Example

Suppose $x \mapsto f(x;\theta)$ is the density of a uniform random variable on $[0,\theta]$. We observe four samples drawn from this distribution: $1.41, 2.45, 6.12$, and 4.9. Find $\mathcal{L}(5)$, $\mathcal{L}(10^6)$, and $\mathcal{L}(7)$.

$\mathcal{L}(5) = 0$. Because $f_5(\underset{6.12}{X_3}) = 0$ on $[0,5]$

$$\mathcal{L}(10^6) = \left(\frac{1}{10^6}\right)^4 = 10^{-24}.$$

$$\mathcal{L}(7) = \left(\frac{1}{7}\right)^4 = \frac{1}{2401}$$

$$\mathcal{L}_x(\theta) = f_\theta(x_1) \cdot f_\theta(x_2) \cdots f_\theta(x_n)$$

As illustrated in this example, likelihood has the property of being zero or small at implausible values of $\boldsymbol{\theta}$, and larger at more reasonable values. Thus we propose the **maximum likelihood estimator**

$$\widehat{\boldsymbol{\theta}}_{\mathrm{MLE}} = \operatorname{argmax}_{\boldsymbol{\theta} \in \mathbb{R}^d} \mathcal{L}_{\mathbf{X}}(\boldsymbol{\theta}).$$

Example

Suppose that $x \mapsto f(x;\mu,\sigma^2)$ is the normal density with mean μ and variance σ^2. Find the maximum likelihood estimator for μ and σ^2.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$\log(f(x_i)) = -\log\sigma - \frac{1}{2}\cdot\log 2\pi - \frac{1}{2}\cdot\left(\frac{x_i-\mu}{\sigma^2}\right)^2$$

$$\downarrow$$

$$\log(\mathcal{L}(\mu,\sigma^2)) = \log\left(\prod_{i=1}^{n}\mathcal{L}(x_i,\mu,\sigma^2)\right) = \sum_{i=1}^{n}\log\mathcal{L}(x_i,\mu,\sigma^2)$$

$$= -n\log\sigma - \frac{n}{2}\log 2\pi - \frac{1}{2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^2}$$

$$= -\frac{n}{2}\log\sigma^2 - \frac{n}{2}\log 2\pi - \frac{1}{2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^2}$$

$$\frac{\partial\log(\mathcal{L}(x_1\ldots x_n;\mu,\sigma^2))}{\partial\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{(\sigma^2)^2}$$

$$= -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\cdot\sum_{i=1}^{n}(x_i-\mu)^2 = 0$$

①

$$\frac{\partial\log(\mathcal{L}(x_1\ldots x_n);\mu,\sigma^2))}{\partial\mu} = \sum_{i=1}^{n}\frac{(x_i-\mu)}{\sigma^2} = 0 \Rightarrow \sum_{i=1}^{n}x_i = n\mu$$

$$\mu = \bar{x}$$

$$① \Rightarrow \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2,$$

log likelihood function monotoney Increase

$\Downarrow$

Maximum Point

plug in

$\bar{x} = \mu$

$$\frac{\partial\log\mathcal{L}_{\mathbf{X}}(\boldsymbol{\theta})}{\partial v} = -\frac{n}{2v} + \frac{1}{2v^2}\sum_{i=1}^{n}(X_i-\mu)^2 = 0$$

$$\frac{\partial\log\mathcal{L}_{\mathbf{X}}(\boldsymbol{\theta})}{\partial\mu} = -\frac{1}{2v}\sum_{i=1}^{n}2(X_i-\mu) = 0,$$

Consider a Poisson random variable X with parameter λ. In other words,

$$\mathbb{P}(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}.$$

Verify that

$$\log\left(\mathcal{L}_{\mathbf{X}}(\lambda)\right) = \log(\lambda)\sum_{i=1}^{n} X_i - n\lambda - \sum_{i=1}^{n}\log(X_i!).$$

Show that it follows the maximum likelihood estimator $\widehat{\lambda}$ is equal to the sample mean $\bar{X}$, and explain why this makes sense intuitively.

$$\log \mathcal{L}_X(\lambda) = \sum_{i=1}^{n} \log(P_X(x_i, \lambda)) = \sum_{i=1}^{n}(X_i \log\lambda - \lambda - \log(X_i!))$$

$$= \log(\lambda) \cdot \sum_{i=1}^{n} X_i - n\lambda - \sum_{i=1}^{n} \log(X_i!)$$

$$\frac{\partial \log \mathcal{L}_X(\lambda)}{\partial \lambda} = \frac{1}{\lambda}\sum_{i=1}^{n} X_i - n = 0 \Rightarrow \lambda = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$$

2^{nd} derivative $= -\frac{1}{\lambda^2}\cdot\sum_{i=1}^{n} X_i < 0$ $(\lambda > 0, \text{Int})$ $\Rightarrow$ Local Maximum at $\lambda = \bar{X}$

↑ decide Max or Min

↓ only one c.p.

