1 What is Bayesian estimation?

While I can not spend much time on it, some mention should be made of Bayesian analysis in any introductory course in Econometrics, as some knowledge of the controversy between Bayesians and classical statisticians is essential to follow some of the discussions in the economic literature.

The strength of Bayesian analysis is that it allows the econometrician to take into account other information than what is in the data, and bring this information into the estimation in a coherent way. The way you bring extra information into the problem is through your *prior* assumption about the parameters.

The difference between Bayesian analysis and the analysis we have so far discussed (what the Bayesians call the *classical* statistical perspective), is partly philosophical.

2 Classical estimation.

In classical statistics we assume that there are some *true* parameters θ that we want to do inference about.

For example, if we want to estimate the mean of set of normal *iid* $\mathcal{N}(\mu, \sigma^2)$ observations, both μ and σ are viewed as fixed, but unknown numbers that we want to estimate. The estimator $\widehat{\theta}$ is a random variable, and we want to find the best estimator based on

- Consistency/Unbiasedness. $\widehat{\theta} \to \theta$
- Mean squared error $E[(\widehat{\theta} - \theta)(\widehat{\theta} - \theta)]$

For example, the Maximum Likelihood estimator of the mean is found by maximizing the sample likelihood,
Exercise 1.

Suppose a variable x_i has a normal distribution with mean μ and variance σ^2.

1. Determine the maximum likelihood estimator of μ.
2. Determine the maximum likelihood estimator of σ^2.

Solution to Exercise 1.

First recall the probability distribution for a normally distributed varible x_i

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2}}$$

The likelihood function is

$$L(x; \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2}}$$

We will instead of the likelihood function maximize the log-likelihood function:

$$\ell(y; \theta) = \sum_{i=1}^{n} \ln\left(\frac{1}{\sqrt{2\pi}\sigma}\right) - \sum_{i=1}^{n} -\frac{1}{2}\frac{1}{\sigma^2}(x_i - \mu)^2$$

Rewrite the log-likelihood function as

$$\ell(y; \theta) = \sum_{i=1}^{n} -\frac{1}{2}\ln(2\pi) - \ln(\sigma) - \sum_{i=1}^{n} -\frac{1}{2}\frac{1}{\sigma^2}(x_i - \mu)^2$$

We find the estimator from the first order conditions, first estimating μ

$$\frac{\partial \ell}{\partial \mu} = \sum_i 2\frac{1}{2\sigma^2}(y_i - \mu) = 0$$

$$\to \sum_i (y_i - \mu) = 0$$

$$\rightarrow \sum_i y_i = n\mu$$

$$\widehat{\mu}^{ml} = \frac{\sum_i y_i}{n}$$

and then estimating σ^2.

$$\frac{\partial \ell}{\partial \sigma} = \sum_{i=1}^{n} -\frac{1}{\sigma} - \sum_{i=1}^{n} \frac{1}{2}\left(\frac{0-(x_i-\mu)^2 2\sigma}{\sigma^4}\right)$$

$$0 = \sum_{i=1}^{n} -\frac{1}{\sigma} + \sum_{i=1}^{n} \frac{1}{2}\left(\frac{2(x_i-\mu)^2\sigma}{\sigma^4}\right)$$

$$0 = \sum_{i=1}^{n} -\frac{1}{\sigma} + \sum_{i=1}^{n} \left(\frac{(x_i-\mu)^2}{\sigma^3}\right)$$

$$0 = -\sum_{i=1}^{n} 1 + \sum_{i=1}^{n} \left(\frac{(x_i-\mu)^2}{\sigma^2}\right)$$

$$0 = -n + \sum_{i=1}^{n} \left(\frac{(x_i-\mu)^2}{\sigma^2}\right)$$

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

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

3 Bayesian Estimation.

In contrast to the classical analysis, Bayesian analysis treats everything in terms of probability distributions. Even θ itself is viewed as a random variable. The goal of Bayesian analysis is to describe the analysts uncertainty about θ in terms of a probability distribution. To do inference, a Bayesian analyst will make probability statements.

Before observing the data, the analyst will summarize his knowledge in the *prior* distribution

$$f(\theta)$$

The *joint* probability density $f(y, \theta)$ describes the probability for observing both the data y and the parameter θ

$$f(y, \theta) = f(y|\theta) \cdot f(\theta)$$

The goal of Bayesian analysis is to find a *posterior density*

$$f(\theta|y),$$

which is the probability distribution of θ we postulate after having observed the data.

