Bayesian Econometrics

Christopher A. Sims
Princeton University
sims@princeton.edu

February 4, 2020

References

Gelman et al. (2014)
MacKay (2003)
Lancaster (2004)

Outline

I. The difference between Bayesian and non-Bayesian inference, and where it matters.

II. Confidence sets and confidence intervals. Why?

III. Bayesian interpretation of frequentist data analysis.

The bare bones of Bayesian inference

- One has a probability model of the observed data Y that depends on unknown parameters θ: $p(Y \mid \theta)$.

- Both Y and θ are random when unknown. But Y will be observed, after which it is no longer unknown and hence no longer "random".

- We have a distribution for θ representing our uncertainty about it before we see Y. It has pdf $\pi(\theta)$.

- They joint distribution is $\pi(\theta)p(Y \mid \theta)$, so the conditional density of $\theta \mid Y$ is

$$\frac{p(Y \mid \theta)\pi(\theta)}{\int p(Y \mid \theta)\pi(\theta)\, d\theta}$$

Bayesian Inference is a Way of Thinking, Not a Basket of "Methods"

- Frequentist inference makes only pre-sample probability assertions.
  - A 95% confidence interval contains the true parameter value with probability .95 only *before* one has seen the data. After the data has been seen, the probability is zero or one.
  - Yet confidence intervals are universally interpreted in practice as guides to *post*-sample uncertainty.
  - They often are reasonable guides, but only because they often are close to posterior probability intervals that would emerge from a Bayesian analysis.

- People want guides to uncertainty as an aid to decision-making. They want to characterize uncertainty *about parameter values*, given the sample that has actually been observed. That it aims to help with this is the distinguishing characteristic of Bayesian inference.

Bayesian quality checks for frequentist inference

- As you know, the complete class theorem suggests that if frequentist inference is not interpretable as Bayesian for some prior, it can usually be improved upon.
- So it worthwhile asking what prior and (possibly implicit) model would support a given frequentist procedure.
- Even if such a prior exists, one can ask whether the prior is reasonable.

Is the difference that Bayesian methods are subjective?

- No.

- The objective aspect of Bayesian inference is the set of rules for transforming an initial distribution into an updated distribution conditional on observations.

- Bayesian thinking makes it clear that for decision-making, pre-sample beliefs are therefore in general important.

- But most of what econometricians do is not decision-making. It is reporting of data-analysis for an audience that is likely to have diverse initial beliefs.

- In such a situation, as was pointed out long ago by Hildreth (1963) and Savage (1977, p.14-15), the task is to present useful information about the shape of the likelihood.

Flat priors, "ignorance"

- Flat priors amount to treating the likelihood, normalized to integrate to one, as the posterior.

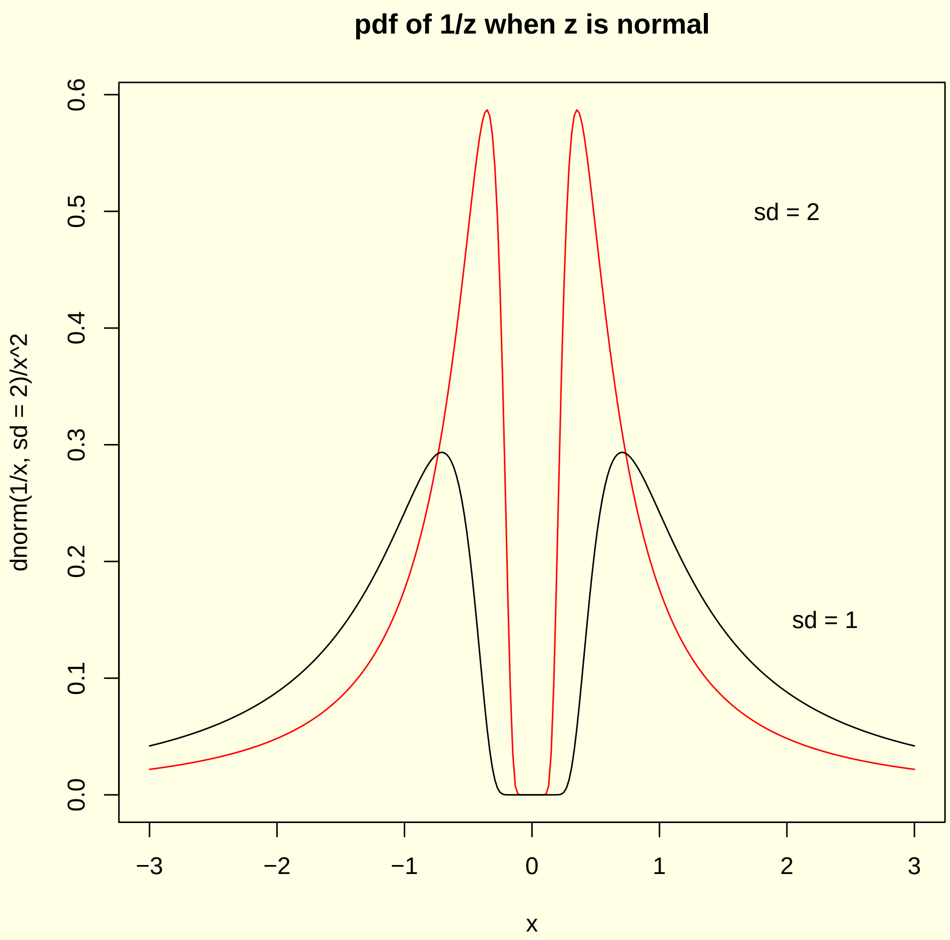
- They do not represent ignorance, or lack of prejudice.

