Bayesian Econometrics

Andrés Ramírez Hassan

Universidad Eafit
Departamento de Economía

January 17, 2017

Outline

1 Introduction

Why you should be a Bayesian Econometrician

Peter Diggle, President of The Royal Statistical Association (2014-2016)

A different trend which has surged upwards in statistics during Peter's career is the popularity of "Bayesian" statistics. Does Peter consider himself a "Bayesian"'? Well, he replies, you can't not believe in Bayes' theorem because it's true. But that doesn't make you a Bayesian in the philosophical sense. When people are making personal decisions – even if they don't formally process Bayes' theorem in their mind – they are adapting what they think they should believe in response to new evidence as it comes in. Bayes' theorem is just the formal mathematical machinery for doing that.

Warming up

Who wants to be a millionaire?

In the first stage in the TV show "Who wants to be a millionaire?" you are asked to answer three very simple questions like:
What is the name of the actor who plays "El Chavo"?

Who wants to be a millionaire?

What is the probability that you overcome the first stage in the TV show "Who wants to be a millionaire?"

Warming up

Confidence Intervals

- $P(\beta \in [\hat{\beta} - t_{N-k}^{\alpha/2}\hat{\sigma}_{\hat{\beta}}, \hat{\beta} + t_{N-k}^{\alpha/2}\hat{\sigma}_{\hat{\beta}}]) = 1 - \alpha$.
- Is $P(\beta \in [0.2, 0.4]) = 0.95$ true?

Warming up

Confidence Intervals

- $P(\beta \in [\hat{\beta} - t^{\alpha/2}_{N-k}\hat{\sigma}_{\hat{\beta}}, \hat{\beta} + t^{\alpha/2}_{N-k}\hat{\sigma}_{\hat{\beta}}]) = 1 - \alpha.$
- Is $P(\beta \in [0.2, 0.4]) = 0.95$ true? NOT

Warming up

Confidence Intervals

- $P(\beta \in [\hat{\beta} - t_{N-k}^{\alpha/2}\hat{\sigma}_{\hat{\beta}}, \hat{\beta} + t_{N-k}^{\alpha/2}\hat{\sigma}_{\hat{\beta}}]) = 1 - \alpha.$
- Is $P(\beta \in [0.2, 0.4]) = 0.95$ true? NOT

Credible Intervals

- $P(\beta \in [\hat{\beta} - t_{N-k}^{\alpha/2}\hat{\sigma}_{\hat{\beta}}, \hat{\beta} + t_{N-k}^{\alpha/2}\hat{\sigma}_{\hat{\beta}}]) = 1 - \alpha.$
- Is $P(\beta \in [0.2, 0.4]) = 0.95$ true?

Warming up

Confidence Intervals

- $P(\beta \in [\hat{\beta} - t_{N-k}^{\alpha/2}\hat{\sigma}_{\hat{\beta}}, \hat{\beta} + t_{N-k}^{\alpha/2}\hat{\sigma}_{\hat{\beta}}]) = 1 - \alpha.$
- Is $P(\beta \in [0.2, 0.4]) = 0.95$ true? NOT

Credible Intervals

- $P(\beta \in [\hat{\beta} - t_{N-k}^{\alpha/2}\hat{\sigma}_{\hat{\beta}}, \hat{\beta} + t_{N-k}^{\alpha/2}\hat{\sigma}_{\hat{\beta}}]) = 1 - \alpha.$
- Is $P(\beta \in [0.2, 0.4]) = 0.95$ true? YES

Why you should be a Bayesian Econometrician

Degrees of belief does not imply to be subjective

The main difference between the Bayesians and the frequentists is not subjectivity.

Degrees of belief does not imply to be subjective

- Subjective Bayesians: Ramsey, de Finetti, Savage and Lindley
- Objective Bayesians: Bayes, Laplace, Jeffreys and Berger

Why you should be a Bayesian Econometrician

Degrees of belief does not imply to be subjective

What is objectivity in a frequentist approach?

Why you should be a Bayesian Econometrician

Degrees of belief does not imply to be subjective

What is objectivity in a frequentist approach?
For example, why should we use a 5% or 1% significance level rather than any other value?

Why you should be a Bayesian Econometrician

Degrees of belief does not imply to be subjective

What is objectivity in a frequentist approach?
For example, why should we use a 5% or 1% significance level rather than any other value?
As someone said, the apparent objectivity is really a consensus (Lindley, 2000).

