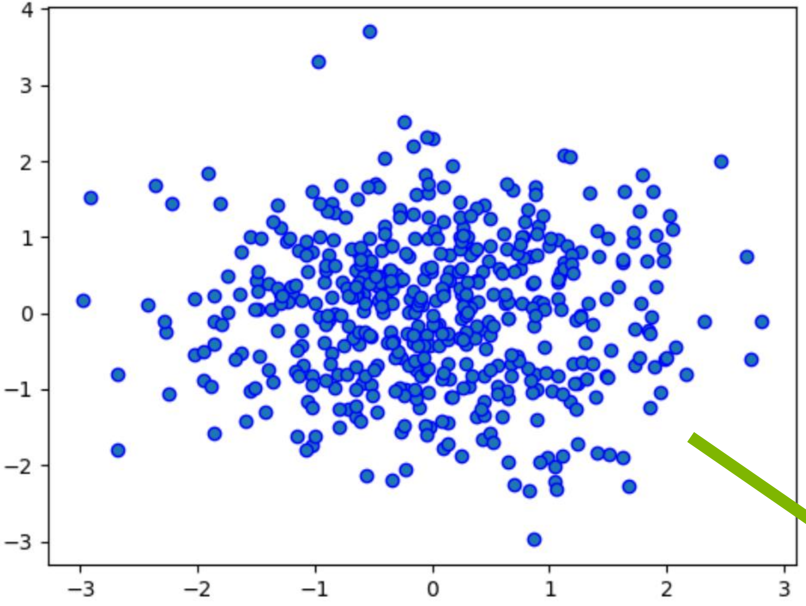
De los Datos a los Modelos: ¿Cómo ajustar Distribuciones?
Clase 06

Investigación Operativa UTN FRBA 2020
Curso: I4051
Elaborado por: Rodrigo Maranzana
Docente: Martín Palazzo

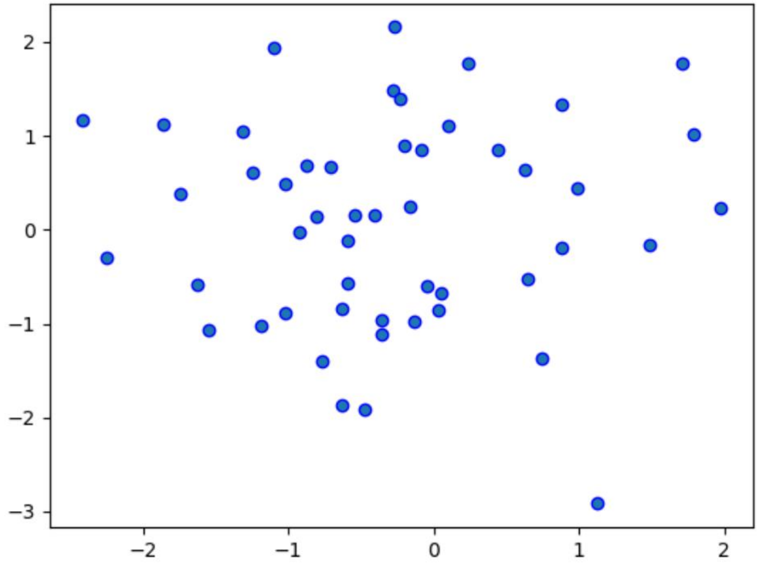
Supuestos:

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



Población con distribución Normal

Muestra

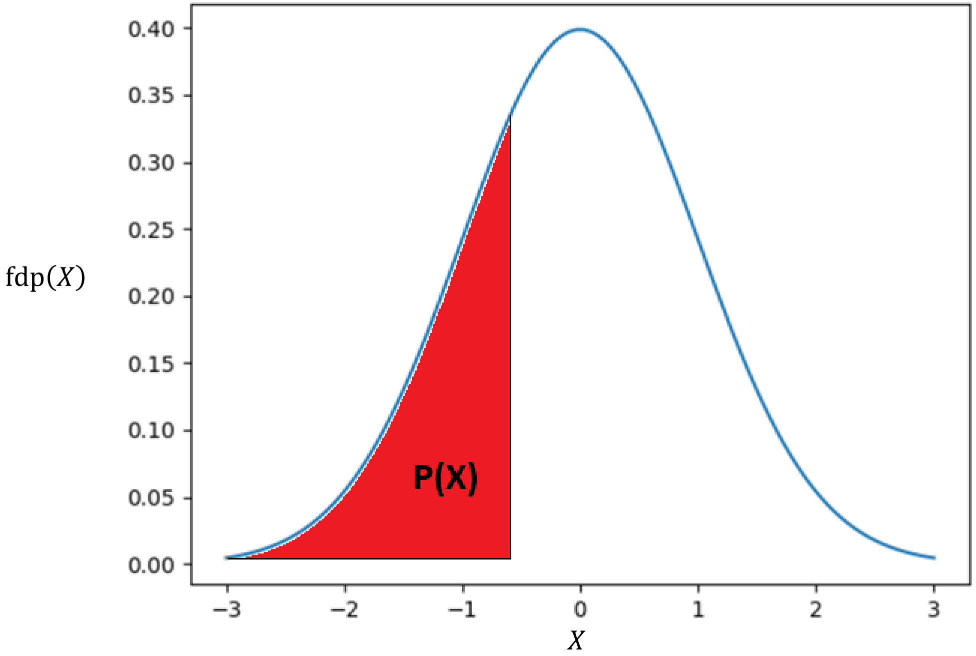


¿Cómo conozco sus parámetros μ y σ?

Recordemos: probabilidad de X

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

Tal que conozco sus parámetros θ

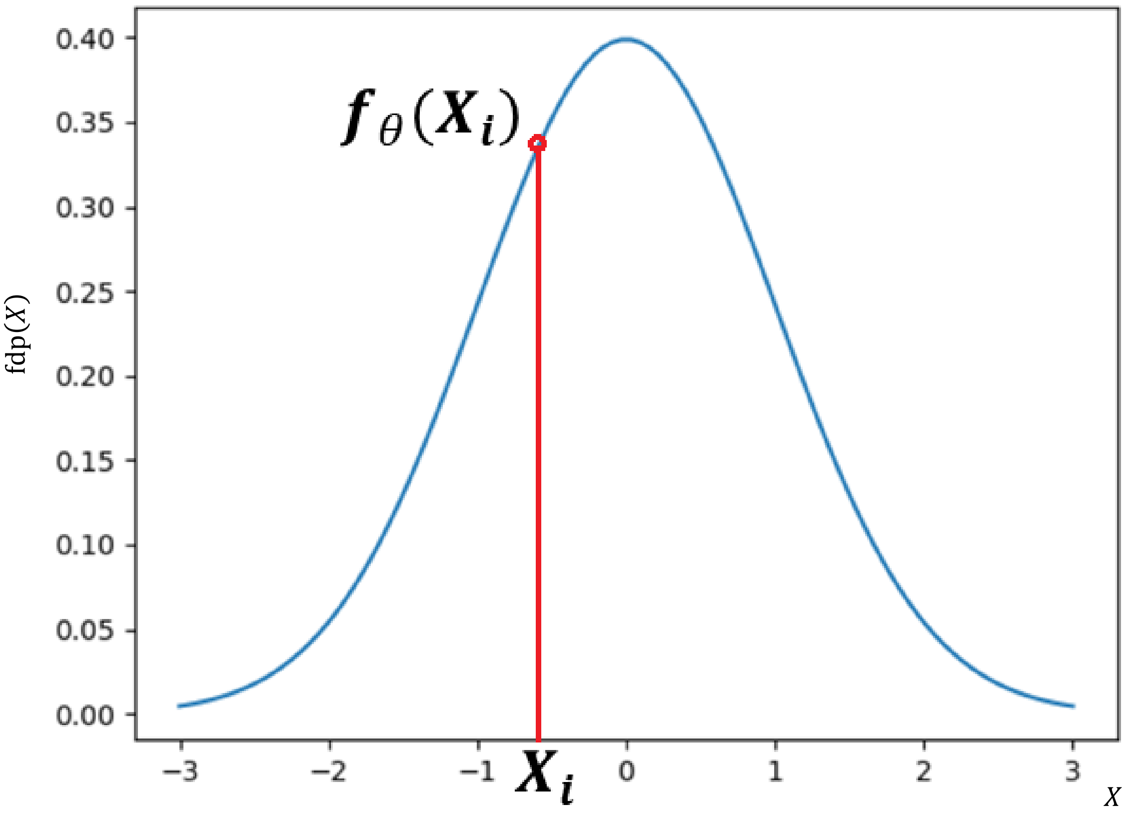


*fdp: función de densidad de probabilidad.

