Bayesian Econometrics

A brief summary of theory

Andrés Ramírez Hassan

Universidad Eafit
Departamento de Economía

February 21, 2021

Outline

The Bayes' rule

Bayes' rule

As expected the point of departure to perform Bayesian inference is the Bayes' rule, that is, the conditional probability of A_i given B is equal to the conditional probability of B given A_i times the marginal probability of A_i over the marginal probability of B,

$$P(A_i|B) = \frac{P(A_i, B)}{P(B)} = \frac{P(B|A_i) \times P(A_i)}{P(B)}, \tag{1}$$

where $P(B) = \sum_i P(B|A_i)P(A_i) \neq 0$, $\{A_i, i = 1, 2, \dots\}$ is a finite or countably infinite partition of a sample space.

The Bayes' rule

The base rate fallacy

Assume that the sample information comes from a positive result from a test whose true positive rate (sensitivity) is 98%, $P(+|\text{Desease}) = 0.98$. On the other hand, the prior information regarding being infected with this disease comes from a base incidence rate that is equal to 0.002, that is $P(\text{Disease}) = 0.002$. Then, **what is the probability of being actually infected?**

$$P(\text{disease}|+) = \frac{P(+|\text{disease}) \times P(\text{disease})}{P(+)},$$

where $P(+) =$
$P(+|\text{disease}) \times P(\text{disease}) + P(+|\neg\text{disease}) \times P(\neg\text{disease})$.

The Bayes' rule

God existence

Let's say that there are two cases of resurrection (Res), Jesus Christ and Elvis, and the total number of people who have ever lived is 108.5 trillion, then the prior base rate is 2/108,500,000,000. On the other hand, the sample information comes from a very reliable witness whose true positive rate is 0.9999999. Then, **what is the probability of this miracle?**

$$P(\text{Res}|\text{Witness}) = \frac{P(\text{Witness}|\text{Res}) \times P(\text{Res})}{P(\text{Witness})},$$

where $P(\text{Witness}) =$
$P(\text{Witness}|\text{Res}) \times P(\text{Res}) + (1 - P(\text{Witness}|\text{Res})) \times (1 - P(\text{Res}))$.

The Bayes' rule

Conditional version of the Bayes' rule

Let's have two conditioning events B and C, then equation 1 becomes

$$P(A_i|B,C) = \frac{P(A_i,B,C)}{P(B,C)}$$
$$= \frac{P(B|A_i,C) \times P(A_i|C) \times P(C)}{P(B|C)P(C)},$$

The Bayes' rule

The Monty Hall problem

This was the situation faced by a contestant in the American television game show *Let's Make a Deal*. There, the contestant was asked to choose a door where behind one door there is a car, and behind the others, goats. Let's say that the contestant picks door No. 1, and the host (Monty Hall), who knows what is behind each door, opens door No. 3, where there is a goat. Then, the host asks the tricky question to the contestant, **do you want to pick door No. 2?**

The Bayes' rule

The Monty Hall problem

Let's name P_i the event *contestant picks door No. i*, H_i the event *host picks door No. i*, and C_i the event *car is behind door No. i*. In this particular setting, the contestant is interested in the probability of the event $P(C_2|H_3, P_1)$. The important point here is that the host knows what is behind each door and randomly picks a door given contestant choice. That is, $P(H_3|C_3, P_1) = 0$, $P(H_3|C_2, P_1) = 1$ and $P(H_3|C_1, P_1) = 1/2$.

The Bayes' rule

$$
\begin{aligned}
P(C_2|H_3, P_1) &= \frac{P(C_2, H_3, P_1)}{P(H_3, P_1)} \\
&= \frac{P(H_3|C_2, P_1)P(C_2|P_1)P(P_1)}{P(H_3|P_1) \times P(P_1)} \\
&= \frac{P(H_3|C_2, P_1)P(C_2)}{P(H_3|P_1)} \\
&= \frac{1 \times 1/3}{1/2} \\
&= \frac{2}{3},
\end{aligned}
$$

A brief summary of theory

Bayes' rule

For two random objects θ and y, the Bayes' rule may be analogously used,

$$\pi(\theta|\mathrm{y}) = \frac{p(\mathrm{y}|\theta) \times \pi(\theta)}{p(\mathrm{y})}, \tag{2}$$

where $\pi(\theta|\mathrm{y})$ is the posterior density function, $\pi(\theta)$ is the prior density, $p(\mathrm{y}|\theta)$ is the likelihood (statistical model), and $p(\mathrm{y}) = \int_{\Theta} p(\mathrm{y}|\theta)\pi(\theta)d\theta$ is the marginal likelihood or prior predictive.

