Bayesian Econometrics

Dutch Book

Andrés Ramírez Hassan

Universidad Eafit
Departamento de Economía

May 20, 2021

Outline

1. Basic Concepts

Introduction

In this course we perform an introduction to Bayesian methods, we show some basic definitions and properties of the bayesian approach. We have taken the content from the work by some authors as Greenberg (2008), Berger (1985), Albert(2007) and Robert & Casella (2004).[1]

1

- Greenberg, E. (2008). Introduction to Bayesian Econometrics. *Springer*.
- Berger, J. (1985). Statistical Decision Theory and Bayesian Analysis. *Springer-Verlag*.
- Robert, C. Casella G. (2004). Monte Carlo Statistical Methods. *Springer*. Second Edition.
- Albert, J. (2007). Bayesian Computation with R.*Springer*.

Introduction

- After this brief epistemological introduction, we show in the first part of this course the basic ideas of the Bayesian approach. In particular, the statistical decision theory foundations of the Bayesian approach, and some simple examples.
- In the second part we introduce some concepts related to elicitation, that is, how to transform experts' knowledge in probabilistic statements.
- In the third part we show some basic conjugate families, which allow to obtaining easily the posterior distribution.

Introduction

- In the fourth part, we study some simulation techniques as Markov chain Monte Carlo (MCMC), which is a flexible simulation method that can deal with a wide variety of models.
- In fifth part we apply MCMC techniques to models commonly encountered in econometrics and statistics. We will emphasize the design of algorithms to analyze these models as a way of preparing the reader to develop algorithms for the new models.
- Finally, the sixth part is devoted to Bayesian Model Averaging, which is a technique to introduce model uncertainty.

Basic Concepts

In probability theory, the sample space, that we denote Ω, of an experiment or random trial is the set of all possible outcomes. A probability is a number assigned to statements or events. Examples of such statements are:

- A_1 : A coin tossed three times will come up heads either two or three times.
- A_2 : A six-sided die rolled once shows an even number of spots.
- A_3 : There will be measurable precipitation on January 1, 2008, at your local airport.

Basic Concepts

Standard notation:

- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
- $P(A \cap B) = P(A) + P(B) - P(A \cup B)$.
- $A \cup A^c = \Omega$

Probability Axioms

1. $P(A) \geqslant 0$.
2. $P(\Omega) = 1$.
3. If $A_1, A_2, \ldots,$ are pairwise disjoint, then $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

1. Let $P(A|B)$ denote the probability of A, given that B is true. Then $P(A|B) = \frac{P(A \cap B)}{P(B)}$. This is the Bayes' rule.

Subjective Probability

The calculus of probability as a quantification of the uncertainty in decision making.

Coherence

A bookmakers' betting odds are coherent if a client cannot place a bet or a combinations of bets such that no matter what outcome occurs, the bookmaker will lose money.

Assumptions

1. The odds are fair to the bookmaker, that is the bookmaker is willing to both sell and buy on any of the events posted.
2. There is no restriction about the number of bets that clients can buy or sell, as long as this is finite.

Subjective Probability

The Dutch book theorem

If assumptions 1 and 2 hold, a necessary condition for a set of prices to be coherent is to satisfy Kolmogorov's axioms.

Betting odds

$$k = \frac{1 - P(A)}{P(A)}$$

where A is the event, and k is is the odds against A. In addition, let's S the number of tickets. $S > 0$ means that we are betting that A occurs, and $S < 0$ means that we are betting against A.

Subjective Probability

Payout

- If A occurs and we bet on it, we would "receive" $S(1+k)$, a positive number because $S > 0$ if we bet on A.
- If A occurs and we bet against it, we would "receive" $S(1+k)$, a negative number because $S < 0$ if we bet against A.

De Finetti betting setup

The price of the ticket p is set by us, the payout is fixed at 1, and your opponent chooses S.
So, a winning ticket on A would pay $p(1+k)$. But in the de Finetti setup, the payout is 1, so $k = (1-p)/p$, which implies $p = P(A)$.

Subjective Probability

Axioms

Event	Our gain
A	$(1-p)S$
A^c	$-pS$

1. If $p < 0$, our opponent, by choosing $S < 0$, will inflict a loss.
2. If $p > 1$, your opponent can set $S > 0$, and you are again sure to lose.