Why you should be a Bayesian Econometrician

Degrees of belief does not imply to be subjective

What is objectivity in a frequentist approach?
For example, why should we use a 5% or 1% significance level rather than any other value?
As someone said, the apparent objectivity is really a consensus (Lindley, 2000).
In fact "Student" (William Gosset) saw statistical significance at any level as being "nearly valueless" in itself ("Student" himself was a philosophical Bayesian) (Ziliak, 2008).

Why you should be a Bayesian Econometrician

Degrees of belief does not imply to be subjective

What is a p–value?

Why you should be a Bayesian Econometrician

Degrees of belief does not imply to be subjective

What is a p–value?

A p–value is the probability of obtaining a statistical summary of the data equal to or "more extreme" than what was actually observed, assuming that the null hypothesis is true. p–value calculations involve not just the observed data, but also more "extreme" hypothetical observations. So,

> *". . . a hypothesis which may be true may be rejected because it has not predicted observable results which have not occurred."*
>
> *Jeffreys (1961)*

Why you should be a Bayesian Econometrician

The p–value fallacy

In fact common frequentist inferential practice intertwined two incompatible theoretical frameworks: the p–value (Fisher, 1958) and hypothesis test (Neyman and Pearson, 1933). The former is an informal short–run criterion, whose philosophical foundation is *reduction to absurdity*, which measures the discrepancy between the data and the null hypothesis. So, the p–value is not a direct measure of the probability that the null hypothesis is false.

Why you should be a Bayesian Econometrician

The p–value fallacy

The latter, whose philosophical foundations is *deduction*, is based on a long–run performance such that controls the overall number of incorrect inferences without care of individual cases. The p–value fallacy consists in interpreting the p–value as the strength of evidence against the null hypothesis and the frequency of type I error under the null hypothesis (Goodman, 1999). Unfortunately, this is a common mistake in applied research.

Why you should be a Bayesian Econometrician

The p–value fallacy

> *"statistical techniques for testing hypothesis... have more flaws than Facebook's privacy policies"*
>
> *Siegfried (2014)*

Why you should be a Bayesian Econometrician

The *p*–value fallacy: Wasserstein and Lazar (2016)

- P–values can indicate how incompatible the data are with a specified statistical model.
- P–values do not measure the probability that the studied hypothesis is true.
- Scientific conclusions and business or policy decisions should not be based only on whether a p–value passes a specific threshold.
- A p–value, or statistical significance, does not measure the size of an effect or the importance of a result.
- By itself, a p–value does not provide a good measure of evidence regarding a model or hypothesis.

Why you should be a Bayesian Econometrician

The *p*–value fallacy

The p–value is associated with the probability of the data given the hypothesis, whereas the main concern in science should be the probability of the hypothesis given the data. The former is associated with the Bayes factor, whereas the p–value is a great answer to a wrong question. It seems that traditional scientific inference suffers from the prosecutor's fallacy.

Why you should be a Bayesian Econometrician

The p–value fallacy

The simultaneous use of the p–value and significance level should take into account that the significance level should dependent on sample size. On the contrary we may not reject the statistical relevance of theoretical meaningless controls.

Why you should be a Bayesian Econometrician

The *p*–value fallacy

Fortunately, there is an approximate link between the orthodox t statistic and the Bayes factor for regression coefficients (Raftery, 1995). In particular, $|t| > (log(N) + 6)^{1/2}$, where N is the sample size, corresponds to strong evidence in favor of rejecting the not relevance of a control in a regression.

Observe that this setting agrees with the idea in experimental designs of selecting the sample size such that we control Type I and Type II errors. In observational studies we cannot control the sample size, but we can select the significance level.

Why you should be a Bayesian Econometrician

What are the differences between Bayesians and frequentists?

In my humble opinion, the differences between these two statistical approaches are not related to subjectivity versus objectivity in scientific research or pragmatic settings. The differences are philosophical, methodological and pedagogical. Although at methodological level, the debate has become considerably muted, except for some aspects of inference, with the recognition that each approach has a great deal to contribute to statistical practice (Good, 1992, Bayarri and Berger, 2004, Kass, 2011).

Why you should be a Bayesian Econometrician

What are the differences between Bayesians and frequentists?

- Deductive Inference: Given that the Central Bank increases intervention interest rate (premise) then inflation rate should decrease (consequence)
- Inductive Inference: What cause explains a decrease of inflation rate (consequence)?

Why you should be a Bayesian Econometrician

What are the differences between Bayesians and frequentists?

There is a huge difference between *effects of causes* (forward causal inference) and *causes of effects* (reverse causal inference) (Gelman, A. and Imbens, G., 2013, Dawid et al., 2016).

Why you should be a Bayesian Econometrician

What are the differences between Bayesians and frequentists?

