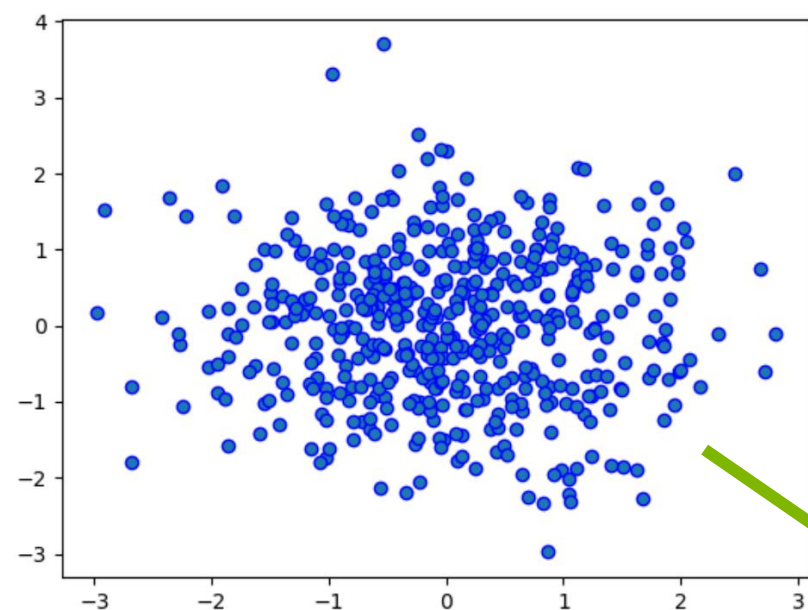


De los Datos a los Modelos: ¿Cómo ajustar Distribuciones? Clase 06

Investigación Operativa UTN FRBA 2020

Elaborado por: Rodrigo Maranzana

Curso: I4051 (Prof. Martin Palazzo)

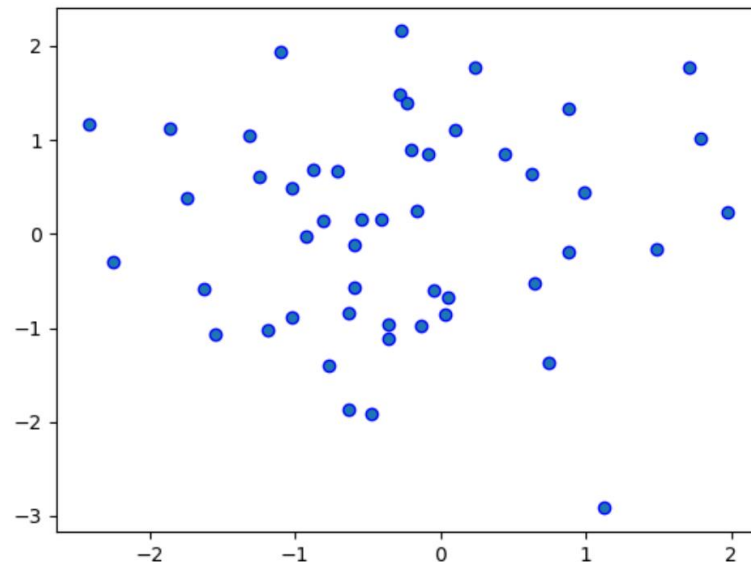


Población con distribución Normal

Supuestos:

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

Muestra



¿Cómo conozco sus parámetros μ y σ ?

