Alberto Andrés Valdés González.
Degree: Mathematical Engineer.
Work position: Data Scientist.
Mail: anvaldes@uc.cl/alberto.valdes.gonzalez.96@gmail.com
Location: Santiago, Chile.

Maximum Likelihood Estimators

Remembering the Bayes theorem we have:

$$\boxed{\mathbb{P}(A \cap B) = \mathbb{P}(A|B) \cdot \mathbb{P}(B)}$$

Using $A = \vec{x}$ and $B = \theta$, we have that:

$$\mathbb{P}(\Theta = \theta \cap X_1 = x_1, ..., X_n = x_n) = \mathbb{P}(X_1 = x_1, ..., X_n = x_n | \Theta = \theta) \cdot \mathbb{P}(\Theta = \theta)$$

We want to maximize $\mathbb{P}(\Theta = x, \vec{X} = \vec{x})$ but how we don't know anything about $\mathbb{P}(\Theta = \theta)$ we going to maximize $\mathbb{P}(\vec{X} = \vec{x} | \Theta = \theta)$.

We define the function:

$$\boxed{L(\theta|\vec{x}) = \mathbb{P}(X_1 = x_1, ..., X_n = x_n | \Theta = \theta)}$$

Note that:

$$\boxed{\text{argmax}_\theta \;\; L(\theta|\vec{x}) = \text{argmax}_\theta \;\; ln(L(\theta|\vec{x}))}$$

For this we will define:

$$\boxed{l(\theta) = ln(L(\theta|\vec{x}))}$$

We define the **maximum likelihood estimator** as follows:

$$\boxed{l(\theta_{MLE}) = \max_{\theta} l(\theta)}$$

Example: $X_i \overset{\text{iid}}{\sim} Exp(\lambda)$.

$$L(\lambda|\vec{x}) = \prod_{i=1}^{n} f(x_i|\lambda) = \prod_{i=1}^{n} \left[\left(\frac{1}{\lambda}\right) \cdot e^{-\left(\frac{1}{\lambda}\right)\cdot x_i}\right] = \left(\frac{1}{\lambda}\right)^n \cdot e^{-\left(\frac{1}{\lambda}\right)\cdot \sum_{i=1}^{n} x_i}$$

$\Rightarrow$

$$l(\lambda) = ln(L(\lambda|\vec{x})) = ln\left(\left(\frac{1}{\lambda}\right)^n \cdot e^{-\left(\frac{1}{\lambda}\right)\cdot \sum_{i=1}^{n} x_i}\right) = -n \cdot ln(\lambda) - \left(\frac{1}{\lambda}\right) \cdot \sum_{i=1}^{n} x_i$$

$\Rightarrow$

$$\boxed{l(\lambda) = -n \cdot ln(\lambda) - \left(\frac{1}{\lambda}\right) \cdot \sum_{i=1}^{n} x_i}$$

$\Rightarrow$

$$\frac{\partial l(\lambda)}{\partial \lambda} = -n \cdot \frac{1}{\lambda} + \frac{1}{\lambda^2} \cdot \sum_{i=1}^{n} x_i$$

Now we want:

$$\left.\frac{\partial l(\lambda)}{\partial \lambda}\right|_{\lambda=\hat{\lambda}} = 0$$

$\Rightarrow$

$$-n \cdot \frac{1}{\hat{\lambda}} + \frac{1}{\hat{\lambda}^2} \cdot \sum_{i=1}^{n} x_i = 0$$

$\Rightarrow$

$$\boxed{\hat{\lambda} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i}$$

Example: $X_i \overset{\text{iid}}{\sim} Poisson(\lambda)$.

$$L(\lambda|\vec{x}) = \prod_{i=1}^{n} f(x_i|\lambda) = \prod_{i=1}^{n} \left[\frac{e^{-\lambda} \cdot \lambda^{x_i}}{(x_i)!}\right] = \frac{e^{-n\cdot\lambda} \cdot \lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} (x_i)!}$$

$\Rightarrow$

$$l(\lambda) = ln(L(\lambda|\vec{x})) = -n \cdot \lambda + \left(\sum_{i=1}^{n} x_i\right) \cdot ln(\lambda) - \sum_{i=1}^{n} ln[(x_i)!\ ]$$

$\Rightarrow$

$$\boxed{l(\lambda) = -n \cdot \lambda + \left(\sum_{i=1}^{n} x_i\right) \cdot ln(\lambda) - \sum_{i=1}^{n} ln[(x_i)!\ ]}$$

$\Rightarrow$

$$\frac{\partial l(\lambda)}{\partial \lambda} = -n + \left(\sum_{i=1}^{n} x_i\right) \cdot \frac{1}{\lambda}$$

Now we want:

$$\frac{\partial l(\lambda)}{\partial \lambda}\Big|_{\lambda=\hat{\lambda}} = 0$$

