ECONOMETRICS I, Spring 2024.

Bent E. Sørensen

Notes on Maximum Likelihood.

Consider a sample of *dependent* variables $x_1, ..., x_T$ with density $f(x_1, ..., x_T; \theta)$, where θ is a $k-$vector of parameters. We will mainly consider the multivariate Normal with mean μ and variance Σ. To estimate μ and variance Σ you find

$$\hat{\mu}, \hat{\Sigma} = \text{argmax}_{\mu,\Sigma} \frac{1}{((2\pi)^N |\Sigma|)^{0.5}} \exp\{-0.5(X-\mu)'\Sigma^{-1}(X-\mu)\}.$$

or, equivalently,

$$\text{argmax}_{\mu,\Sigma} - \log(|\Sigma|) - (X-\mu)'\Sigma^{-1}(X-\mu)\,.$$

This model quickly becomes impossible to estimate as the sample size grows because of the large number of elements in Σ so we need to postulate a lower-dimensional model for Σ. We often assume that the data follows an ARMA model, which makes Σ a function of a low number of parameters.

In most cases, we make use of conditioning. Recall that $f(x|y)$ is $f(x,y)/f(y)$ or $f(x,y)$ is $f(x|y)\,f(y)$ so

$$f(x_1, ..., x_T; \theta) = f(x_T|x_1, ..., x_{T-1}; \theta) f(x_1, ..., x_{T-1}; \theta)$$

But you can do the same for $f(x_1, ..., x_{T-1}; \theta)$ and on and on, till you have

$$f(x_1, ..., x_T; \theta) = f(x_T|x_1, ..., x_{T-1}; \theta) f(x_{T-1}|x_1, ..., x_{T-2}; \theta) .... f(x_2|x_1; \theta) f(x_1; \theta)\,.$$

We usually write $f(x_t|x_1, ..., x_{t-1}$ as $f(x_t|X^{t-1})$ or sometimes more compactly as $f_{t-1}(x_t; \theta)$. You can think of AR models as ways to write probabilities conditional on the past in a simple way (we will elaborate in the next handout).

We can take logs as before and get the log likelihood function

$$\log f(x_1; \theta) + \log f_1(x_2; \theta) + ... + \log f_{T-1}(x_T; \theta)\,.$$

For stationary models, the LLN and CLT we used in the ML theory still holds (not for a random walk). So the powerful results regarding the asymptotic variance and the Hessian etc. still holds.

For models with an MA-component, the conditional likelihood is actually complicated and we will not look at this for now. AR(models, on the other hand, are formulated directly in terms of the conditional distribution. To bring home the intuition, we will just look at the AR(1) model here (and that is what you should know for the exam). We will return to more complicated models in Econometrics II. For the AR(1) model :

$$y_t = \mu + a y_{t-1} + u_t \, ,$$

we have for a given t that conditioning on the past, the mean is $\mu + a y_{t-1}$ and the random part is u_t which are i.i.d. and independent of past u'S and (therefore) past y's. So the log-likelihood function becomes the sum of terms involving u_t (as a function of unobserved a and μ) and the unconditional likelihood for the first observation: $l(\mu, a, \sigma^2) = \log f(y_1; \mu, a, \sigma^2) + \log f(y_2|y_1; \mu, a, \sigma^2) + ... + \log f(y_T|y_{T-1}; \mu, a, \sigma^2)$ which is (after multiplying by 2, to get rid of all the 0.5 factors):

$$-\frac{1}{2}[\log(\sigma^2) - \log(1-a^2)] - \frac{(y_1 - \frac{\mu}{1-a})^2}{2\sigma^2/(1-a^2)} - \Sigma_{t=2}^{T}[\frac{1}{2}\log(\sigma^2) + \frac{(y_t - \mu - a y_{t-1})^2}{2\sigma^2}]\ ]$$

The Maximum Likelihood estimator of a may differ substantially from a least squares estimator, where you ignore the first term, if a is near unity where the term with the log of $1 - a^2$ can affect the estimate a lot (in particular, the ML estimator will not allow $\hat{a}$ to be unity as this would blow up the log term.)