Verosimilitud (Likelihood) $\xrightarrow{\text{De 1 valor } X_i}$ $\boldsymbol{L(\theta|X_i) = f_\theta(X_i)}$

Densidad de probabilidad

Tal que conozco un valor X_i de la muestra



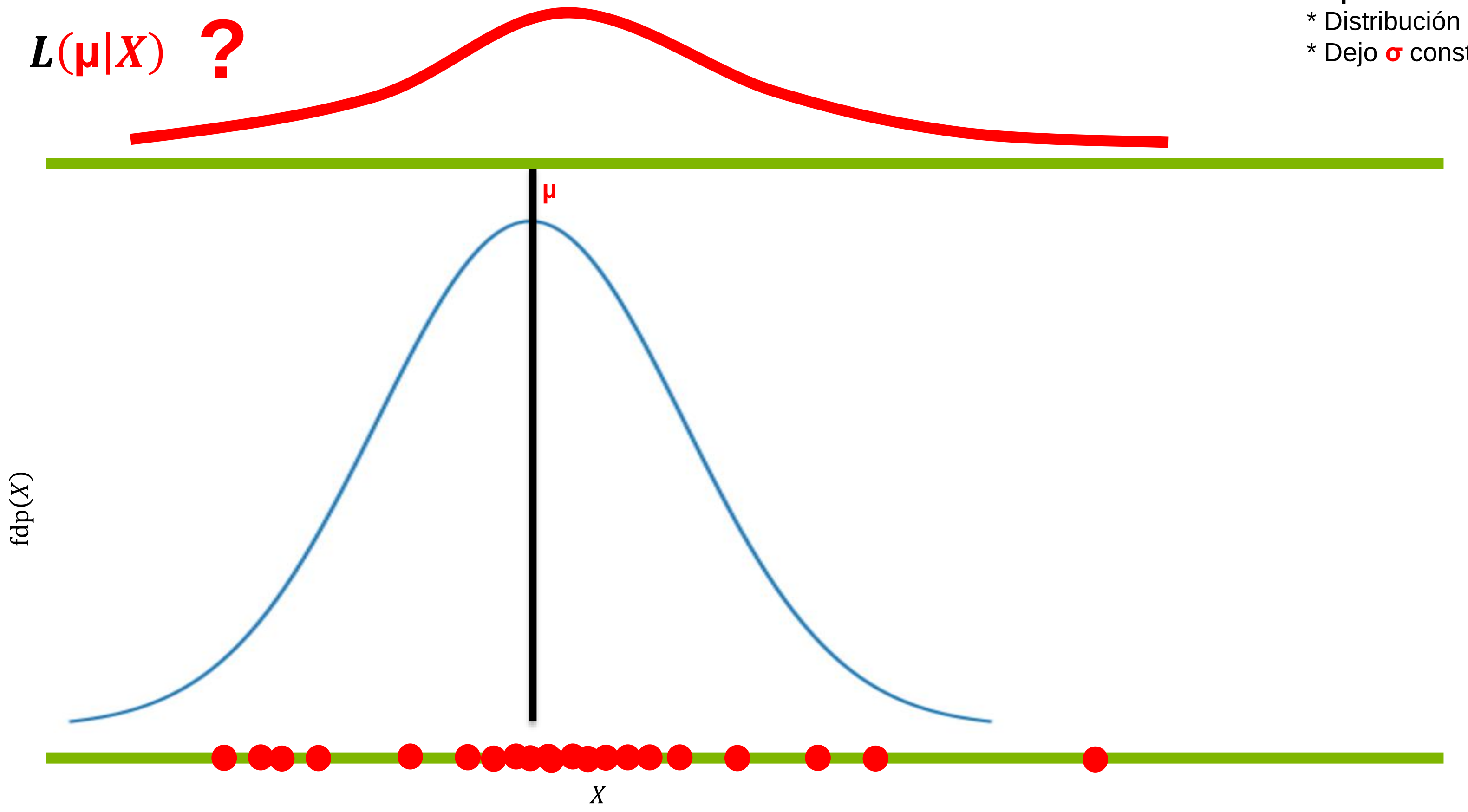
Verosimilitud (Likelihood)