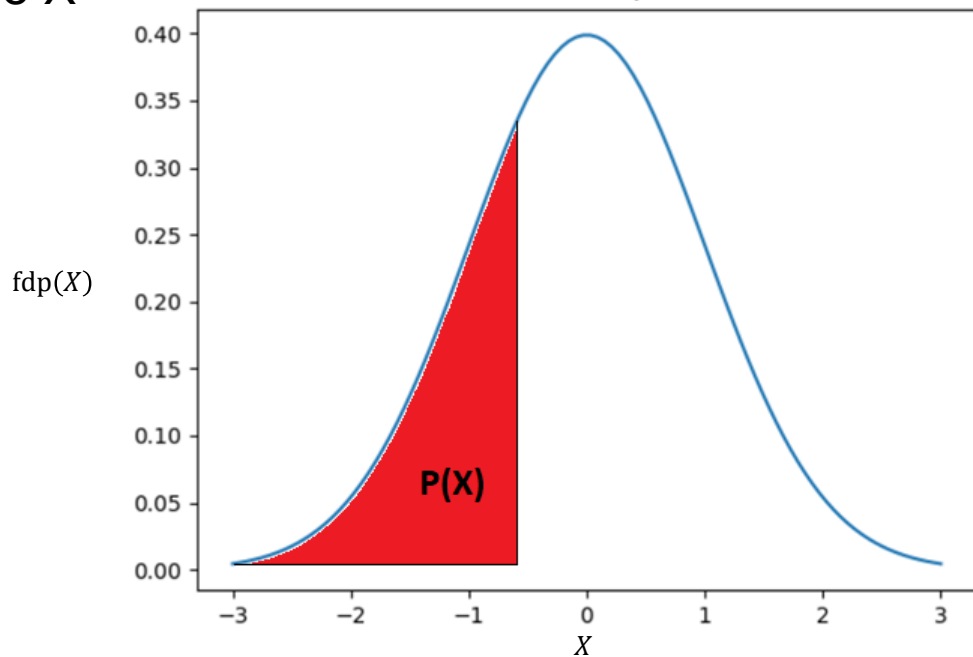
Recordemos: probabilidad
de X



$$P(X) = P(X|\theta)$$



Tal que conozco sus parámetros θ



*fdp: función de densidad de probabilidad.

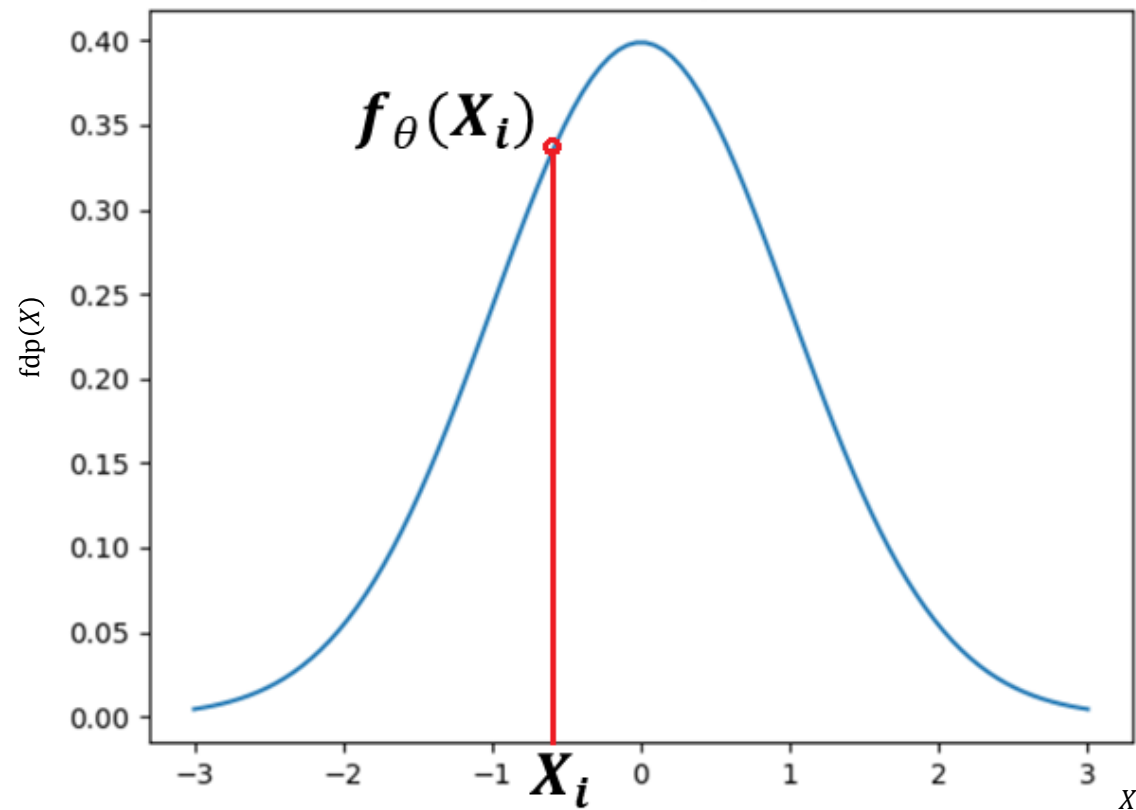
Verosimilitud (Likelihood)