A brief summary of theory

Bayes' rule

For two random objects θ and y, the Bayes' rule may be analogously used,

$$\pi(\theta|\mathrm{y}) = \frac{p(\mathrm{y}|\theta) \times \pi(\theta)}{p(\mathrm{y})} \tag{3}$$

$$\propto p(\mathrm{y}|\theta) \times \pi(\theta), \tag{4}$$

where $\pi(\theta|\mathrm{y})$ is the posterior density function, $\pi(\theta)$ is the prior density, $p(\mathrm{y}|\theta)$ is the likelihood (statistical model), and $p(\mathrm{y}) = \int_{\Theta} p(\mathrm{y}|\theta)\pi(\theta)d\theta$ is the marginal likelihood or prior predictive.

A brief summary of theory

Model uncertainty

Observe that the Bayesian inferential approach is conditional, that is, what can we learn about an unknown object θ given that we already observed y? The answer is also conditional on the probabilistic model, that is $p(\mathrm{y}|\theta)$. So, what if we want to compare different models, let's say $\mathcal{M}_m$, $m = \{1, 2, \dots, M\}$.

$$\pi(\theta|\mathrm{y}, \mathcal{M}_m) = \frac{p(\mathrm{y}|\theta, \mathcal{M}_m) \times \pi(\theta|\mathcal{M}_m)}{p(\mathrm{y}|\mathcal{M}_m)}. \tag{5}$$

A brief summary of theory

The posterior model probability is

$$\pi(\mathcal{M}_m|\mathrm{y}) = \frac{p(\mathrm{y}|\mathcal{M}_m) \times \pi(\mathcal{M}_m)}{p(\mathrm{y})}, \tag{6}$$

where $p(\mathrm{y}|\mathcal{M}_m) = \int_{\Theta} p(\mathrm{y}|\theta, \mathcal{M}_m) \times \pi(\theta|\mathcal{M}_m) d\theta$ due to equation 5, and $\pi(\mathcal{M}_m)$ is the prior model probability.

A brief summary of theory

Posterior odds

We can avoid calculating $p(\text{y})$ when performing model selection (hypothesis testing) using posterior odds ratio, that is, comparing models $\mathcal{M}_1$ and $\mathcal{M}_2$,

$$PO_{12} = \frac{\pi(\mathcal{M}_1|\text{y})}{\pi(\mathcal{M}_2|\text{y})} = \frac{\pi(\text{y}|\mathcal{M}_1)}{\pi(\text{y}|\mathcal{M}_2)} \times \frac{\pi(\mathcal{M}_1)}{\pi(\mathcal{M}_2)}, \tag{7}$$

where the first term in equation 7 is named the Bayes Factor, and the second term is the prior odds.

A brief summary of theory

Posterior probabilities from posterior odds

Given two models $\mathcal{M}_1$ and $\mathcal{M}_2$ such that $\pi(\mathcal{M}_1|\text{y}) + \pi(\mathcal{M}_2|\text{y}) = 1$. Then, $\pi(\mathcal{M}_1|\text{y}) = \frac{PO_{12}}{1+PO_{12}}$ and $\pi(\mathcal{M}_2|\text{y}) = 1 - \pi(\mathcal{M}_1|\text{y})$.
In general, $\pi(\mathcal{M}_m|\text{y}) = \frac{\pi(\text{y}|\mathcal{M}_m)\times\pi(\mathcal{M}_m)}{\sum_{l=1}^{M}\pi(\text{y}|\mathcal{M}_l)\times\pi(\mathcal{M}_l)}$.

A brief summary of theory

$2\log(PO_{12})$	PO_{12}	Evidence against M_2
0 to 2	1 to 3	Not worth more than a bare mention
2 to 6	3 to 20	Positive
6 to 10	20 to 150	Strong
>10	>150	Very strong

Table: Kass and Raftery guidelines (1995)

A brief summary of theory

Probabilistic predictions

We can also obtain a posterior predictive distribution,

$$\pi(\mathrm{y}_0|\mathrm{y},\mathcal{M}_m) = \int_\Theta \pi(\mathrm{y}_0,\theta|\mathrm{y},\mathcal{M}_m)d\theta$$
$$= \int_\Theta \pi(\mathrm{y}_0|\theta,\mathrm{y},\mathcal{M}_m)\pi(\theta|\mathrm{y},\mathcal{M}_m)d\theta. \quad (8)$$

Observe that equation 8 is a posterior expectation $\mathbb{E}[\pi(\mathrm{y}_0|\theta,\mathrm{y},\mathcal{M}_m)]$. This is a very common feature in Bayesian inference that is suitable for computation based on Monte Carlo integration. In addition, the Bayesian approach takes estimation error into account.

