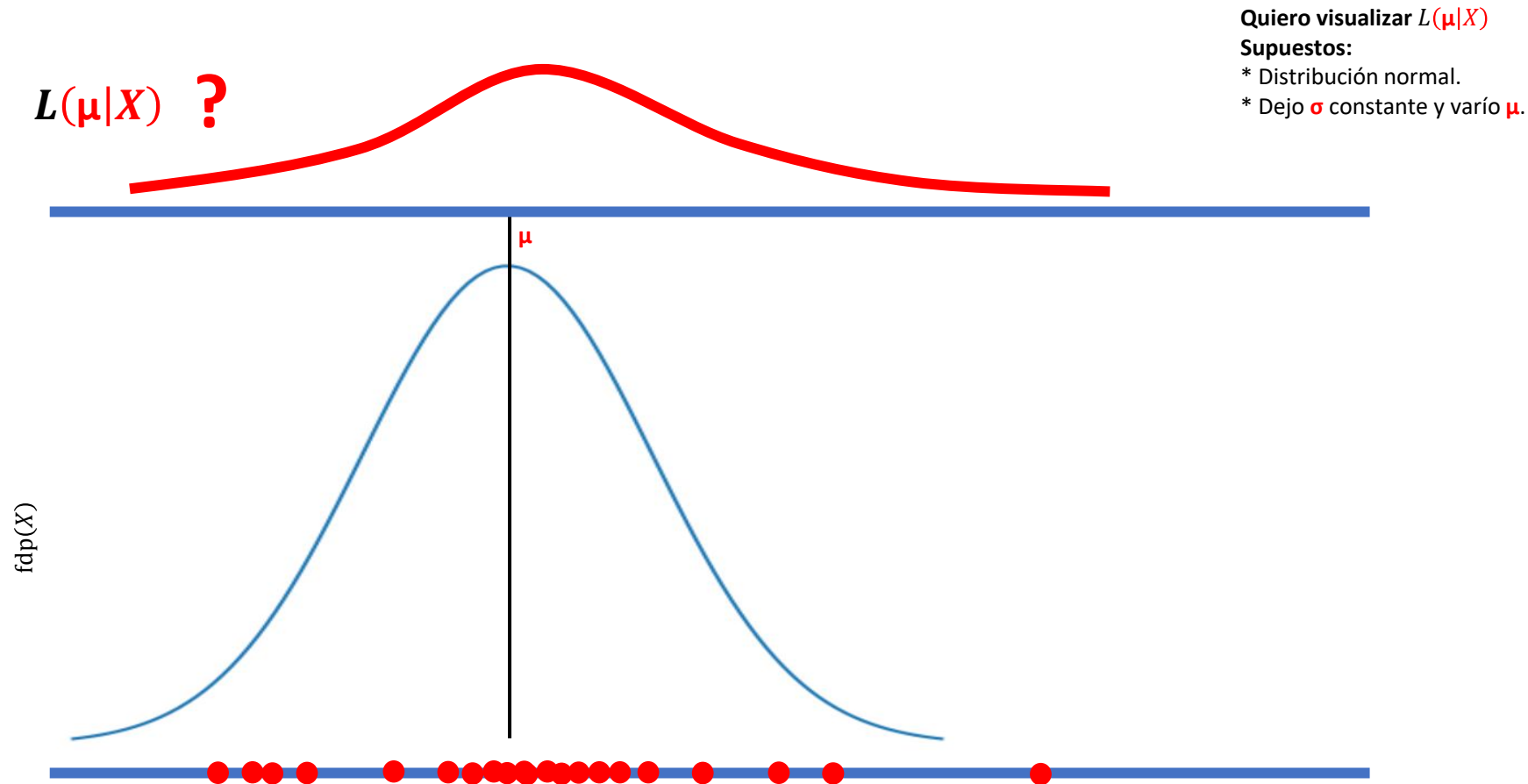
De la muestra:

$$L(\theta|X_1, X_2, \dots, X_n)$$

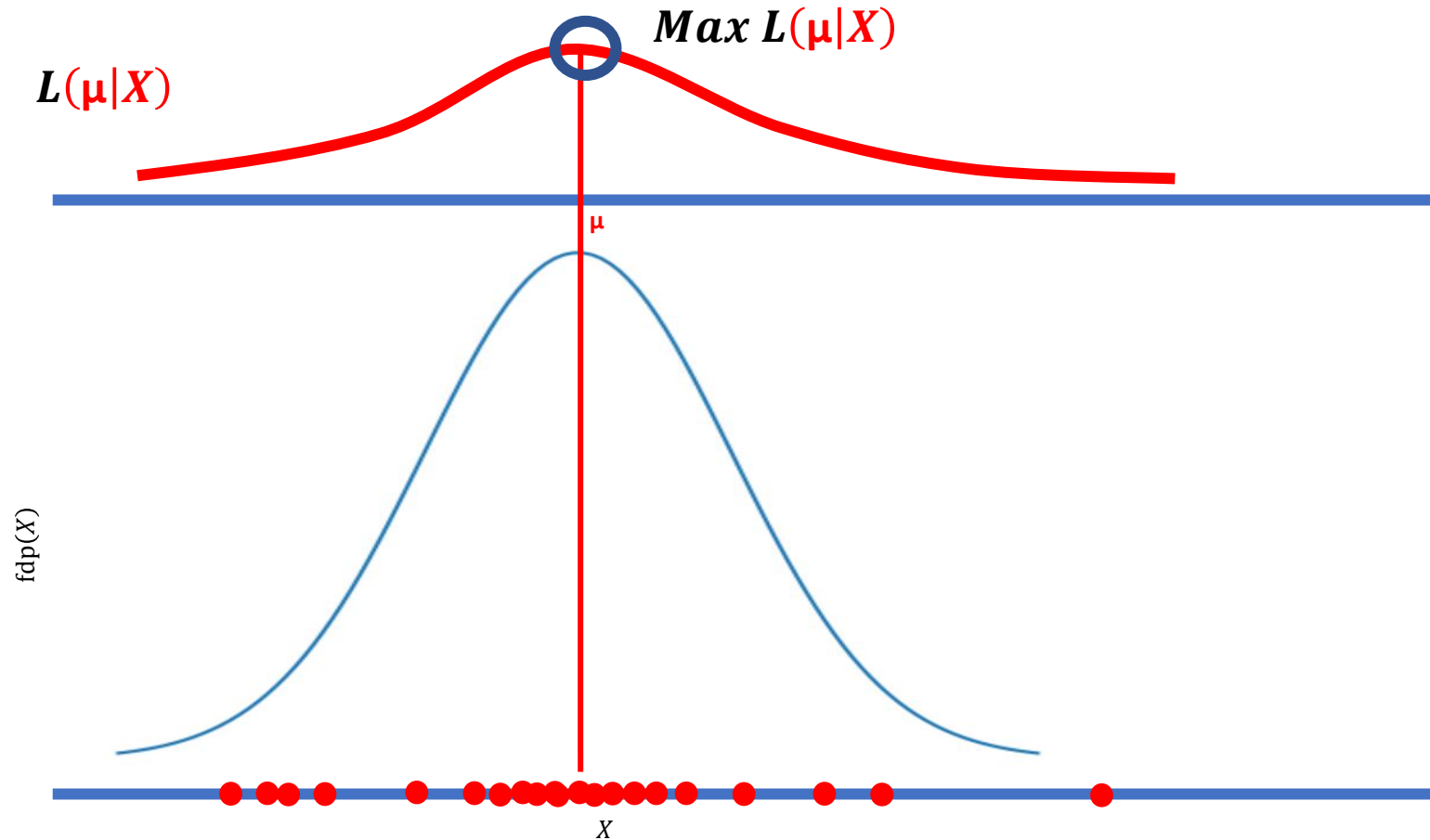
$$= L(\theta|X_1)L(\theta|X_2) \dots L(\theta|X_n)$$