De 1 valor X_i

$$L(\theta|X_i) = f_{\theta}(X_i)$$

Densidad de probabilidad

Tal que conozco un valor
 X_i de la muestra



Verosimilitud (Likelihood)

De la muestra:

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

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

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

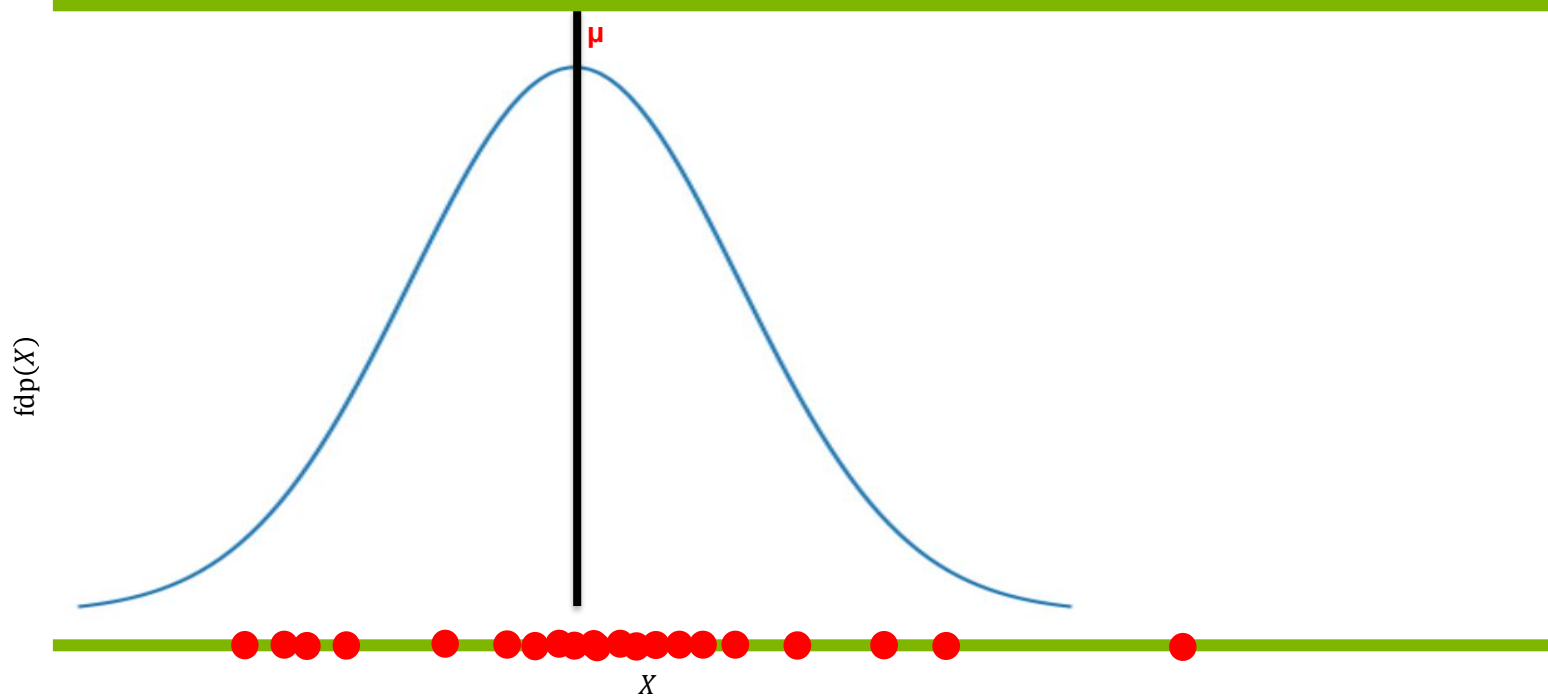
$L(\mu|X)$?