Global Maximum

Example

Suppose $Y = X\beta + \epsilon$ for $i = 1, 2, \cdots, n$, where ϵ has distribution $\mathcal{N}(0, I\sigma^2)$. Treat σ as known and β as the only unknown parameter. Suppose that n observations $(X_1, Y_1), \ldots, (X_n, Y_n)$ are made.

Show that the least squares estimator for β is the same as the MLE for β by making observations about your log likelihood.

$$\epsilon \sim N(0, I\sigma^2) \Rightarrow f_\epsilon(Y_i - X_i\beta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(Y_i - X_i\beta)^2}{\sigma^2}}$$

$$\log \mathcal{L}_{(X,Y)}(\beta) = \log\left(\frac{n}{\sigma\sqrt{2\pi}}\right) - \sum_{i=1}^{n}\frac{(Y_i - X_i\beta)^2}{\sigma^2}$$

$$\text{Maximum}(\log \mathcal{L}_{(X,Y)}(\beta)) \Rightarrow \text{Maximum}\left(-\frac{(Y_i - X_i\beta)^2}{\sigma^2}\right) \Rightarrow \text{Mini}\left(\frac{(Y_i - X_i\beta)^2}{\sigma^2}\right)$$

↳ Least Square

Exercise

(a) Consider the family of distributions which are uniform on $[0, b]$, where $b \in (0, \infty)$. Explain why the MLE for the distribution maximum b is the sample maximum.

(b) Show that the MLE for a Bernoulli distribution with parameter p is the empirical success rate $\frac{1}{n}\sum_{i=1}^{n} X_i$.

(a) $f(x_i) = \frac{1}{b}$

$\mathcal{L}_X f(x_i) = (\frac{1}{b})^n \Rightarrow$ Maximum $= b \Rightarrow$ minimum $(b \geq \mathrm{Max}(x_i))$

$\therefore b = \mathrm{Max}(X_i) \Rightarrow \mathcal{L}_X f(x_i)$ reach Maximum.

(b) $f(x_i) = p^{x_i} \cdot (1-p)^{1-x_i}$, $x_i \in \{0,1\}$

$\mathcal{L}_X f(x_i) = p^{\sum x_i} \cdot (1-p)^{n-\sum x_i}$, $\log \mathcal{L} = \sum x_i \cdot \log p + (n - \sum x_i) \cdot \log(1-p)$

$\frac{\partial \log \mathcal{L}}{\partial p} = \frac{1}{p}\sum x_i - \frac{n}{1-p} + \frac{1}{1-p} \cdot \sum x_i = 0$

why Maximum?

$= \frac{1}{p(1-p)} \sum x_i = \frac{1}{1-p} \cdot n \Rightarrow p = \frac{1}{n}\sum x_i$

2^{nd} derivative $= -\frac{1}{p^2} \cdot \sum x_i + \frac{n}{(1-p)^2} - \frac{\sum x_i}{(1-p)^2}$

Solution.

(a) The likelihood associated with any value of b smaller than the sample maximum is zero, since at least one of the density values is zero in that case. The likelihood is a decreasing function of b as b ranges from the sample maximum to ∞, since it's equal to $(1/b)^n$. Therefore, the maximal value is at the sample maximum.

(b) The derivative of the log likelihood function is

$$\frac{\mathrm{d}}{\mathrm{d}p} \log \frac{p^s}{(1-p)^{n-s}} = \frac{s}{p} - \frac{n-s}{1-p},$$

where s is the number of successes. Setting the derivative equal to zero and solving for p, we find $p = s/n$.

[MLE]

(a) Find the maximum likelihood estimator for the family of geometric distributions with parameter $0 < p < 1$. (You don't need to prove that the value you find is actually a maximum; just differentiate the log-likelihood and solve for the zero).

(b) I simulated 10 independent samples from a geometric distribution with parameter p and got

$$[0, 4, 1, 3, 4, 3, 1, 14, \ 0, 13]$$

Use the maximum likelihood estimator to estimate the value of p that I used.

$$\Pr(X=k) = (1-p)^{k-1} \cdot p$$

$$\mathcal{L}_{X_1 \ldots X_n}(X = X_i) = p^n \cdot \prod_{i=1}^{n} (1-p)^{X_i - 1}$$

$$\log \mathcal{L}_X() = n \cdot \log p + \sum_{i=1}^{n} (X_i - 1) \cdot \log(1-p) = n \cdot \log p - n \log(1-p) + \log(1-p) \sum_{i=1}^{n} X_i$$

$$\frac{\partial \log \mathcal{L}_X}{\partial p} = \frac{n}{p} + \frac{n}{1-p} - \frac{\sum_{i=1}^{n} X_i}{1-p} = 0$$

$$\Rightarrow \frac{1}{p \cdot (1-p)} = \frac{\frac{1}{n} \cdot \sum_{i=1}^{n} X_i}{1-p} \Rightarrow p = \frac{n}{\sum_{i=1}^{n} X_i} = \frac{1}{\bar{X}}$$

(b) $\bar{X} = 4.3$

$$\hat{p} = \frac{1}{4.3}$$