Bayesian Econometrics

Andrés Ramírez Hassan

Universidad Eafit
Departamento de Economía

January 19, 2017

Outline

Bayesian Inference
Loss function

Given a set of actions, $\mathcal{A}$, whose consequences depend on some state of the nature, Θ, the loss function $L(\theta, a) : \Theta \times \mathcal{A} \rightarrow \mathcal{R}^{+}$, expresses the relative importance of the error committed by selecting $a \in \mathcal{A}$ when $\theta \in \Theta$ is the true.

Bayesian Inference

Bayesian expected loss

Bayesian expected loss

If $\pi(\theta)$ is the believed probability distribution of θ. The Bayesian expected loss of an action a is
$\rho(\pi, a) = E^{\pi} L(\theta, a) = \int_{\Theta} L(\theta, a) dF^{\pi}(\theta)$

Bayesian Inference

Frequentist risk

Decision rule

A decision rule $\delta(x)$ is a function from $\mathcal{X}$, the set of possible outcomes in the sample space, into $\mathcal{A}$.

Frequentist risk

The risk function of a decision rule $\delta(x)$ is defined by
$R(\theta, \delta) = E_\theta^X[L(\theta, \delta(X))] = \int_{\mathcal{X}} L(\theta, \delta(X)) dF^X(x|\theta)$

Bayesian Inference
Inadmissibility

R-better

A decision rule δ_1 is R-better than a decision rule δ_2 if $R(\theta, \delta_1) \leq R(\theta, \delta_2)$ for all $\theta \in \Theta$, with strict inequality for some θ.

Inadmissibility

A decision rule δ is admissible if there exist no R-better decision rule. A decision rule δ is inadmissible if there does exist an R-better decision rule.

Bayesian Inference
The conditional Bayes principle

Bayes action

Choose an action $a \in \mathcal{A}$ which minimizes $\rho(\pi, a)$. Such action will be called a Bayes action and will be denoted a^{π^*}.

Bayesian Inference
The Bayes risk principle

Bayes risk

The Bayes risk of a decision rule δ, with respect to a prior distribution π on Θ, is defined as $r(\pi, \delta) = E^{\pi}[R(\theta, \delta)]$

Bayes rule

A decision rule δ_1 is preferred to a rule δ_2 if $r(\pi, \delta_1) < r(\pi, \delta_2)$. A decision rule which minimizes $r(\pi, \delta)$ is optimal; it is called a Bayes rule (δ^{π}).

Bayesian Inference
The posterior expected loss

The posterior expected loss

The posterior expected loss of an action a, when the posterior distribution is $\pi(\theta|x)$, is $\rho(\pi(\theta|x), a) = \int_{\Theta} L(\theta, a) dF^{\pi(\theta|x)}(\theta)$. A posterior Bayes action ($\delta^{\pi}(x)$) is any action $a \in \mathcal{A}$ which minimizes $\rho(\pi(\theta|x), a)$, or equivalently which minimizes $\int_{\Theta} L(\theta, a) f(x|\theta) dF^{\pi(\theta)}(\theta)$, where $f(x|\theta)$ is the density function.

Bayesian Inference

Bayes rules and posterior expected loss

Result 1

A Bayes rule δ^π can be found by choosing, for each x such that $m(x) > 0$ (the marginal), an action which minimizes the posterior expected loss.

Result 2

If δ is a nonrandomized estimator, then $r(\pi, \delta) = \int_{x:m(x)>0} \rho(\pi(\theta|x), \delta(x))dF^m(x)$.

Bayesian Inference
Estimation problems

Result 3

If $L(\theta, a) = (\theta - a)^2$, the Bayes rule is $\delta^\pi(x) = E^{\pi(\theta|x)}[\theta]$

Result 4

If $L(\theta, a) = w(\theta)(\theta - a)^2$, the Bayes rule is
$\delta^\pi(x) = \frac{E^{\pi(\theta|x)}[w(\theta)\theta]}{E^{\pi(\theta|x)}[w(\theta)]}$

Bayesian Inference
Estimation problems

Result 5

If $L(\theta, a) = |\theta - a|$, any median is a Bayesian estimate of θ.

Result 6

If $L(\theta, a) = \begin{Bmatrix} K_0(\theta - a), \theta - a \geq 0 \\ K_1(a - \theta), \theta - a < 0 \end{Bmatrix}$ any $K_0/(K_0 + K_1)$-fractile of $\pi(\theta|x)$ is a Bayes estimate of θ.

Bayesian Inference
Hypothesis test

Result 7

In testing $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_1$, the actions of interest are a_0 and a_1, where a_i denotes no rejection of H_i. If $L(\theta, a_i) = \begin{Bmatrix} 0, \theta \in \Theta_i \\ K_i, \theta \in \Theta_j (j \neq i) \end{Bmatrix}$ The posterior expected losses of a_0 and a_1 are $K_0 P(\Theta_1|x)$ and $K_1 P(\Theta_0|x)$, respectively. The Bayes decision is that corresponding to the smallest posterior expected loss.