Subjective Probability

Axioms

If you are certain that A will occur, coherency dictates that we set $p = 1$.

Axioms

Event	Our gain
A_1	$(1 - p_1)S_1 - p_2 S_2 + (1 - p_3)S_3$
A_2	$-p_1 S_1 + (1 - p_2)S_2 + (1 - p_3)S_3$
$(A_1 \cup A_2)^c$	$-p_1 S_1 - p_2 S_2 - p_3 S_3$

Consider betting in three events: A_1, A_2 and $A_1 \cup A_2$. We should set p_1, p_2 and p_3 such that we do not have a sure loss.

Subjective Probability

Axioms

If you are certain that A will occur, coherency dictates that we set $p = 1$.

Axioms

Event	Our gain
A_1	$(1 - p_1)S_1 - p_2 S_2 + (1 - p_3)S_3$
A_2	$-p_1 S_1 + (1 - p_2)S_2 + (1 - p_3)S_3$
$(A_1 \cup A_2)^c$	$-p_1 S_1 - p_2 S_2 - p_3 S_3$

Consider betting in three events: A_1, A_2 and $A_1 \cup A_2$. We should set p_1, p_2 and p_3 such that we do not have a sure loss. This implies $p_1 + p_2 = p_3$.

Subjective Probability

Axioms

Event	Our gain
AB	$(1-p_1)S_1 + (1-p_2)S_2 + (1-p_3)S_3$
BA^c	$-p_1S_1 + (1-p_2)S_2 - p_3S_3$
AB^c	$-p_1S_1 - p_2S_2$
$(A \cup B)^c$	$-p_1S_1 - p_2S_2$

Consider betting in three events: AB, B and $A|B$. We should set p_1, p_2 and p_3 such that we do not have a sure loss.

Subjective Probability

Axioms

Event	Our gain
AB	$(1-p_1)S_1 + (1-p_2)S_2 + (1-p_3)S_3$
BA^c	$-p_1S_1 + (1-p_2)S_2 - p_3S_3$
AB^c	$-p_1S_1 - p_2S_2$
$(A \cup B)^c$	$-p_1S_1 - p_2S_2$

Consider betting in three events: AB, B and $A|B$. We should set p_1, p_2 and p_3 such that we do not have a sure loss. This implies $p_1 = p_2p_3$.

Notation for Density and Distributions Functions

- $\pi(\cdot)$ denotes a prior and $\pi(\cdot|y)$ denotes posterior density function of parameters; these densities are continuous random variables in the statistical models we discuss.
- $f(y|\theta)$ denotes the likelihood function.
- $p(\cdot)$ denotes the probability mass function (p.m.f) of a discrete random variable.
- $f(\cdot)$ denotes the probability density function (p.d.f) for continuous data. $F(\cdot)$ denotes the distribution function (d.f) for continuous data.

Prior, Likelihood, and Posterior

Now, we show how to obtain the posterior distribution from the likelihood function and the prior distribution. By the Bayes' rule:

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

where

$$f(x) = \int f(x|\theta)\pi(\theta)d\theta.$$

The interesting thing in this approach is that we can mix two information sources: Prior information available or knowledge from the expert source (through prior distribution) and data obtained from an experiment or observation (through the likelihood function).

Prior, Likelihood, and Posterior

To illustrate how we can obtain a posterior distribution from the likelihood function and prior distribution we show the following example taken from Albert:[2]

- Suppose a person is interested in learning about the sleeping habits of American college students. This person hears that doctors recommend eight hours of sleep for an average adult. What proportion of college students get at least eight hours of sleep?
- Here we think of a population consisting of all American college students and let p represent the proportion of this population who sleep at least eight hours. We are interested in learning about the location of p.

[2] Albert, J. (2007). Bayesian Computation with R. *Springer*. pag 19.

Prior, Likelihood, and Posterior

- To learn about p, a first step is to take a random sample of students from some university.
- But, before taking the sample the person can do an initial research to learn about the sleeping habits of college students. This research will help her in constructing a prior distribution.
- A simple approach for assessing a prior for p is to write down a list of plausible proportion values and then assign weights to these values. In this example, the person believes that:

p	0.05	0.15	0.25	0.35	0.45	0.55	0.65	0.75	0.85	0.95
weights	0.0625	0.125	0.25	0.25	0.125	0.0625	0.03125	0.03125	0.03125	0.03125

Based on her beliefs, she assigns these values the corresponding weights.

