Bayesian Econometrics: Meet-up 1

Kevin Arthur Tittel

February 9, 2023

1 Discussion Theory

In this Section we start off with a brief introduction of Bayesian Econometrics given the slides of Lecture 1. The corresponding theory will be divided by a selection of key concepts.

What is the Bayesian approach. Opposed to the traditional, frequentist approach, the novel Bayesian approach is faced with many advantages: Bayesian analysis

- allows the inclusion of prior knowledge on economic theory and expert's opinion;
- is exact in small samples without having to rely on asymptotic results, especially useful for modelling rare events in small samples;
- can be feasibly and easily implemented for advanced econometrics models following (simple) simulation methods;
- is highly operational due to increasing computing power nowadays;
- manages to extract (limited) information whilst assuring parsimonious models;
- is, from a computational perspective, more suitable for hierarchical statistical models than maximum likelihood-based frequentist approaches'
- provides a natural way to deal with uncertainty, both by including parameter uncertainty and modelling uncertainty;

- is shown to be optimal for decision making under uncertainty subject to certain conditions.

Given all these advantages of Bayesian analysis over the frequentist approach, we move on to characterizing the Bayesian approach. In short, the Bayesian approach establishes a probability distribution of the parameters given the data (i.e. $p(\theta|y)$), instead of (the traditional approach of) a probability distribution for the data given the parameters (i.e. $p(y|\theta)$). However, the latter is still of aid; namely, we start off excluding the data y, solely starting from the point that only prior beliefs about θ are known. Then, we start to include some actual data, or likelihood on $y|\theta$, to update the final knowledge about θ. It is essential to realize that the extent to which the prior $p(\theta)$ is informative, and the amount of data points y, play a role in the final shape of the posterior distribution $p(\theta|y)$.

What is a kernel. A kernel function $k(\theta|y)$ belongs to the posterior distribution $p(\theta|y)$, as it uniquely identifies the distribution of the parameters θ. Namely, the posterior distribution is proportional to the kernel (i.e. $p(\theta|y) = ck(\theta|y)$) for a certain constant c independent of θ. Note that the posterior distribution is a proper density function, whereas the kernel function is no density function, and hence does not integrate to 1. Similarly, prior beliefs are no proper densities either. A kernel function is of interest since detecting known distributions makes it easier to obtain analytical solutions.

Example: linear regression models. There are multiple take-aways from this example, which is divided into a case without an informative prior, and with an informative prior. For both examples, we introduce the following distributions: multivariate normal (for $p(y|\theta)$), the multivariate t-distribution (for the marginal posterior $p(\beta|y)$), and the inverted Gamma-2 distribution (for the marginal posterior $p(\sigma^2|y)$).

What is a conjugate prior. A conjugate prior is a prior distribution $p(\theta)$ that ensures that the posterior distribution $p(\theta|y)$ is of the same type, or same family, as $p(\theta)$. This makes it easier to obtain analytical solutions for the posterior distribution (given known distributions), and provides an easier view on how the data updates the prior information into the posterior information.

2 True or False Questions

1. Both the frequentist approach and Bayesian approach require a model with true parameter vector θ and one realization of data y.
2. More informative prior distributions will have greater influence on the posterior distribution for a given sample size.
3. Larger sample sizes are required for more accurate results in case we use less informative, or diffuse, priors.
4. Suppose that the kernel function $k(\theta|y)$ is known. In order to derive the posterior distribution for a particular value of θ, the value of constant c is necessary in the derivation.
5. In linear regression models with parameter set $\theta = (\beta, \sigma^2)$, conjugate priors are a normal distribution for $\beta|\sigma^2$ and an inverted Gamma-2 distribution for σ^2.
6. In linear regression models with parameter set $\theta = (\beta, \sigma^2)$, a conjugate prior with large variance B has a large impact on the mean of the the marginal posterior distribution $p(\beta|y)$.
7. If the number of observations increases, the influence of the prior beliefs always becomes less.