Use the definition of conditional probability

$$f(\theta|y) = \frac{f(y,\theta)}{f(y)}$$

Since

$$f(y) = \int_{-\infty}^{\infty} f(y,\theta)d\theta$$

We can write it as

$$f(\theta|y) = \frac{f(y,\theta)}{\int_{-\infty}^{\infty} f(y,\theta)d\theta}$$

Also, since

$$f(y|\theta) = \frac{f(y,\theta)}{f(\theta)}$$

then

$$f(y,\theta) = f(y|\theta)f(\theta)$$

This can be used to formulate the well known *Bayes formula*:

$$f(\theta|y) = \frac{f(y|\theta)f(\theta)}{\int_{-\infty}^{\infty} f(y,\theta)d\theta}$$

Example

Bayesian estimation of mean of normal. Let us look at how a Bayesian will proceed in the estimation of the mean of a normal distribution.

$$y \sim \mathcal{N}(\mu, \sigma^2)$$

Recall the normal distribution function

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(y-\mu)^2}{\sigma^2}}$$

A Bayesian will summarize his prior assumptions about the distribution into the parameters m and ν in the following *prior distribution*.

$$f(\mu|\sigma) = \frac{1}{\sqrt{2\pi\sigma^2/\nu}} e^{-\frac{1}{2}\frac{(\mu-m)^2}{\sigma^2/\nu}}$$

The parameter m is the best guess on the mean based on prior knowledge, and the parameter ν describing the *precision* of the prior knowledge.

The *sample* distribution

$$\begin{aligned} f(y|\mu,\sigma) &= \prod_{t=1}^{n} \left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{1}{\sigma^2}(y_t-\mu)^2} \right) \\ &= \left(\frac{1}{2\pi\sigma^2} \right)^{\frac{T}{2}} e^{-\frac{1}{2}\frac{1}{\sigma^2}\sum_{t=1}^{T}(y_t-\mu)^2} \\ &= \left(2\pi\sigma^2\right)^{-\frac{T}{2}} e^{-\frac{1}{2}\frac{1}{\sigma^2}(\mathbf{y}-\mu\mathbf{1})(\mathbf{y}-\mu\mathbf{1})'} \end{aligned}$$

The goal is to produce the *posterior* distribution

$$f(\mu|y;\sigma)$$

Without going through the calculations,[1] it turns out that this will be

$$f(\mu|y;\sigma) = \frac{1}{\sqrt{2\pi\frac{\sigma^2}{\nu+T}}} e^{-\frac{1}{2}\frac{(\mu-m^*)^2}{\frac{\sigma^2}{\nu+T}}}$$

where

$$m^* = \left(\frac{\nu}{\nu+T}\right) m + \left(\frac{T}{\nu+T}\right) \frac{\sum_t y_t}{T}$$

[1] For this example, look in Hamilton (1994) for the details.

To find the estimator, we minimize the *loss function*

$$E[(\mu - \widehat{\mu})^2]$$

Doing the same steps as when doing Maximum Likelihood before, we find that $\widehat{\mu} = m^*$ is the Bayesian estimator

If you look at

$$\widehat{\mu} = \left(\frac{\nu}{\nu + T}\right) m + \left(\frac{T}{\nu + T}\right) \frac{\sum_t y_t}{T}$$

you note that this is a weighted average of your prior m and the sample mean $\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$. Also note that the lower the precision ν, the more weight is put on the data, and the less on the prior.

Note that the limit when $\nu \to 0$ is the classical (sample average) estimate. This is what is called an *improper* or *diffuse* prior.[2]

3.1 Comments

The major strength of Bayesian analysis, the ability to account for prior information, is also one of its major weaknesses, because of the subjectivity that goes into the specification of the prior. To some degree, to agree with the conclusion of a Bayesian estimation, one has to agree that the priors are justified. The issue is the one of the *source* of the prior. This is similar to the problem of differences of opinion in Aumann's game theoretic concept of "correlated equilibrium," that at some level we all need to have similar priors.[3] The issue of subjectivity is an important deterrent to the acceptance of Bayesian results, one often find it easier to accept results based on classical statistical methods because of the aura of objectivity that follows from "only" considering the data.[4]

The messiness of the calculations involved in the transform from the prior to the posterior is actually also a deterrent to doing Bayesian analysis. If you look at the two pages of small type in Hamilton that does the transform from a prior to a posterior for a simple (OLS) regression, you see why some people may prefer the (relatively) simpler calculations involved in finding Maximum Likelihood estimators.

On a lighter note, (but still relevant), another problem for the acceptance of Bayesian Analysis is that the population of Bayesian statisticians have more than its fair share of cranks, who argue the Bayesian viewpoint with religious fervor. According to this brand of econometrician, Bayesian statistics is the *only* thing to do, and their method of teaching econometrics have more in common with Jehova's witnesses on the scent of a new convert(convict) than with coherent argument. This kind of experience tends to turn off many people curious about using Bayesian method in cases where it may be sensible.

4 Readings

Zellner

(Hamilton, 1994, ch 12: 12.1)

5 Further Readings

DeGroot (1971)

[2]Note that taking $\nu = 0$ is not well-defined, you need to look at

$$\lim_{\nu \to 0} f(\mu; \sigma^2)$$

[3]If one goes back to Adam and Eve there is some sense in which that is true, but that argument is somewhat remote from practical estimation problems. . .

[4]Of course, any classical analysis is in some sense using prior knowledge through the choice of the particular model specification, but that choice is usually implied, it is (almost) never explicitly accounted for in the analysis.

References

Morris DeGroot. *Optimal Statistical Decisions*. McGraw–Hill, 1971.

James D Hamilton. *Time Series Analysis*. Princeton University Press, 1994.