$$= \prod_i^n L(\theta|X_i)$$

Visualización de función de Likelihood

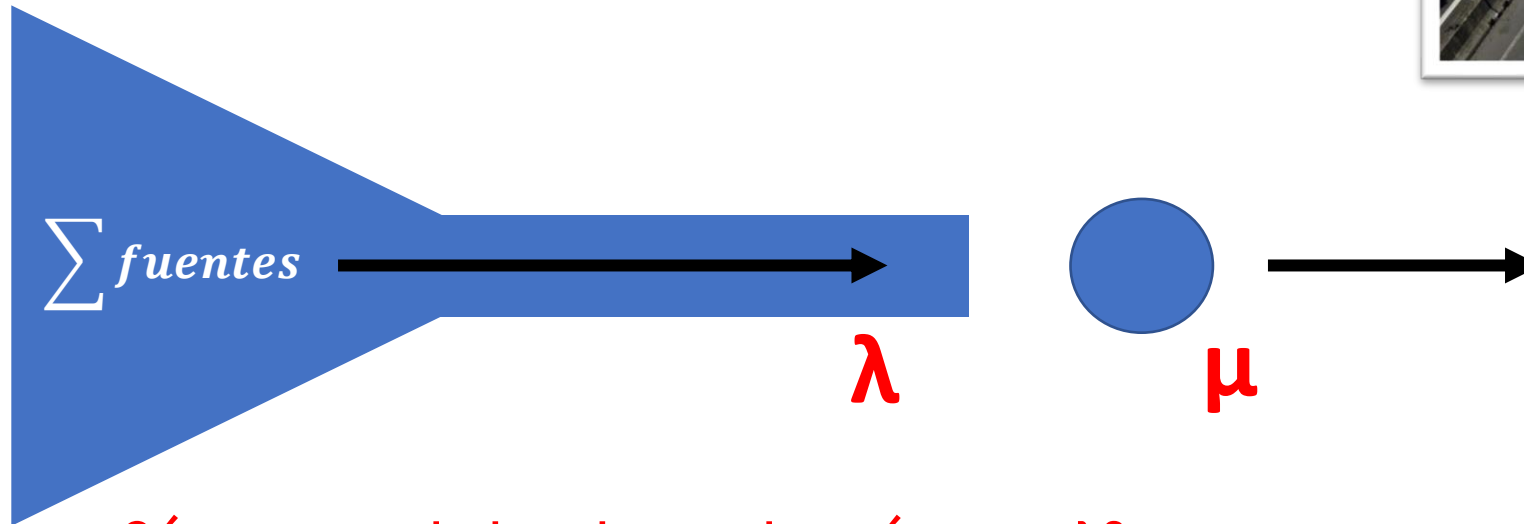


Visualización de función de Likelihood



Ejemplo: estimación de flujo vehicular

- Conozco el parámetro de servicio de agentes μ , ya que es un proceso controlable.
- En un control de tránsito, **necesito estimar el parámetro de llegada de vehículos λ** .
- La distribución de “tiempo entre llegadas” se supone exponencial.
- Puedo aplicar ajuste paramétrico de distribución.