It is argued that the philosophical grounds of Bayesian statistics are stronger than the frequentist approach, and this is achieved at negligible logical cost (Jeffreys, 1931, 1961). In particular, the Bayesian framework is based on *inductive* inference (*Inverse probability*); on the basis of what we see, we evaluate what hypothesis is most tenable, whereas the frequentist framework is based on *deductive* inference.

Why you should be a Bayesian Econometrician

What are the differences between Bayesians and frequentists?

The frequentist approach just uses sample information to perform inference, whereas the Bayesian approach uses sample and non-sample information to accomplish such task (Judge et al., 1985).

Why you should be a Bayesian Econometrician

What are the differences between Bayesians and frequentists?

In making inference or decisions about the state of the nature in the Bayesian paradigm, all the relevant *experimental* information is given by the observed data. Then the relevance of the *Likelihood principle* in this statistical approach; this framework is conditioned to data. A characteristic that is not present in the orthodox frequentist approach (Berger, 1993).

Why you should be a Bayesian Econometrician

Likelihood principle

We are given a coin and are interested in the probability, θ, of having it come up heads when flipped. It is desired to test $H_0 : \theta = 1/2$ versus $H_1 : \theta > 1/2$. An experiment is conducted by flipping the coin (independently) in a series of trials, the results of which is the observation of 9 heads and 3 tails. This is not yet enough information to specify $f(y|\theta)$, since the series of trials was not explained. Two possibilities:

The experiment consisted of a predetermine 12 flips, so that $X = [Heads]$ would be $\mathcal{B}(12, \theta)$.

The experiment consisted of flipping the coin until 3 tails were observed, so that X would be $\mathcal{NB}(3, \theta)$.

(Berger, 1993) following (Lindley and Phillips, 1976)

Why you should be a Bayesian Econometrician

Likelihood principle

$l_1(\theta) = f_1(y|\theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y} = 220\theta^9(1-\theta)^3$
$l_2(\theta) = f_2(y|\theta) = \binom{n+y-1}{y}\theta^y(1-\theta)^{n-y} = 55\theta^9(1-\theta)^3$
Using a frequentist approach, the significance level of $y = 9$ using the Binomial model against $\theta = 1/2$ would be:

$$\alpha_1 = P_{1/2}(Y \geq 9) = f_1(9|1/2) + f_1(10|1/2) + f_1(11|1/2) + f_1(12|1/2) = 0.075$$

For the Negative Binomial model, the significance level would be:

$$\alpha_2 = P_{1/2}(Y \geq 9) = f_2(9|1/2) + f_2(10|1/2) + \ldots = 0.0325$$

We arrive to a different conclusions using a significance level equal to 5%.

Why you should be a Bayesian Econometrician

Differences

- frequentists consider probability as a physical phenomenon v.s Bayesians consider that probability lives in the mind of scientists, like most of the theoretic constructs in science (Parmigiani and Inoue, 2008).
- frequentists consider probability associated with variability v.s. Bayesians consider that probability is associated with uncertainty.
- The good properties of the frequentist estimators are asymptotic v.s. The statistical inference in the Bayesian approach is exact.

Why you should be a Bayesian Econometrician

Differences

Finally, we might say that *roughly speaking* statistics is the study of uncertainty (Lindley, 2000), although some statisticians may disagree, under this definition the only coherent way of handing uncertainty is through the theory of probability. Bayesian statistics follows this approach, whereas the frequentists prefer not to follow this principle.

Why you should be a Bayesian Econometrician

Bayesian perspective

We should have in mind that from a Bayesian perspective the true state of the nature is deterministic but unknown. Thus, the prior distribution is a way to reflect knowledge regarding nature before getting data. Then, the posterior distribution is an update of previous knowledge using new experimental information. This is precisely the way that human beings use to learn.

Why Bayesian Inference is not the dominant paradigm?

Explanations

- Subjectivity
- Bayesian Leaders
- History
- Computation
- Theory Bias

Why you should be a Bayesian Econometrician

Decision Theory under Uncertainty

Following Marx we must recognize that sometimes it is not enough to understand the world (inference), sometimes we must change it (action). In this regard is where statistical decision theory is fundamental, as long as we follow the advice of Fisher, we should distinguish between inferences about specific events and inferences about theories. The former is basic for applied work where decisions have to be made. Remember,

> *"...Knowledge is useful if it helps to make the best decisions."*
>
> *Marschak (1960, pp 1)*

Why you should be a Bayesian Econometrician

Decision Theory under Uncertainty

Decision theory is concerned about making optimal decisions under uncertainty. For instance, the Minimum Least Squared estimator is a decision rule under uncertainty assuming a Squared Loss function.