Prior, Likelihood, and Posterior

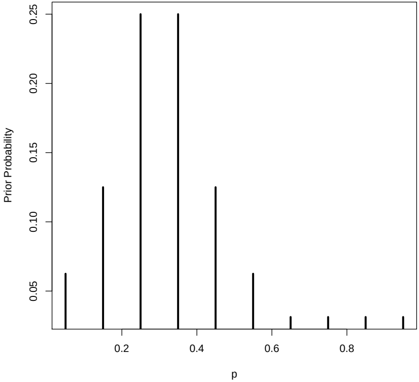

Figure: Discrete Prior for p, proportion of who sleep at least eight hours.

Prior, Likelihood, and Posterior

- A sample of 27 students is taken in this group, 11 record that they had at least eight hours of sleep the previous night. Based on the prior information and this observed data, the person is interested in estimating the proportion p.
- If we regard a success that a person slept at least eight hours and we take a random sample with s success and f failures, then the likelihood function is given by

$$L(y|p) \propto p^s(1-p)^f.$$

Prior, Likelihood, and Posterior

Suppose that our prior density for p is denoted by $g(p)$. The posterior density for p, by Bayes' rule, is obtained, up to a proportionality constant, by multiplying the prior density by the likelihood

$$g(p|y) \propto g(p)L(y|p).$$

Prior, Likelihood, and Posterior

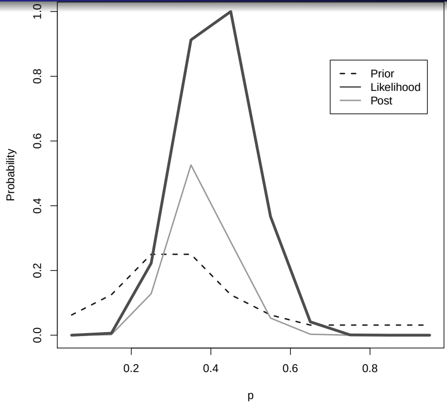

Figure: Prior, Likelihood and Posterior distribution for p, proportion of who sleep at least eight hours.

Prior, Likelihood, and Posterior

Distribution Beta as Prior

Because the proportion is a continuous parameter, we can use an alternative approach, defining a density $g(p)$ on interval $(0, 1)$ that represents the person's initial beliefs. Suposse she believes that the porportion is equally likely to be smaller or larger than $p = 0.3$. Moreover, she is 90% confident that p is less than 0.5. A convenient family of densities for a proportion is the beta with kernel proportional to

$$g(p) \propto p^{a-1}(1-p)^{b-1},$$

where the election of the hyperparameters a y b are chosen to reflect the user's prior beliefs about p.

Prior, Likelihood, and Posterior

Distribution Beta as Prior

Here the person believes that the median and 90th percentiles are given, respectively, by 0.3 and 0.5 and this can be matched, by trial and error, with a beta density with $a = 3.4$ and $b = 7.4$. Combining this beta prior with the likelihood function, one can show that the posterior density is also of the beta form update parameters $a + s$ and $b + f$

$$g(p|y) \propto p^{a+s-1}(1-p)^{b+f-1},$$

$$0 < p < 1.$$

Prior, Likelihood, and Posterior

Distribution Beta as Prior

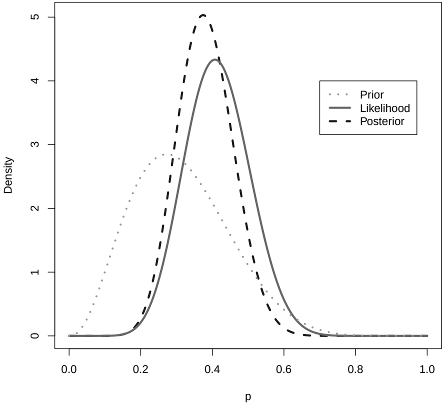

Figure: Prior, likelihood and Posterior distribution for p, proportion of who sleep at least eight hours.