The maximum likelihood estimator (MLE)

Say X has pdf $f(x;\theta)$ which depends on some parameter θ that we wish to estimate.

We are given n independent samples $X_1, X_2, \ldots, X_n$ of X. Given these samples, we define the <u>likelihood</u> function as

$$L(X_1, X_2, \ldots, X_n; \theta) = f(X_1;\theta) \cdot f(X_2;\theta) \cdots f(X_n;\theta)$$

product of the density evaluated at each sample

The <u>maximum likelihood estimate</u> (MLE) $\hat{\theta}$ is the value of θ which maximizes L:

$$\hat{\theta} = \max_{\theta} L(X_1, X_2, \ldots, X_n; \theta)$$

How do we find the maximum of a function?

If $L(X_1, \ldots, X_n; \theta)$ is differentiable and concave in θ (not always the case) we find $\hat{\theta}$ such that

$$\left.\frac{\partial L(X_1, \ldots, X_n; \theta)}{\partial \theta}\right|_{\theta = \hat{\theta}} = 0$$

Handy trick: since $\ln(x)$ is a strictly increasing function, $\ln L(\cdots)$ has the same max as $L(\cdots)$. Thus we could also find $\hat{\theta}$ s.t.

$$\left.\frac{\partial \ln L(X_1, \ldots, X_n; \theta)}{\partial \theta}\right|_{\theta = \hat{\theta}} = 0$$

Example 1

pdf:

$$X \sim \mathrm{Exp}(1/\beta) \qquad f(x;\beta) = \frac{1}{\beta} e^{-x/\beta} \quad x \geq 0$$

Observe n samples $X_1, \ldots, X_n$

Likelihood function

$$L(X_1, X_2, \ldots, X_n; \beta) = \left(\frac{1}{\beta} e^{-X_1/\beta}\right)\left(\frac{1}{\beta} e^{-X_2/\beta}\right) \cdots\cdots$$

$$= \frac{1}{\beta^n} e^{-\frac{1}{\beta}\sum_{i=1}^{n} X_i}$$

$$\ln L(X_1, X_2, \ldots; \beta) = -n \ln \beta - \frac{1}{\beta}\sum_{i=1}^{n} X_i$$

$$\frac{\partial \ln L}{\partial \beta} = \frac{-n}{\beta} + \frac{1}{\beta^2}\sum_{i=1}^{n} X_i$$

$$= 0 \quad \text{for}$$

$$\frac{n}{\hat{\beta}} = \frac{1}{\hat{\beta}^2}\sum_i X_i \Rightarrow \boxed{\hat{\beta} = \frac{1}{n}\sum_{i=1}^{n} X_i}$$

(sample mean)

Example 2

X is Gaussian with unknown mean η and known variance σ^2

$$f(x;\eta) = \frac{1}{\sqrt{2\pi}\cdot\sigma_x} e^{-(x-\eta)^2/2\sigma^2}$$

Observe $X_1, X_2, \ldots, X_n$

$$L(X_1,\ldots,X_n;\eta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} \cdot e^{-(X_i-\eta)^2/2\sigma^2}$$

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} \cdot e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\eta)^2}$$

$$\ln L(\cdots;\eta) = -\frac{n}{2}\ln 2\pi\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\eta)^2$$

$$\frac{\partial \ln L}{\partial \eta} = \frac{1}{2\sigma^2}\sum_{i=1}^{n}(2\eta - 2X_i)$$

$$= 0 \quad \text{when} \quad \sum_{i=1}^{n}(\hat{\eta} - X_i) = 0$$

$$\Rightarrow \sum_{i=1}^{n}\hat{\eta} = \sum_{i=1}^{n} X_i$$

$$\Rightarrow n\hat{\eta} = \sum_{i=1}^{n} X_i$$

$$\Rightarrow \boxed{\hat{\eta} = \frac{1}{n}\sum_{i=1}^{n} X_i}$$

sample mean again

Problem

Suppose X has pdf

$$f(x;\theta) = \theta x^{\theta-1} \qquad 0 \le x \le 1$$

for some $0 < \theta < \infty$

Observe $X_1, \ldots, X_n$.