Why you should be a Bayesian Econometrician

Decision Theory under Uncertainty

The statistical decision framework must be natural to economist because its origin is a sum zero game (Neumann and Morgenstern, 1944, Wald, 1945) where opponents are econometricians and state of nature. The Bayesian paradigm is particular appealing in decision theory under uncertainty because both approaches use non experimental information and have deep theoretical ties.

Why you should be a Bayesian Econometrician

Statistical Decision Theory

For instance, we do obtain admissible rules using informative priors (Berger, 1993), whereas common frequentist rules, like Minimum Least Squares, are inadmissible under a squared loss function and a parametric space greater than 2 (James, 1961, James and Stein, 1961). This last statement is also true for Bayesian analysis using non-informative distributions.

We should recognize that the risk function introduced by (Wald, 1947) was focused on the long-term performance of a decision rule in a series of repetitions, whereas the Bayesian perspective of this theory is based just on observed outcomes.

Why you should be a Bayesian Econometrician

Why you should be a Bayesian Econometrician

Therefore, from a theoretical standpoint, the main consideration for adopting the Bayesian approach is that it allows us to establish a statistical framework that simultaneously unifies decision theory, statistical inference, and probability theory under a single philosophically and mathematically consistent structure.

References I

Bayarri, M. J. and Berger, J. (2004). The interplay of bayesian and frequentist analysis. *Statistical Science*, 19(1):58–80.

Berger, J. (1993). *Statistical Decision Theory and Bayesian Analysis*. Springer, third edition edition.

Dawid, A. P., Musio, M., and Fienberg, S. E. (2016). From statistical evidence to evidence of causality. *Bayesian Analysis*, 11(3):725–752.

Fisher, R. (1958). *Statistical Methods for Research Workers*. Hafner, New York, 13th edition.

Good, I. J. (1992). The bayes/non bayes compromise: A brief review. *Journal of the American Statistical Association*, 87(419):597–606.

Goodman, S. N. (1999). Toward evidence-based medical statistics. 1: The P value fallacy. *Annals of internal medicine*, 130(12):995–1004.

References II

James, W. (1961). Inadmissibility of of the usual estimator for the mean of a multivariate normal distribution. *Proccedings of the third Berkeley symposium on mathematical statistics and probability*, 1:197–206.

James, W. and Stein, C. (1961). Estimation with quadratic loss. *Proccedings of the fourth Berkeley symposium on mathematical statistics and probability*, 1:361–380.

Jeffreys, H. (1931). *Scientific Inference*. Cambridge University Press, London.

Jeffreys, H. (1961). *Theory of Probability*. Oxford University Press, London.

Judge, G., Griffiths, W., Carter, R., Lutkepohl, H., and Lee, T.-C. (1985). *The theory and practice of econometrics*. John Wiley & Sons, second edition edition.

Kass, R. (2011). Statistical inference: the big picture. *Statistical science*, 26(1):1–9.

References III

Lindley, D. V. (2000). The philosophy of statistics. *The Statistician*, 49(3):293–337.

Lindley, D. V. and Phillips, L. D. (1976). Inference for a Bernoulli process (a Bayesian view). *American Statistician*, 30:112–119.

Marschak, J. (1960). Economic measurements for policy and prediction. In Holt, C. C., Modigliani, F., Muth, J. F., and Simon, H. A., editors, *Planning production, inventories and workforce*. Prentice Hall, Englewood Cliffs. pp. 1–26.

Gelman, A. and Imbens, G. (2013). Why ask why? forward causal inference and reverse causal questions. Technical Report w19614, National Bureau of Economic Research.

Neumann, J. V. and Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton University Press, Princeton.

References IV

Neyman, J. and Pearson, E. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society, Series A*, 231:289–337.

Parmigiani, G. and Inoue, L. (2008). *Decision theory principles and approaches*. John Wiley & Sons.

Raftery, A. (1995). Bayesian model selection in social research. *Sociological Methodology*, 25:111–163.

Siegfried, T. (2014). To make science better, watch out for statistical flaws. Technical report, ScienceNews.

Wald, A. (1945). Generalization of a theorem by v. neumann concerning zero sum two-person games. *Annals of Mathematics*, 46.

Wald, A. (1947). Foundations of a general theory of sequential decision functions. *Econometrica*, 15.

Wasserstein, R. L. and Lazar, N. A. (in press 2016). The ASA's statement on p–values: context, process and purpose. *The American Statistician*.

References V

Ziliak, S. (2008). Guinnessometrics; the Economic Foundation of student's t. *Journal of Economic Perspectives*, 22(4):199–216.