Lectures 17-18-19

1 A primer on Bayesian Analysis

1.1 Decision Problems, Decision Rules, and Admissibility.

Bayesian analysis plays an important role in modern econometrics. The starting point of classical Bayesian analysis is a parametric statistical model:

$$f(x|\theta), \quad \theta \in \Theta.$$

In Bayesian analysis there is typically a concrete decision problem to solve (with actions a in some set $\mathcal{A}$) and a loss-function $\mathcal{L}(a;\theta)$ used to rank actions.

We would like to use the loss function to find a "reasonable" decision rule d for the decision problem at hand.[1]

In order to define what a reasonable decision rule, we need a few more definitions. Define the risk of a decision as the expected loss of a decision at parameter θ:

$$R(d;\theta) = \mathbb{E}_{\theta}[\mathcal{L}(d(x);\theta)] = \int \mathcal{L}(d(x);\theta) f(x|\theta) dx.$$

Our notion of reasonable decision rules is as follows:

Definition 1: A decision function d is admissible if there is no other decision d' such that:

$$R(d';\theta) \leq R(d;\theta)$$

for every θ (with at least some strict inequality).

[1] A decision rule is a map from data to actions

1.2 Bayes Rules are Admissible

Note that the risk function depends on θ. Therefore, some decision rules may have low risk at some points of the parameter space but hight risk in others. A Bayes Decision Rule—for a given "weight "function π on Θ—is one that minimizes average risk defined as:

$$r_\pi(d) \equiv \int_\Theta R(d(x);\theta)\pi(\theta)d_\theta. \tag{1}$$

Consequently, we say that:

Definition 2: d^* is a Bayesian decision rule for the weight π is:

$$r_\pi(d^*) \leq r_\pi(d')$$

for any other rule d'. The weight function π is usually called a prior.

Result: Suppose that the risk function $R(:,\theta)$ is continuous for any decision rule and let π be any prior with full support on Θ. The Bayes rule d^* corresponding to π is admissible.

PROOF: Suppose d^* is not admissible. Then there exists d' such that:

$$R(d',\theta) \leq R(d^*,\theta),$$

with strict inequality for some $\theta^* \in \Theta$. Since $R(d^*,\theta)$ is continuous in θ, there exists a neighborhood $N(\theta^*)$ such that

$$R(d',\theta) < R(d^*,\theta) \quad \forall \quad \theta \in N(\theta^*).$$

Since π has full support, this means that:

$$r_\pi(d') < r_\pi(d^*).$$

A contradiction.

Thus, Bayes Rules are reasonable choices for decision problems.

1.3 Minimizing Bayes Risk and Minimizing Posterior Loss

Minimizing Bayes risk is a complicated problem: we are optimizing over a space of functions. Note however that:

$$
\begin{aligned}
r_\pi(d) &\equiv \int_\Theta R(d(x);\theta)\pi(\theta)d\theta \\
&= \int_\Theta \Big(\int_X \mathcal{L}(d(x);\theta) f(x|\theta)dx \Big) \pi(\theta)d\theta \\
&\quad \text{(by definition of Risk)} \\
&= \int_X \Big(\int_\Theta \mathcal{L}(d(x);\theta) f(x|\theta)\pi(\theta)d\theta \Big) dx \\
&= \int_X \Big(\int_\Theta \mathcal{L}(d(x);\theta)\pi(\theta|x)d\theta \Big) f^*(x)dx
\end{aligned}
$$

where $f^*(x) = \int_\theta f(x|\theta)\pi(\theta)d\theta$. So, minimizing (ex-ante) Bayes Risk is the same as choosing the action $d(x)$ that minimizes:

$$\int_\Theta \mathcal{L}(d(x);\theta)\pi(\theta|x)d\theta.$$

The latter quantity is referred to as <u>Posterior Loss</u>.

2 Bayesian Analysis of Linear Regression

In order to illustrate the concepts discussed in the previous section, we now present a Bayesian analysis of Gaussian Linear Regression:

$$\underbrace{y}_{T\times 1} = \underbrace{X}_{T\times k} \underbrace{\beta}_{k\times 1} + \underbrace{\epsilon}_{T\times 1}. \tag{2}$$

For simplicity, we assume that the analysis is "conditional" on X and that $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$. The parameters of this model are β and σ^2. Hence, a prior will need to specify a joint distribution over (β, σ^2).

2.1 Bayesian Estimation of β

We consider first the problem of estimating the vector β. An action is an element a of $\mathbb{R}^p$. The parameter space Θ is also $\mathbb{R}^p$. The loss function is *quadratic loss*

$$\mathcal{L}(a, \beta) = (a - \beta)'(a - \beta). \tag{3}$$

An estimator—denoted $\widehat{\beta}_T$—is a function that takes data (y, X) to an action a. A Bayes Estimator given a prior $\pi(\cdot)$ on Θ is the estimator (or decision rule) that minimizes average risk given the prior.