De la muestra:

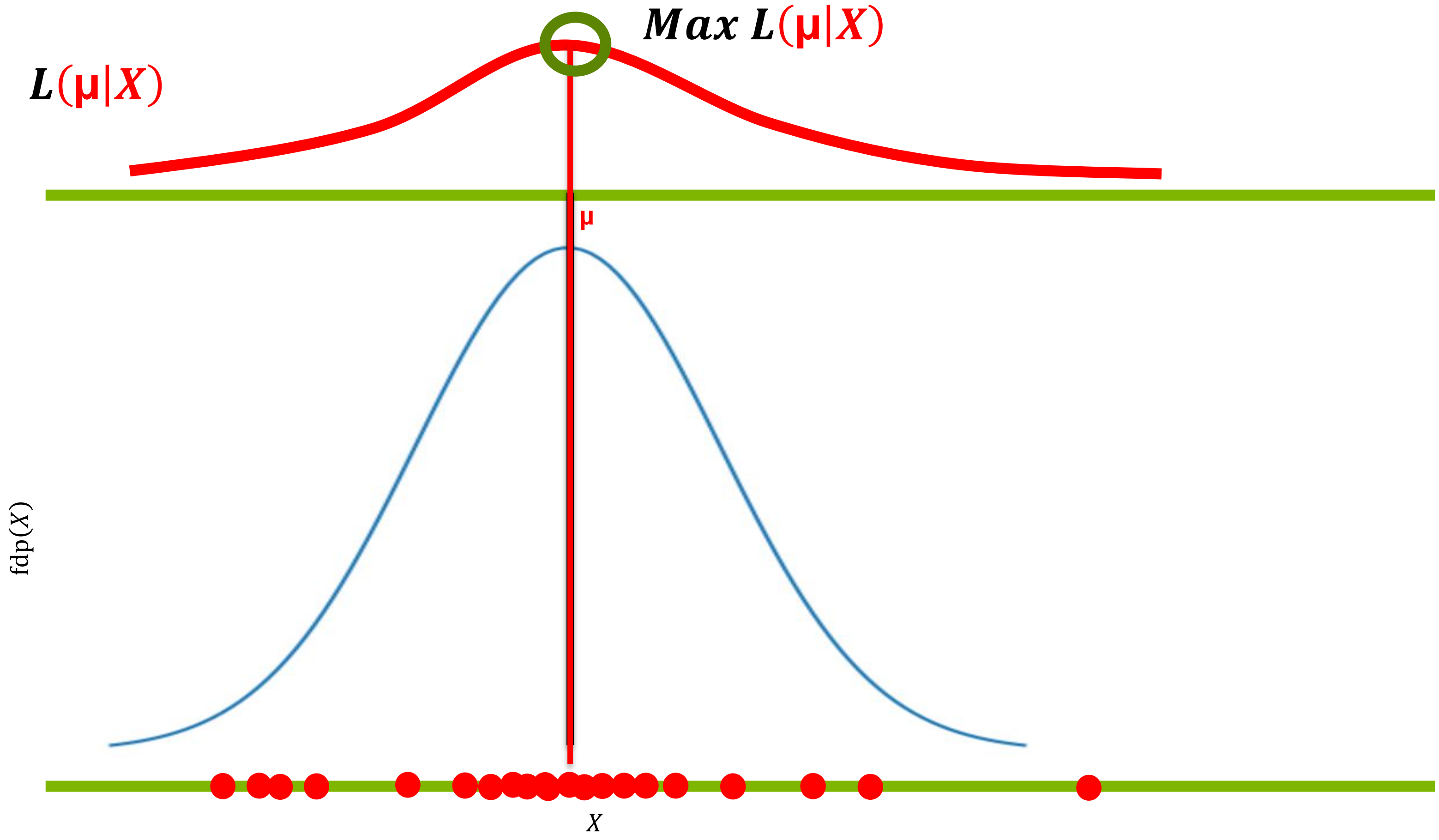
$$L(\theta|X_1, X_2, \dots, X_n)$$

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

$$= \prod_{i}^{n} L(\theta|X_i)$$



Quiero visualizar $L(\mu|X)$
Supuestos:
* Distribución normal.
* Dejo **σ** constante y varío **μ**.