Bayesian Inference

Hypothesis test

Result 7

In the Bayesian test, the null hypothesis is rejected, that is, action a_1 is taken, when $\frac{K_0}{K_1} > \frac{P(\Theta_0|x)}{P(\Theta_1|x)}$, where usually $\Theta = \Theta_0 \cup \Theta_1$, then $P(\Theta_1|x) > \frac{K_1}{K_1+K_0}$.
In classical terminology, the rejection region of the Bayesian test is $C = \left\{x : P(\Theta_1|x) > \frac{K_1}{K_1+K_0}\right\}$.

Bayesian Inference

Inference losses

Credible sets

If C denotes a credible rule, that is, when x is observed, the set $C(x) \subset \Theta$ will be the credible set for θ, and given the loss function $L(\theta, C(x)) = 1 - I_{C(x)}(\theta)$, then $\rho(\pi(\theta|x), C(x)) = 1 - P^{\pi(\theta|x)}(\theta \in C(x))$.

Measure of credibility

Given $\alpha(x)$ as a measure of the credibility with which it is felt that θ is in $C(x)$, it would be reasonable to measure the accuracy of the report by $L_C(\theta, \alpha(x)) = (I_{C(x)}(\theta) - \alpha(x))^2$. This loss function could be used to suggest a choice of the report $\alpha(x)$. So, the Bayes choice of $\alpha(x)$ is then $P^{\pi(\theta|x)}(\theta \in C(x))$.

Bayesian Inference
Posterior credible sets

Credible sets

Given the posterior $\pi(\theta|x)$, it is generally possible to compute the probability that the parameter θ lies in a particular region Θ_R of the parameter space Θ:
$P(\theta \in \Theta_R|x) = \int_{\Theta_R} \pi(\theta|x)d\theta$.
This is a measure of degree of belief that $\theta \in \Theta_R$ given the sample and prior information.

Credible sets

The set $\Theta_C \in \Theta$ is a $100(1-\alpha)\%$ credible set w.r.t $\pi(\theta|x)$ if:
$P(\theta \in \Theta_C|x) = \int_{\Theta_C} \pi(\theta|x)d\theta = 1-\alpha$.

Bayesian Inference
Highest Posterior Density sets

HPD

A $100(1-\alpha)\%$ Highest Posterior Density set for θ is a $100(1-\alpha)\%$ credible interval for θ with the property that it has a smaller space than any other $100(1-\alpha)\%$ credible set for θ.
$C = \{\theta : \pi(\theta|x) \geq k\}$, where k is the largest number such that $\int_{\theta:\pi(\theta|x)\geq k} \pi(\theta|x)d\theta = 1-\alpha$.
HPDs are very general tool in that they will exist any time the posterior exists. However, they are not rooted firmly in probability theory.

Bayesian Inference
Predictive inference

Loss function

Suppose that one has a loss $L(z, a)$ involving the prediction of Z, so $L(\theta, a) = E_{\theta}^{Z} L(Z, a) = \int L(z, a) g(z|a) dz$, where $g(z|a)$ is the density of Z. So, the prediction problem is reduced to one involving just θ.

Bayesian Inference
Predictive inference

Predictive density

Prediction should be based on the predictive density
$\pi(Z|x) = \int \pi(Z, \theta|x) d\theta = \int \pi(Z|x, \theta)\pi(\theta|x) d\theta$.
The predictive pdf can be used to obtain a point prediction given a loss function $L(Z, z^*)$, where z^* is a point prediction for Z. We can seek z^* that minimizes the mathematical expectation of the loss function.

Bayesian Inference
Model selction

Posterior Model Probability

In addition to learning about parameters or predictions, an econometrician might be interested in comparing different models. Given a set of models $\mathcal{M}\left\{M_1, M_2, \ldots, M_K\right\}$ then,
$\pi(\theta^i|x, M_i) = \frac{f(x|\theta^i, M_i)\pi(\theta^i|M_i)}{f(x|M_i)}, \left\{i = 1, 2, \ldots, K\right\}$.
So, the posterior model probability is $\pi(M_i|x) = \frac{\pi(x|M_i)\pi(M_i)}{f(x)}$,
where the marginal likelihood is equal to
$\pi(x|M_i) = \int f(x|\theta^i, M_i)\pi(\theta^i|M_i)d\theta^i$.

