Maximum Likelihood Parameter Estimation

April 8, 2013

Consider random variable X with probability distribution function $f(x)$ that depends on a single distribution parameter θ . It takes n independent samples $x_1, ..., x_n$.

Suppose we know $x_1, ...x_n$, but not θ , how to find θ ?

What is θ ? θ is a parameter that is important in distributions, such as the λ in Poisson Distribution $\left(f(k;\lambda) = \frac{\lambda^k e^{-\lambda}}{k!} \right)$, or the p in Binomial distribution $\left(f(k:n,p) = C_k^n p^k (1-p)^{n-k} \right)$

How to find θ ? Let's consider the probability that n-samples selected randomly consists of exactly that n value.

$$P = \begin{cases} f(x_1)f(x_2)...f(x_n) = \prod_{i=1}^{n} f(x_i) & \text{Discrete case} \\ f(x_1)\delta x \cdot f(x_2)\delta x...f(x_n)\delta x = \prod_{i=1}^{n} f(x_i) \cdot \delta x^n & \text{Continuous case} \end{cases}$$

P thus depends on $x_1, x_2, ..., x_n$ and θ. But we don't know θ. We can use that n sample to estimate θ.

For given set of $x_1, x_2, ..., x_n$, P is a function of θ , called Likelihood function.

Since θ will affect P , there are lots of possible θ , then it is good to choose some specific θ.

The principle of maximum likelihood function isto choose a approximation for unknow value θ that P is as large as possible.

For P to be maximum, when P is differentiable w.r.t θ : $\frac{\partial P}{\partial \theta} = 0$

The solution depending on given set of x_1 , x_2 , ... , x_n is called maximum likelihood estimate.

Since $\ln P$ is monotonic increasing for P ($P \in [0,1]$) , so it will be better to consider

$$\frac{\partial \ln P}{\partial \theta} = 0$$

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

Then

$$P = \prod_{i=1}^{n} f(x_i)$$

$$P = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_1-\mu}{\sigma}\right)^2} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_2-\mu}{\sigma}\right)^2} ... \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_n-\mu}{\sigma}\right)^2}$$

$$P = \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{1}{\sigma}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

The propbability P is a function of x_i , μ, σ

Assume we have an Gaussian Variable that only x_i are known, how to find μ and σ ?

Recall from Statistics, μ is the mean and σ is the variance.

$$\mu = \frac{\sum_{i=1}^{n} x_i}{n} \qquad \sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)}$$

Let's try to "derive them"

First, take logarithms

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

Then , to find μ that maximize P , consider the first order condition

$$\begin{aligned} \frac{\partial \ln P}{\partial \mu} &= -\frac{\partial}{\partial \mu}\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2 \\ &= \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) \end{aligned}$$

For maximum P ,

$$\begin{aligned} \frac{\partial \ln P}{\partial \mu} &= \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) = 0 \\ \iff & \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) = 0 \\ \iff & \mu = \frac{1}{n}\sum_{i=1}^{n} x_i \end{aligned}$$

This is the estimation of μ , which is called *mean* !

To find σ that maximize P , do the same, consider the first order condition

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

$$\frac{\partial \ln P}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i - \mu)^2$$

For maximum P

$$\begin{aligned} \frac{\partial \ln P}{\partial \sigma} &= -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i - \mu)^2 = 0 \\ \iff & \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2 \end{aligned}$$

i.e.

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

$-END-$