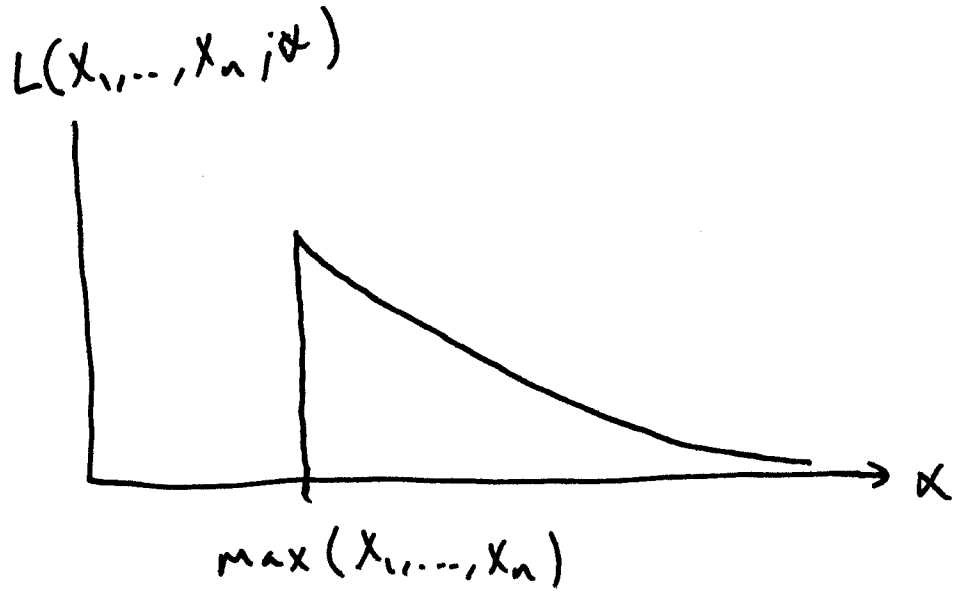
Calculate the MLE for θ.

Calculating the MLE is not always as easy as taking a derivative. For example, say X is Uniform$([0,\alpha])$ where α is unknown

$$f(x;\alpha) = \begin{cases} 1/\alpha & 0 \le x \le \alpha \\ 0 & \text{else} \end{cases}$$

If we observe $X_1, X_2, \ldots, X_n$ the likelihood function is

$$L(X_1,\ldots,X_n;\alpha) = \begin{cases} (1/\alpha)^n & \underbrace{\alpha \ge X_i \text{ for all } i}_{\text{i.e. } \alpha \ge \max_i X_i} \\ 0 & \text{else} \end{cases}$$



This function is clearly not differentiable in α, but it is clear that it is maximized at

$$\hat{\alpha} = \max(X_1, \ldots, X_n)$$

<u>Note</u>: The MLE $\hat{\theta}$ is a random variable itself (since it is a function of r.v.s $X_1, \ldots, X_n$)

In the case above, we can actually find the distribution for the MLE $\hat{\alpha}$. It is simply the max of n uniform random variables

$$\Rightarrow f(\hat{\alpha}) = \frac{n\hat{\alpha}^{n-1}}{\alpha^n} \qquad 0 \le \hat{\alpha} \le \alpha$$

From this we can compute the <u>mean</u> and <u>variance</u> of $\hat{\alpha}$.

to see if its biased

to see if its consistent

$$E[\hat{\alpha}] = \int_0^{\alpha} \hat{\alpha} \cdot \frac{n\,\hat{\alpha}^{n-1}}{\alpha^n}\, d\hat{\alpha}$$

$$= \frac{n}{\alpha^n}\left(\frac{\hat{\alpha}^{n+1}}{n+1} \Big|_0^{\alpha} \right) = \frac{n}{n+1} \cdot \alpha$$

So $\hat{\alpha}$ is <u>biased</u> (a slight under estimate)

The second moment is

$$E[\hat{\alpha}^2] = \int_0^{\alpha} \hat{\alpha}^2 \frac{n\,\hat{\alpha}^{n-1}}{\alpha^n}\, d\hat{\alpha}$$

$$= \frac{n}{\alpha^n}\left(\frac{\hat{\alpha}^{n+2}}{n+2} \Big|_0^{\alpha} \right) = \frac{n}{n+2}\alpha^2$$

$$\Rightarrow Var[\hat{\alpha}] = \frac{n}{n+2}\alpha^2 - \left[\frac{n}{n+1}\alpha\right]^2$$

$$= \alpha^2 \cdot \frac{n}{(n+2)(n+1)^2}$$

Since $Var[\hat{\alpha}] \to 0$ as $n \to \infty$, the MLE is consistent.

Properties of the MLE

① It can be biased or unbiased

② Under fairly general conditions, the MLE is <u>consistent</u>

(i.e. gets arbitrarily close to the "right answer" as n gets big)

③ The MLE is <u>invariant</u>.

If $\hat{\theta}$ is the MLE for θ

then $g(\hat{\theta})$ is the MLE for $g(\theta)$

This property does not hold for general estimators — it is one of the things that make the MLE so nice

④ The MLE $\hat{\theta}$ is asymptotically Gaussian

As $n \to \infty$

$$\hat{\theta} \approx \text{Normal}\left(\theta, \frac{1}{B}\right)$$

where $$B = n\, E\left[\frac{\partial}{\partial \theta} \ln f(X;\theta)\right]$$