Financial Econometrics
Prof Olivier Scaillet
TA Gaetan Bakalli

Maximum Likelihood Estimators

Let $x_1, ..., x_T$ be the sample we are given with, i.e., our observations. If we assume that: (1) the observations are *independently* sampled and (2) the process that generated this sample is a normal distribution with mean μ and variance σ^2, then the joint density (i.e., the likelihood, the probability) of obtaining the sample $x_1, ..., x_T$ is given by the likelihood function:

$$L(\mu, \sigma^2) = f(x_1) \times ... \times f(x_T)$$

where $f(x_i) = \left(2\pi\sigma^2\right)^{-\frac{1}{2}} exp\left(-\frac{1}{2\sigma^2}\left(x_i - \mu\right)^2\right)$. Plugging in, we get

$$\begin{aligned} L(\mu, \sigma^2) &= \prod_{i=1}^{T} \left(2\pi\sigma^2\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2}\left(x_i - \mu\right)^2\right) \\ &= \left(2\pi\sigma^2\right)^{-\frac{T}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{T}\left(x_i - \mu\right)^2\right) \end{aligned}$$

If you look closely, you may notice that as T grows large, the value of $L(\mu, \sigma^2)$ will tend to zero, and this will make the function too flat, and therefore any numerical optimizer such as fminsearch will have trouble finding even a local maximum or minimum. Fortunately, from calculus we know that if we take a monotonic transform $l(\mu, \sigma^2)$ of $L(\mu, \sigma^2)$, the values of μ and σ^2that maximize $l(\mu, \sigma^2)$ will be the same as those that maximize $L(\mu, \sigma^2)$. To make things simple, therefore, we choose as our monotone transformation the (natural) log function. So, we get:

$$\begin{aligned} l(\mu, \sigma^2) &= \log\left[L(\mu, \sigma^2)\right] \\ &= \log\left[\left(2\pi\sigma^2\right)^{-\frac{T}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{T}\left(x_i - \mu\right)^2\right)\right] \\ &= -\frac{T}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{T}\left(x_i - \mu\right)^2 \end{aligned}$$

The maximum likelihood principle tells us that we should choose the values of μ and σ^2that maximize $l(\mu, \sigma^2)$ and $L(\mu, \sigma^2)$ in order to obtain the parameters of the normal distribution that *most likely* generated the observed sample $x_1, ..., x_T$. To maximize $l(\mu, \sigma^2)$ we may do two things:

1. Brute force method: In Matlab, use fminsearch on the function $-l(\mu, \sigma^2)$ to identify the parameters that minimize the negative of $l(\mu, \sigma^2)$, that is, the parameters that maximize $l(\mu, \sigma^2)$ itself. This was done in class, or you may check the solutions.

2. Analytical method: In this case, we may find the values of μ and σ^2 by setting the partial derivatives of $l(\mu, \sigma^2)$with respect to the variables μ and σ^2 equal to zero and solving for the corresponding variables.

The analytical method gives:

$$\begin{aligned}\frac{\partial l(\mu, \sigma^2)}{\partial \mu} &= \frac{1}{\sigma^2} \sum_{i=1}^{T} (x_i - \mu) \\ \frac{\partial l(\mu, \sigma^2)}{\partial \sigma^2} &= -\frac{T}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{T} (x_i - \mu)^2\end{aligned}$$

Setting the partial derivatives equal to zero and solving gives $\hat{\mu} = \frac{1}{T} \sum_{i=1}^{T} x_i$ and $\widehat{\sigma^2} = \frac{1}{T} \sum_{i=1}^{T} (x_i - \hat{\mu})^2$. Clearly, for more complicated likelihood functions it will be more difficult or even impossible to obtain explicit solutions for the parameters as in this case. Finally, note that the estimator for σ^2 is biased since the denominator is T instead of $T - 1$. So, Maximum Likelihood Estimators are not necessarily unbiased in general.