Maximal Likelihood Estimation

$X_1, X_2, X_3, \ldots X_n$ have joint density denoted

$$f_\theta(x_1, x_2, \ldots, x_n) = f(x_1, x_2, \ldots, x_n|\theta)$$

Given observed values $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$, the likelihood of θ is the function

$$lik(\theta) = f(x_1, x_2, \ldots, x_n|\theta)$$

considered as a function of θ.

If the distribution is discrete, f will be the frequency distribution function.

In words: $\boxed{\text{lik}(\theta)\text{=probability of observing the given data as a function of } \theta.}$

Definition:

The maximum likelihood estimate (mle) of θ is that value of θ that maximises $lik(\theta)$: it is the value that makes the observed data the "most probable".

If the X_i are iid, then the likelihood simplifies to

$$lik(\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

Rather than maximising this product which can be quite tedious, we often use the fact that the logarithm is an increasing function so it will be equivalent to maximise the log likelihood:

$$l(\theta) = \sum_{i=1}^{n} \log(f(x_i|\theta))$$

9.0.1 Poisson Example

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

For $X_1, X_2, \ldots, X_n$ iid Poisson random variables will have a joint frequency function that is a product of the marginal frequency functions, the log likelihood will thus be:

$$\begin{aligned} l(\lambda) &= \textstyle\sum_{i=1}^{n}(X_i log\lambda - \lambda - logX_i!) \\ &= log\lambda \textstyle\sum_{i=1}^{n} X_i - n\lambda - \sum_{i=1}^{n} logX_i! \end{aligned}$$

We need to find the maximum by finding the derivative:

$$l'(\lambda) = \frac{1}{\lambda}\sum_{i=1}^{n} x_i - n = 0$$

which implies that the estimate should be

$$\hat{\lambda} = \bar{X}$$

(as long as we check that the function l is actually concave, which it is).

The mle agrees with the method of moments in this case, so does its sampling distribution.

9.0.2 Normal Example

If $X_1, X_2, \ldots, X_n$ are iid $\mathcal{N}(\mu, \sigma^2)$ random variables their density is written:

$$f(x_1, \ldots, x_n|\mu, \sigma) = \prod_i^n \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}[\frac{x_i - \mu}{\sigma}]^2\right)$$

Regarded as a function of the two parameters, μ and σ this is the likelihood:

$$\ell(\mu, \sigma) = -nlog\sigma - \frac{n}{2}log2\pi - \frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \mu)^2$$

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

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

so setting these to zero gives $\bar{X}$ as the mle for μ, and $\hat{\sigma}^2$ as the usual.

9.0.3 Gamma Example

$$f(x|\alpha, \lambda) = \frac{1}{\Gamma(\alpha)}\lambda^\alpha x^{\alpha-1}e^{-\lambda x}$$

giving the log-likelihood:

$$l(x|\alpha, \lambda) = \sum_{i=1}^n [\alpha log\lambda + (\alpha - 1)logx_i - \lambda x_i - log\Gamma(\alpha)]$$

One ends up with a nonlinear equation in $\hat{\alpha}$ this cannot be solved in closed form, there are basically two methods and they are called root-finding methods, they are based on the calculus theorem that says that when a function is continuous, and changes signs on an interval, it is zero on that interval.

For this particular problem there already coded in matlab a mle method called gamfit, that also provides a confidence interval.

For general optimization, the function in Matlab is `fmin` for one variable, and `fmins` you could also look at how to use `optimize` in Splus.