A brief summary of theory

Model uncertainty in prediction

If we want to consider model uncertainty in prediction or any unknown probabilistic object, we can follow same arguments. In the prediction case,

$$\pi(\mathrm{y}_0|\mathrm{y}) = \sum_{m=1}^{M} \pi(\mathcal{M}_m|\mathrm{y})\pi(\mathrm{y}_0|\mathrm{y}, \mathcal{M}_m), \tag{9}$$

A brief summary of theory

Model uncertainty in parameters' inference

In parameters,

$$\pi(\theta|\mathrm{y})=\sum_{m=1}^{M}\pi(\mathcal{M}_m|\mathrm{y})\pi(\theta|\mathrm{y},\mathcal{M}_m), \tag{10}$$

where $\mathbb{E}(\theta|\mathrm{y})=\sum_{m=1}^{M}\hat{\theta}_m\pi(\mathcal{M}_m|\mathrm{y})$,, $Var(\theta|\mathrm{y})=\sum_{m=1}^{M}\pi(\mathcal{M}_m|\mathrm{y})\widehat{Var}(\theta|\mathrm{y},\mathcal{M}_m)+\sum_{m=1}^{M}\pi(M_m|\mathrm{y})(\hat{\theta}_m-\mathbb{E}[\theta|\mathrm{y})]^2$, $\hat{\theta}_m$ and $\widehat{Var}(\theta|\mathrm{y},\mathcal{M}_m)$ are the posterior mean and variance under model m, respectively.

A brief summary of theory

Bayesian updating

A nice advantage of the Bayesian approach, which is very useful in *state space models*, is the way that the posterior distribution updates with new sample information. Given $\mathrm{y} = \mathrm{y}_{1:t+1}$ a sequence of observations, then

$$\begin{aligned}
\pi(\theta|\mathrm{y}_{1:t+1}) &\propto p(\mathrm{y}_{1:t+1}|\theta) \times \pi(\theta) \\
&= p(y_{t+1}|\mathrm{y}_{1:t}, \theta) \times p(\mathrm{y}_{1:t}|\theta) \times \pi(\theta) \\
&\propto p(y_{t+1}|\mathrm{y}_{1:t}, \theta) \times \pi(\theta|\mathrm{y}_{1:t}).
\end{aligned}$$

A brief summary of theory

Bayesian updating

This is particular useful under the assumption of *conditional independence*, that is, $y_{t+1} \perp \mathrm{y}_{1:t}|\theta$, then $p(y_{t+1}|\mathrm{y}_{1:t}, \theta) = p(y_{t+1}|\theta)$ such that the posterior can be recovered recursively. This facilities online updating due to all information up to t being in θ. Then, $\pi(\theta|\mathrm{y}_{1:t+1}) \propto p(y_{t+1}|\theta) \times \pi(\theta|\mathrm{y}_{1:t}) \propto \prod_{h=1}^{t+1} p(y_h|\theta) \times \pi(\theta)$.

A brief summary of theory

Sampling properties of Bayesian "estimators"

$$\begin{aligned}\pi(\theta|y) &\propto \exp\{l(y|\theta)\} \times \pi(\theta)\\ &\approx \exp\left\{l(y|\hat{\theta}) - \frac{N}{2\sigma^2}(\hat{\theta}-\theta_0)^2\right\} \times \pi(\theta)\\ &\propto \exp\left\{-\frac{N}{2\sigma^2}(\hat{\theta}-\theta_0)^2\right\} \times \pi(\theta)\end{aligned}$$

Observe that we have that the posterior density is proportional to the kernel of a normal density with mean $\hat{\theta}$ and variance σ^2/N as long as $\pi(\hat{\theta}) \neq 0$. This kernel dominates as the sample size gets large due to N in the exponential term.

Bayesian Inference

Estimation problems

Result 1

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

Result 2

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 3

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

Result 4

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 5

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 $L(\theta, a_i) = \left\{ \begin{matrix} 0, \theta \in \Theta_i \\ K_i, \theta \in \Theta_j (j \neq i) \end{matrix} \right\}$ 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 5

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 $L(\theta, a) = E_\theta^Z L(Z, a) = \int L(z, a) g(z|\theta) dz$, where $g(z|\theta)$ 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.

A brief summary of theory

Summary

I presented the basic theory concepts of Bayesian inference.
We are done in this course!!!