Max L(μ|X)
L(μ|X)
μ
fdp(X)
X

hacia CABA

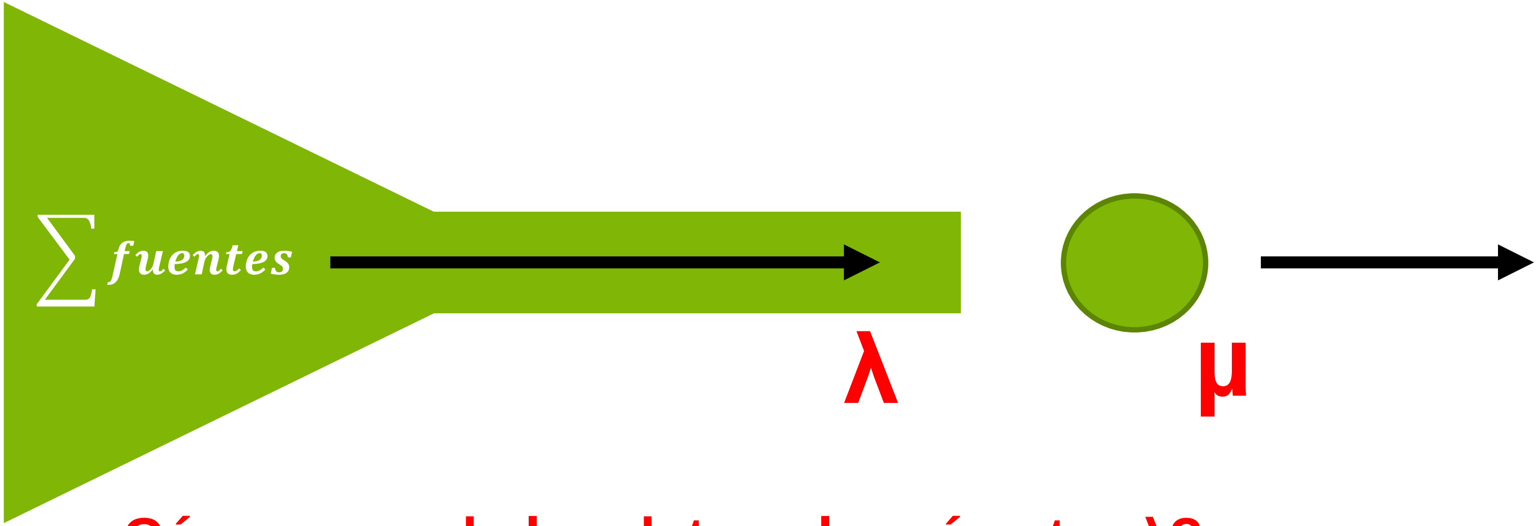
Necesito cuantificar el CAOS.

No puedo solucionarlo pero puedo adelantarlo.

Mejorar la toma de decisiones.