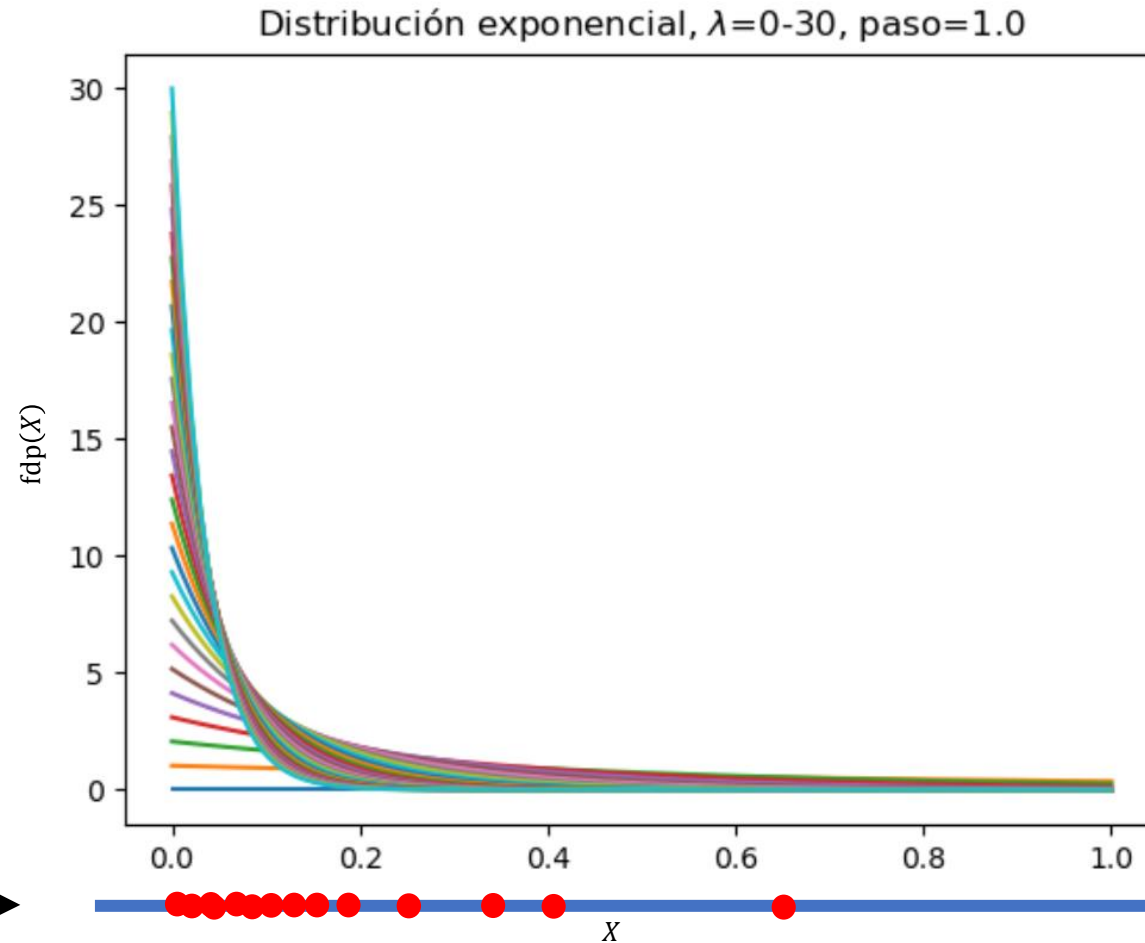


¿Cómo paso de los datos al parámetro λ ?

Distribución exponencial gráficamente

¿Cuál es el λ óptimo para esta distribución?

Muestra →



Solución analítica de Maximum Likelihood

Para este caso: **¡existe solución analítica de la Verosimilitud Máxima!**

$$L(\theta|X_i \dots X_n) = \prod_i^n L(\theta|X_i) \quad \text{Dado que: } L(\theta|X_i) = f_{\theta}(X_i)$$

En la exponencial:

$$f_{\beta}(X) = \beta e^{-\beta x} \quad \downarrow$$
$$L(\beta|X_i \dots X_n) = \prod_i^n L(\beta|X_i) = \prod_i^n \beta e^{-\beta x_i}$$

Solución analítica de Maximum Likelihood

$$L(\beta|X_i \dots X_n) = \prod_i^n \beta e^{-\beta x_i}$$

Queremos calcular:

$$\max L(\beta|X_i \dots X_n)$$

¿Cómo encontramos el máximo analíticamente?

$$\frac{d L(\beta|X_i \dots X_n)}{d\beta} = 0$$

Solución analítica de Maximum Likelihood

$$\begin{aligned}L(\beta|X_i \dots X_n) &= \prod_i^n \beta e^{-\beta x_i} \\&= \beta e^{-\beta x_1} \beta e^{-\beta x_2} \dots \beta e^{-\beta x_n} \\&= \beta^n e^{-\beta x_1 - \beta x_2 - \dots - \beta x_n}\end{aligned}$$

$$L(\beta|X_i \dots X_n) = \beta^n e^{-\beta(x_1 + x_2 + \dots + x_n)}$$

$$\frac{dL(\beta|X_i \dots X_n)}{d\beta} = \frac{d[\beta^n e^{-\beta(x_1 + x_2 + \dots + x_n)}]}{d\beta} = 0$$

Solución analítica de Maximum Likelihood

Sabemos que:

$$\max L(\beta|X_i \dots X_n) = \max \log L(\beta|X_i \dots X_n)$$

¡Las propiedades del logaritmo facilitan la optimización!

$$\log L(\beta|X_i \dots X_n) = \log \beta^n e^{-\beta(x_1+x_2+\dots+x_n)}$$

$$= \log \beta^n + \log e^{-\beta(x_1+x_2+\dots+x_n)}$$

$$\log L(\beta|X_i \dots X_n) = n \log \beta - \beta(x_1 + x_2 + \dots + x_n)$$

Solución analítica de Maximum Likelihood

$$\frac{dL(\beta|X_i \dots X_n)}{d\beta} = \frac{d(n \log \beta - \beta(x_1 + x_2 + \dots + x_n))}{d\beta} = 0$$

$$= \frac{n}{\beta} - (x_1 + x_2 + \dots + x_n) = 0$$

$$\beta = \frac{(x_1 + x_2 + \dots + x_n)}{n}$$

$$\lambda = \frac{1}{\beta}$$

En otras distribuciones

- ¡No todas son tan simples!
- No todas tienen solución analítica
- Uso de métodos numéricos de ajuste -> Distribución Beta
- Lo importante es: **conocer cómo armar la Función de densidad**

Resumen

- Puente entre **datos** -> **parámetros** de una Distribución
- Ajuste de datos a una densidad de probabilidad **conocida**
- Parámetros **desconocidos**

Ajuste en Python

Distribución Beta:

```
from scipy.stats import beta  
  
a, b, loc, scale = beta.fit(x)
```

**** Documentación: Solución numérica, óptimos locales!**