- They do not in general make Bayesian probability statements look like frequentist confidence statements.

- Example: Flat prior on θ as the limit as $\sigma \to \infty$ of a $N(0, \sigma^2)$ prior.

Flat priors, "ignorance"

- Flat priors amount to treating the likelihood, normalized to integrate to one, as the posterior.
- They do not represent ignorance, or lack of prejudice.
- They do not in general make Bayesian probability statements look like frequentist confidence statements.
- Example: Flat prior on θ as the limit as $\sigma \to \infty$ of a $N(0, \sigma^2)$ prior.
- Implies probability on large $|\theta|$ values increasing, therefore prior on $1/\theta \xrightarrow{\mathcal{D}} 0$


pdf of 1/z when z is normal
dnorm(1/x, sd = 2)/x^2
0.0
0.1
0.2
0.3
0.4
0.5
0.6
−3
−2
−1
0
1
2
3
x
sd = 2
sd = 1

Flat priors as an approximation

- Sometimes, the likelihood dominates the prior.
- That is, likelihood $p(Y \mid \theta)$ is very small outside a set S such that the prior $\pi(\theta)$ is nearly constant on S.
- Then a flat prior and $\pi()$ will deliver nearly the same posterior.
- In regular cases, the likelihood does concentrate in smaller and smaller regions as sample size increases, so with any prior expressible as a density function, in large enough samples a flat prior becomes a good approximation.

How to characterize the likelihood

- Present its maximum.
- Present a local approximation to it based on a second-order Taylor expansion of its log. (Standard MLE asymptotics.)
- Plot it, if the dimension is low.
- If the dimension is high, present slices of it, marginalizations of it, and implied expected values of functions of the parameter. The functions you choose might be chosen in the light of possible decision applications.

- The marginalization is often more useful if a simple, transparent prior is used to downweight regions of the parameter space that are widely agreed to be uninteresting.

Confidence set simple example

- $X \sim U(\beta + 1, \beta - 1)$
- β known to lie in (0,2)
- 95% confidence interval for β is $(X - .95, X + .95)$
- Can truncate to (0,2) interval or not — it's still a 95% interval.

Example cont.

- Bayesian with $U(0,2)$ prior has 95% posterior probability ("credibility") interval that is generally a subset of the intersection of the $X \pm 1$ interval with (0,2), with the subset 95% of the width of the interval.

- Frequentist intervals are simply the intersection of $X \pm .95$ with (0,2). They are wider than Bayesian intervals for X in $(0,2)$, but narrower for X values outside that range, and in fact simply vanish for X values outside $(-.95, 2.95)$.

Are narrow confidence intervals good or bad?

- A 95% confidence interval is always a collection of all points that fail to be rejected in a 5% significance level test.

- A completely empty confidence interval, as in example 1 above, is therefore sometimes interpreted as implying “rejection of the model”.

- As in example 1, an empty interval is approached as a continuous limit by very narrow intervals, yet very narrow intervals are usually interpreted as implying very precise inference.

Bettable confidence sets

- Müller and Norets (2016)

- Suppose we allow someone to look at the $(1 - \alpha)$ confidence interval, then make a decision as to whether to make a bet at $(1 - \alpha)/\alpha$ odds that parameter is *not* in the interval.

Bettable confidence sets

- Müller and Norets (2016)

- Suppose we allow someone to look at the $(1-\alpha)$ confidence interval, then make a decision as to whether to make a bet at $(1-\alpha)/\alpha$ odds that parameter is *not* in the interval.

- If this bettor can make money on average, maybe we don't like the CI.

Bettable confidence sets

- Müller and Norets (2016)

- Suppose we allow someone to look at the $(1-\alpha)$ confidence interval, then make a decision as to whether to make a bet at $(1-\alpha)/\alpha$ odds that parameter is *not* in the interval.

- If this bettor can make money on average, maybe we don't like the CI.

- Every non-bettable confidence set contains a $1-\alpha$ credibility set for some prior.