Last class we showed that minimizing average risk is the same as *minimizing expected posterior loss for every data realization.* Expected posterior loss of action a is given by:

$$\mathbb{E}[(a - \beta)'(a - \beta) \mid y, X]$$

By adding and subtracting $E[\beta|y, x]$ in the terms inside the brackets, it is not difficult to see that the action that minimizes posterior expected loss is the posterior mean:

$$\mathbb{E}[\beta \mid y, X]$$

Consider then the following prior (or weight function) on β:

$$\beta|\sigma^2 \sim \pi(\beta|\sigma^2) \equiv \mathcal{N}_k(\mu\ ,\ \sigma^2 V), \quad \sigma^2 \sim \pi(\sigma^2) \tag{4}$$

The prior assumes that—conditional on the variance parameter—all the coefficients are approximately normal with values close to the vector μ. There is typically no good way of selecting a prior. More often than not, the selection of a prior trades-off interpretation and convenience in its implementation.

As first step in the analysis, we derive the posterior distribution of β. We obtain this distribution in two-steps. First, we consider the posterior distribution of β but conditional on σ^2. Then, we integrate over σ^2.

The posterior distribution of $\beta|\sigma^2$ is usually obtained by applying Bayes Theorem:

$$\pi(\beta \mid \sigma^2, y, X) = \frac{f(y, X \mid \beta, \sigma^2)\pi(\beta|\sigma^2)}{\int_\Theta f(y, X \mid \beta, \sigma^2)\pi(\beta|\sigma^2) d\beta}$$

Since we have assumed that the analysis is conditional on X we can write:

$$\pi(\beta \mid \sigma^2, y, X) = \frac{f(y \mid X, \beta, \sigma^2)\pi(\beta|\sigma^2)}{\int_\Theta f(y \mid X, \beta, \sigma^2)\pi(\beta, \sigma^2) d\beta}, \tag{5}$$

where $f(y|X, \beta, \sigma^2)$ is the conditional distribution of y given (X, β) and $\pi(\cdot)$ is the p.d.f. of the prior for β. Note that:

$$y|X, \beta, \sigma^2 \sim \mathcal{N}_T(X\beta\ ,\ \sigma^2 \mathbb{I}_T), \quad \beta|\sigma^2 \sim \mathcal{N}_k(0_{k\times 1}, \sigma^2 V).$$

Consequently, $f(y|X, \beta, \sigma^2)\, \pi(\beta|\sigma^2)$ equals:

$$\frac{1}{(\sqrt{2\pi\sigma^2})^T} \exp\left(-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)\right) \frac{1}{\sqrt{2\pi\sigma^2}^k \det(V)^{1/2}} \exp\left(-\frac{1}{2\sigma^2}(\beta - \mu)' V^{-1}(\beta - \mu)\right).$$

Typically, one can compute (or take computer draws from) the posterior distribution of β without worrying about the numerator in (4). In our example, it suffices to manipulate the expression:

$$\exp\left(-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)\right) \exp\left(-\frac{1}{2\sigma^2}(\beta - \mu)' V^{-1}(\beta - \mu)\right),$$

where we can forget about the constants. The expression above equals:

$$\exp\left(-\frac{1}{2\sigma^2} y' y\right) \exp\left(-\frac{1}{2\sigma^2}\beta \left(V^{-1} + X'X\right) \beta + \frac{(y'X + V^{-1}\mu)\beta}{\sigma^2}\right).$$

Completing the square and ignoring all the terms that do not have β on them, gives the posterior distribution as a constant times the exponential of:

$$-\frac{1}{2\sigma^2}\left(\beta-\left(V^{-1}+X'X\right)^{-1}\left(X'y+V^{-1}\mu\right)\right)\left(V^{-1}+X'X\right)\left(\beta-\left(V^{-1}+X'X\right)^{-1}\left(X'y+V^{-1}\mu\right)\right).$$

This implies that:

$$\beta|\sigma^2,y,X\sim\mathcal{N}_k\left(\left(\frac{1}{T}V^{-1}+\frac{X'X}{T}\right)^{-1}\frac{\left(X'y+V^{-1}\mu\right)}{T},\ \frac{\sigma^2}{T}\left(\frac{1}{T}V^{-1}+\frac{X'X}{T}\right)^{-1}\right). \tag{6}$$

This means that the Bayesian Estimator of β given the Gaussian prior $\pi(\beta)$ is:

$$\left(\frac{1}{T}V^{-1}+\frac{X'X}{T}\right)^{-1}\left(\frac{X'y}{T}+\frac{V^{-1}\mu}{T}\right)$$