Bayesian Inference
Model selction

Posterior Odds ratio

The posterior odds ratio can be used to compare two models,
$PO_{ij} = \frac{\pi(M_i|x)}{\pi(M_j|x)} = \frac{\pi(x|M_i)\pi(M_i)}{\pi(x|M_j)\pi(M_j)} = BF_{ij} \times Prior\ Odds\ ratio_{ij}$.

Table: Jeffreys Guidelines. See also Kass and Raftery (1995) page 777.

$Log_{10}(PO_{ij}) > 2$	Decisive support for M_i
$3/2 < Log_{10}(PO_{ij}) < 2$	Very strong evidence for M_i
$1 < Log_{10}(PO_{ij}) < 3/2$	Strong evidence for M_i
$1/2 < Log_{10}(PO_{ij}) < 1$	Substantial evidence for M_i
$0 < Log_{10}(PO_{ij}) < 1/2$	Weak evidence for M_i

Bayesian Updating

Now we show the way in which posterior distributions are updated as new information becomes available. Let θ represent one parameter or a vector of parameters, and let x_1 represent the first set of data obtained in a experiment.

$$\pi(\theta|x_1) \propto f(x_1|\theta)\pi(\theta).$$

Next, suppose that a new experiment is perform and new set of data x_2 is obtained. Then the posterior distribution given the complete data set $\pi(\theta|x_1, x_2)$ by the bayes' rules:

Bayesian Updating

$$\begin{aligned}\pi(\theta|x_1, x_2) &\propto f(x_1, x_2|\theta)\pi(\theta) \\ &= f(x_2|x_1, \theta)f(x_1|\theta)\pi(\theta) \\ &= f(x_2|x_1, \theta)\pi(\theta|x_1).\end{aligned} \tag{1}$$

If the data sets are independent, $f(x_2|x_1, \theta)$ simplifies to $f(x_2|\theta)$. Whether or not the data sets are independent, however, note that (1) has the form of a likelihood times a density for θ, but that the latter density is $\pi(\theta|x_1)$: the posterior distribution based on the initial set of data occupies the place where a prior distributions is expected.[1]

[1] Greenberg, E. (2008). *Introduction to Bayesian Econometrics, pag 24.*

Large Samples: Bernstein–von Mises theorem

Consider the case of independent trials, where the likelihood function is:

$$L(\theta|x) = \prod f(x_i|\theta) = \prod L(\theta|x_i).$$

$L(\theta|x_i)$ is the likelihood contribution of x_i. Also define the log likelihood function as:

$$\begin{aligned} l(\theta|x) &= \log L(\theta|x) \\ &= \sum l(\theta|x_i) \\ &= n\bar{l}(\theta|x), \end{aligned}$$

where $\bar{l}(\theta|x) = (1/n) \sum l(\theta|x_i)$ is the mean log likelihood contribution.

Large Samples: Bernstein–von Mises theorem

The posterior distribution can be written as:

$$\begin{aligned} \pi(\theta|x) &\propto \pi(\theta)L(\theta|x) \\ &\propto \pi(\theta)\exp\left(n\bar{l}(\theta|x)\right). \end{aligned}$$

For large n, the exponential term dominates $\pi(\theta)$, which does not depend on n. Accordingly, we can expect that the prior distribution will play a relatively smaller role than do the data, as reflected in the likelihood function, when the sample size is large.

Large Samples: Bernstein–von Mises theorem

If we denote the true value of θ by θ_0, it can be show that

$$\lim_{n\to\infty} \bar{l}(\theta|x) \to \bar{l}(\theta_0|x).$$

Accordingly, for large n, the posterior distribution collapses to a distribution with all its probability at θ_0. This property is similar to the criterion of **consistency** in the frequentist literature and extends to the multiparameter case.

Large Samples: Bernstein–von Mises theorem

For large sample, the posterior distribution can therefore be written approximately as:

$$\pi(\theta|x) \propto \pi(\theta) \exp\left[-\frac{n}{2\nu}(\theta-\hat{\theta})^2\right],$$

where $\nu = [-\bar{l}''(\hat{\theta}|x)]^{-1}$. The second term is in the form of a normal distribution with mean $\hat{\theta}$ and variance ν/n, and it dominates $\pi(\theta)$ because of the n in the exponential. See Greenberg.[2])

[2] Greenberg, E. (2008). *Introduction to Bayesian Econometrics, pag 27*

Large Samples: Bernstein–von Mises theorem

In summary, when n is large:[3]

1. The prior distribution plays a relatively small role in determining the posterior distribution.
2. The posterior distribution converges to a degenerate distribution at the true value of the parameter.
3. The posterior distribution is approximately normally distributed with mean $\hat{\theta}$.

[3] Greenberg, E. (2008). *Introduction to Bayesian Econometrics, pag 27.*

References I

Kass, R. and Raftery, A. (1995). Bayes factors. *Journal of The American Statistical Association*, 90(430):773–795.