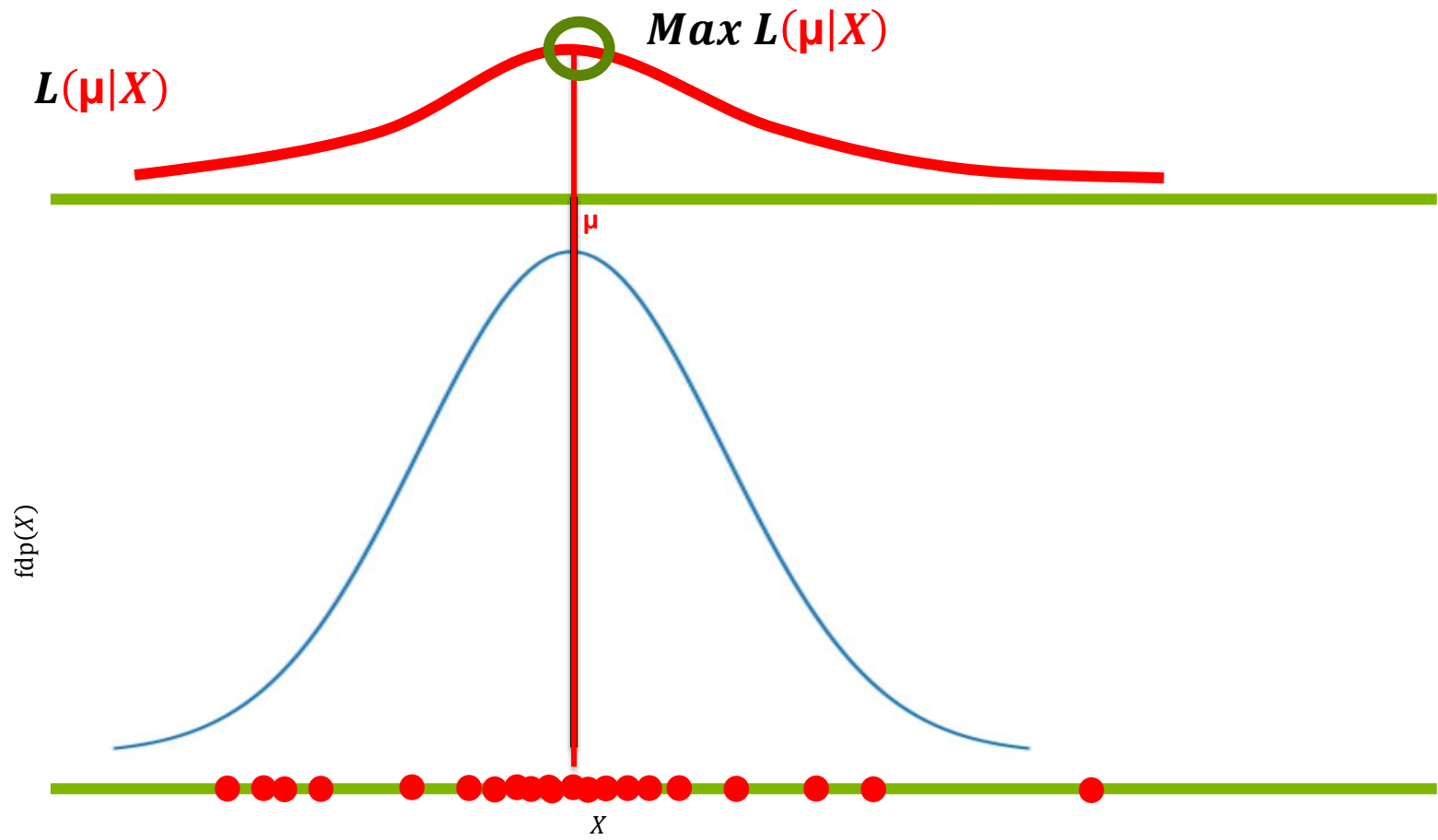
Quiero visualizar $L(\mu|X)$

Supuestos:

* Distribución normal.

* Dejo σ constante y varío μ .





Cuarentena: Control en Panamericana hacia CABA