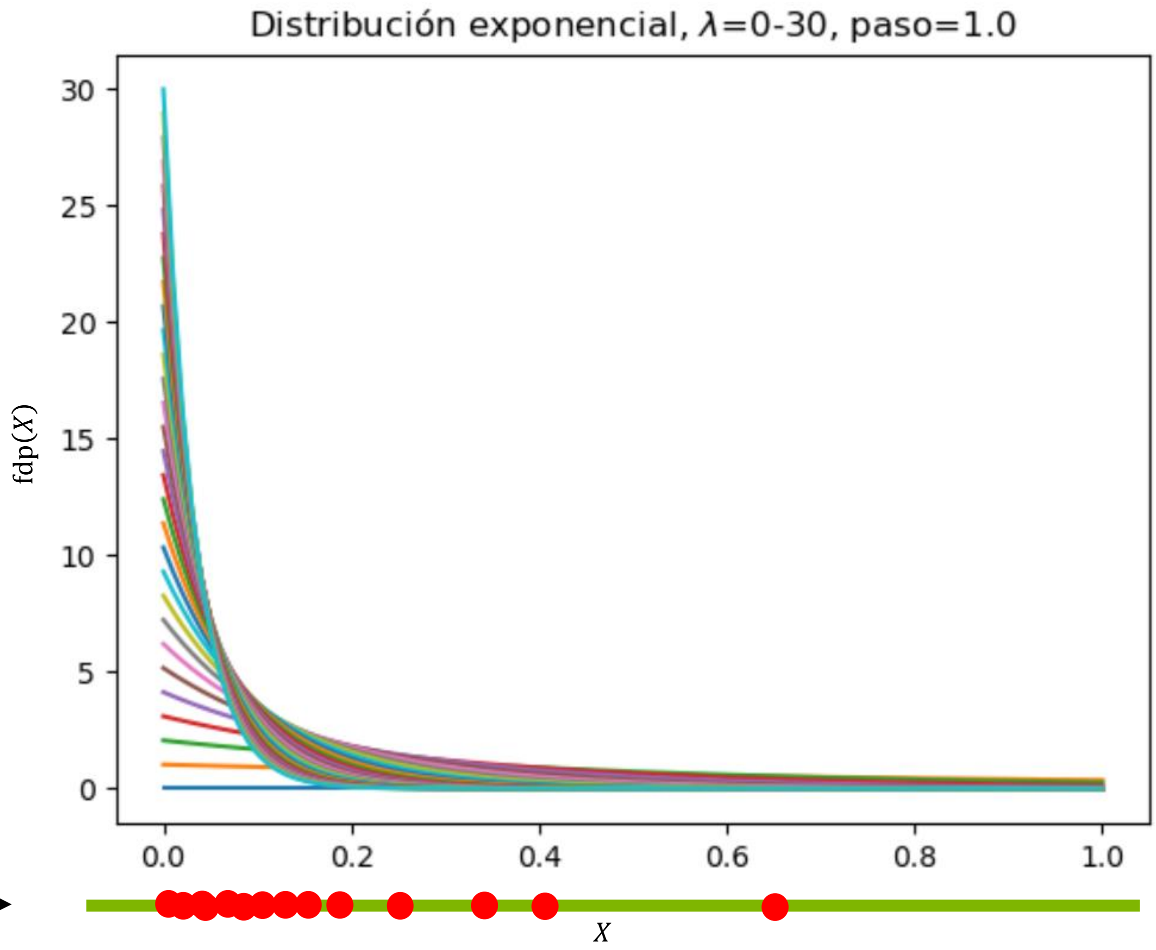
Datos:

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 Máxi**

$$L(\theta|X_i \dots X_n) = \prod_{i}^{n} L(\theta|X_i) \qquad Dado\ que: L(\theta|X_i) = \boldsymbol{f_\theta(X_i)}$$

En la exponencial:

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

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

Caso: Distribución exponencial

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

Queremos calcular:

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

¿Cómo encontramos el máximo analíticamente?

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

Caso: Distribución exponencial

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

$$= \boldsymbol{\lambda e^{-\lambda x_1} \lambda e^{-\lambda x_2} \dots \lambda e^{-\lambda x_n}}$$

$$= \boldsymbol{\lambda^n e^{-\lambda x_1 - \lambda x_2 - \cdots - \lambda x_n}}$$

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

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

Caso: Distribución exponencial

Sabemos que:

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

¡Las propiedades del logaritmo facilitan la optimización!

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

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

$$\log L(\lambda|X_i \ldots X_n) = \mathrm{n} \log \boldsymbol{\lambda} - \boldsymbol{\lambda}(\boldsymbol{x_1} + \boldsymbol{x_2} + \cdots + \boldsymbol{x_n})$$

Caso: Distribución exponencial

$$\frac{dL(\lambda|X_i \ldots X_n)}{d\lambda} = \frac{d(\mathrm{n} \log \lambda - \lambda(x_1 + x_2 + \cdots + x_n))}{d\lambda} = 0$$

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

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

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