- If the confidence set is non-bettable and “similar” (coverage $1 - \alpha$ for all true parameter values), or if it is non-bettable “from either side”, it is a credibility set under some prior.

While CS's usually do, CI's often don't, exist

- To create a CS, find a test of $H_0 : \theta = \theta_0$ at the α level for each $\theta_0 \in \Theta$. The CS is $\{\theta_0 \mid \theta = \theta_0\}$ not rejected.

- It's intuitively appealing to integrate the likelihood/posterior to obtain a "marginal" distribution for the parameter for which a CI is desired. This is completely correct for a Bayesian analysis. It also happens to work for pure location-shift problems with symmetry in the likelihood (like normal models, or asymptotically regular models). But generally, these integrated likelihoods do not allow derivation of CI's.

- The coverage probability for a random interval for a parameter generally depends on the values of other parameters.

Exploring the parameter space

- In applied work we almost always try many models, or versions of models.

- We may add or subtract explanatory variables, check for serial correlation, check whether nonlinearity is present, check whether there are outliers or indications of non-normality.

- In frequentist inference, the fact that this exploration occurs and affects final displayed results makes the logical foundations of tests and confidence sets dubious. An example is “pre-test bias”.

- In Bayesian inference, this experimentation with model variants can be thought of as exploration of the likelihood surface. So long as the explorations are reported, there is no logical problem in interpreting the resulting characterization of the likelihood shape.

Breaks

- Suppose we have a model specifying Y_t is i.i.d. with a pdf $f(y_t;\mu_t)$, and with μ_t taking on just two values, $\mu_t = \underline{\mu}$ for $t < t^*$, $\mu_t = \bar{\mu}$ for $t \geq t^*$.

- The pdf of the full sample $\{Y_1, \ldots, Y_T\}$ therefore depends on the three parameters, $\underline{\mu}, \bar{\mu}, t^*$.

- If f has a form that is easily integrated over μ_t and we choose a prior $\pi(\underline{\mu}, \bar{\mu})$ that is conjugate to f (meaning it has the same form as the likelihood), then the posterior marginal pdf for t^* under a flat prior on t^* is easy to calculate: for each possible value of t^*, integrate the posterior over $\underline{\mu}, \bar{\mu}$.

- The plot of this integrated likelihood as a function of t^* gives an easily interpreted characterization of uncertainty about the break date.

- Frequentist inference about the break date has no simple interpretation for any observed complications (like multiple local peaks, or narrow peaks) in the global shape of the likelihood. Generally coverage probabilities for t^* depend on the values of μ_1 and μ_2.

Another example where Bayesian and frequentist inference diverges: Time series with possible unit roots.

- This is another “non-regular” problem.
- If we are sure, in an autoregressive model, that the data are stationary, then inference is regular: in large samples the likelihood depends on parameters mainly through location shifts. Therefore the usual parallelism between Bayesian and frequentist conclusions holds.
- But if we believe that possibly the data are non-stationary, the problem is non-regular. Asymptotic distribution theory is different when there is a unit root than in the stationary case. This leads to complicated difficulties in inference for a frequentist.

- Bayesian inference in this case is much more straightforward, and focuses attention on different aspects of the modeling assumptions.

- We'll discuss these issues in more detail later.

Resistance

1. It's subjective.

2. It requires strong assumptions about small sample distributions.

3. Frequentists can come up with intuitively appealing and easily computed estimates, examine their properties. Bayesians always find optimal estimates, which may be hard to understand and compute.

4. The workhorse framework of econometrics, GMM, has no Bayesian interpretation, except asymptotically.

*

References

FERGUSON, T. S. (1967): *Mathematical Statistics: A Decision Theoretic Approach*, New York and London: Academic Press.

GELMAN, A., J. B. CARLIN, H. S. STERN, D. B. DUNSON, A. VEHTARI, AND D. B. RUBIN (2014): *Bayesian Data Analysis*, CRC Press.

HILDRETH, C. (1963): "Bayesian Statisticians and Remote Clients," *Econometrica*, 31, 422–438.

KIM, J.-Y. (2002): "Limited information likelihood and Bayesian analysis," *Journal of Econometrics*, 107, 175–193.

Kwan, Y. K. (1998): "Asymptotic Bayesian analysis based on a limited information estimator," *Journal of Econometrics*, 88, 99–121.

Lancaster, T. (2004): *An Introduction to Modern Bayesian Econometrics*, Blackwell.

MacKay, D. J. C. (2003): *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press.

Müller, U. K. and A. Norets (2016): "Credibility of Confidence Sets in Nonstandard Econometric Problems," *Econometrica*, 84, 2183–2213.

Savage, L. J. (1977): "The Shifting Foundations of Statistics," in *Logic, Laws and Life*, ed. by R. Colodny, University of Pittsburgh Press, 3–18.