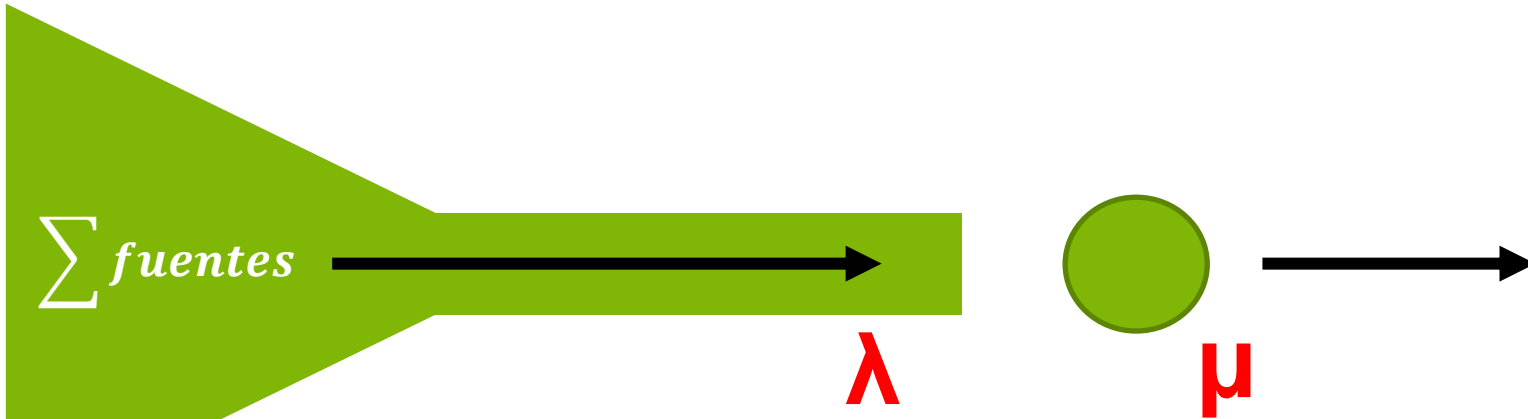
Necesito cuantificar el CAOS.

No puedo solucionarlo pero puedo adelantarlo.

Mejorar la toma de decisiones.

Datos:

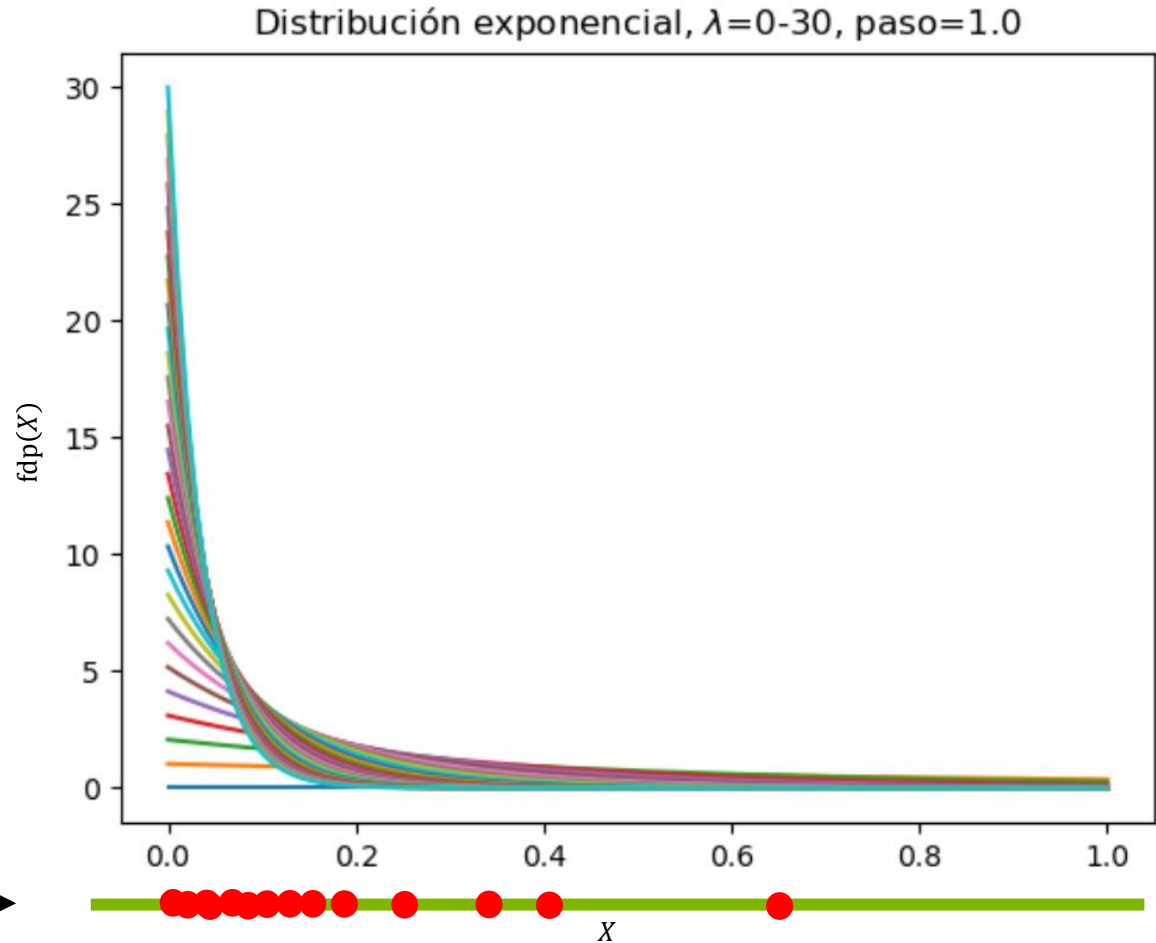
Patentes de autos y hora en que entraron a CABA en meses normales.
Parámetro de servicio ya medido, fácil de obtener.



¿Cómo paso de los datos al parámetro λ ?

¿Cuál es el λ
óptimo
para esta
distribución?

Muestra →




Caso: Distribución exponencial

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:

$$f_\lambda(X) = \lambda e^{-\lambda x}$$

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

Caso: Distribución exponencial

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

Queremos calcular:

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

¿Cómo encontramos el máximo analíticamente?

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

Caso: Distribución exponencial

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

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

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

Caso: Distribución exponencial

Sabemos que:

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

¡Las propiedades del logaritmo facilitan la optimización!

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

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

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

Caso: Distribución exponencial

$$\frac{dL(\lambda|X_1 \dots X_n)}{d\lambda} = \frac{d(n \log \lambda - \lambda(x_1 + x_2 + \dots + x_n))}{d\lambda} = 0$$

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

$$\lambda = \frac{(x_1 + x_2 + \cdots + x_n)}{n}$$

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**

En resumen: ¿Qué hicimos?

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

En Python:

Distribución Beta:

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

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

https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.rv_continuous.fit.html