$\Rightarrow$

$$\boxed{\hat{\lambda} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i}$$

Example: $X_i \stackrel{\text{iid}}{\sim} Bernoulli(p)$.

$$L(p|\vec{x}) = \prod_{i=1}^{n} f(x_i|p) = \prod_{i=1}^{n} [p^{x_i} \cdot (1-p)^{1-x_i}] = p^{\sum_{i=1}^{n} x_i} \cdot (1-p)^{n-\sum_{i=1}^{n} x_i}$$

$\Rightarrow$

$$l(p) = ln(L(p|\vec{x})) = \left(\sum_{i=1}^{n} x_i\right) \cdot ln(p) + \left(n - \sum_{i=1}^{n} x_i\right) \cdot ln(1-p)$$

$\Rightarrow$

$$\boxed{l(p) = \left(\sum_{i=1}^{n} x_i\right) \cdot ln(p) + \left(n - \sum_{i=1}^{n} x_i\right) \cdot ln(1-p)}$$

$\Rightarrow$

$$\frac{\partial l(p)}{\partial p} = \left(\sum_{i=1}^{n} x_i\right) \cdot \frac{1}{p} - \left(n - \sum_{i=1}^{n} x_i\right) \cdot \frac{1}{1-p}$$

Now we want:

$$\frac{\partial l(p)}{\partial p}\Big|_{p=\hat{p}} = 0$$

$\Rightarrow$

$$\left(\sum_{i=1}^{n} x_i\right) \cdot \frac{1}{\hat{p}} - \left(n - \sum_{i=1}^{n} x_i\right) \cdot \frac{1}{1-\hat{p}} = 0$$

$\Rightarrow$

$$\left(\sum_{i=1}^{n} x_i\right) \cdot (1-\hat{p}) = \left(n - \sum_{i=1}^{n} x_i\right) \cdot \hat{p}$$

$\Rightarrow$

$$\left(\sum_{i=1}^{n} x_i\right) - \hat{p} \cdot \left(\sum_{i=1}^{n} x_i\right) = n \cdot \hat{p} - \left(\sum_{i=1}^{n} x_i\right) \cdot \hat{p}$$

$\Rightarrow$

$$\left(\sum_{i=1}^{n} x_i\right) = n \cdot \hat{p}$$

$\Rightarrow$

$$\boxed{\hat{p} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i}$$

Example: $X_i \overset{\text{iid}}{\sim} Normal(\mu, \sigma)$.

$$L(\mu, \sigma|\vec{x}) = \prod_{i=1}^{n} f(x_i|\mu, \sigma) = \prod_{i=1}^{n} \left[\frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot exp\left(-\frac{1}{2} \cdot \left[\frac{x_i - u}{\sigma} \right]^2 \right) \right]$$

$$= (2\pi)^{-\frac{n}{2}} \cdot \sigma^{-n} \cdot exp\left(-\frac{1}{2} \cdot \sum_{i=1}^{n} \left[\frac{x_i - u}{\sigma} \right]^2 \right)$$

$\Rightarrow$

$$l(\mu, \sigma) = ln(L(\mu, \sigma|\vec{x})) = -\frac{n}{2} \cdot ln(2\pi) - n \cdot ln(\sigma) - \frac{1}{2} \cdot \sum_{i=1}^{n} \left[\frac{x_i - u}{\sigma} \right]^2$$

$\Rightarrow$

$$\boxed{l(\mu, \sigma) = -\frac{n}{2} \cdot ln(2\pi) - n \cdot ln(\sigma) - \frac{1}{2} \cdot \sum_{i=1}^{n} \left[\frac{x_i - u}{\sigma} \right]^2}$$

$\Rightarrow$

$$\frac{\partial l(\mu, \sigma)}{\partial \mu} = \frac{1}{\sigma^2} \cdot \sum_{i=1}^{n} [x_i - \mu] = \frac{1}{\sigma^2} \cdot \left(\sum_{i=1}^{n} x_i - n \cdot \mu \right)$$

$$\frac{\partial l(\mu, \sigma)}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \cdot \sum_{i=1}^{n} [x_i - \mu]^2 = \frac{1}{\sigma} \cdot \left[-n + \frac{1}{\sigma^2} \cdot \sum_{i=1}^{n} (x_i - \mu)^2 \right]$$

Now we want:

$$\left. \frac{\partial l(\mu, \sigma)}{\partial \mu} \right|_{\mu = \hat{\mu}} = 0 \qquad \left. \frac{\partial l(\mu, \sigma)}{\partial \sigma} \right|_{\sigma = \hat{\sigma}} = 0$$

$\Rightarrow$

$$\boxed{\hat{\mu} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i} \qquad \boxed{\hat{\sigma^2} = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2}$$
