Bayesian and Frequentist Approaches

G. Jogesh Babu

Penn State University
http://sites.stat.psu.edu/∼babu

http://astrostatistics.psu.edu

All models are wrong

But some are useful

– George E. P. Box (son-in-law of Sir Ronald Fisher)

Box is well known for his contributions to time series analysis, design of experiments and Bayesian inference

Introduction

- The Bayesian vs. Frequentist debate is largely muted in statistics community.
- The debate often misses some salient features on both sides.
- Both the Bayesian and frequentist ideas have a lot to offer to practioners.
- It is always good statistical practice to analyze the data by several methods and compare.

- Inherently joint Bayesian-frequentist situations.
- Hypothesis testing and Estimation

Bayesian methods are based on Decision theoretic principle.
Actions are dictated by risk management.
Minimizing the expected loss.

Both frequentist and Bayesian methods often lead to the same solutions, when no external information (other than the data and model itself) is to be introduced into the analysis.

Frequentist methods are computationally easy.
No need for numerical integration in most cases.

Interpretations are different.

Parts of the presentation is based on the paper by Bayarri and Berger in Statistical Science, entitled "The Interplay of Bayesian and Frequentist Analysis."

Frequentist Principle

In repeated practical use of a statistical procedure, the long-run average actual accuracy should not be less than (and ideally should equal) the long-run average reported accuracy.

This is actually a joint frequentist-Bayesian principle.

Example (95% classical confidence interval for a normal mean)

The text book statement focuses on fixing the value of the normal mean and imagining repeatedly drawing data from the given model and utilizing the confidence procedure repeatedly on this data.

However, in practice the confidence procedure is used on a series of different problems. Hence, in evaluating the procedure, we should simultaneously be averaging over the differing means and data. Conceptually, it is the combined frequentist-Bayesian average that is practically relevant.

Hypothesis testing

Type I error

This is the error made when a test rejects a true null hypothesis.
A type I error is false positive, in signal detection.
These are philosophically a focus of skepticism and Occam's razor.

Type II error

Error made when a test fails to reject a false null hypothesis.
A type II error may be compared with a so-called false negative.
Such an error occurs if we believe there is no signal when in actuality there is a signal.

Null $\mathrm{H}_0 : \theta \in \Theta_0$ verses alternative $\mathrm{H}_1 : \theta \in \Theta_1$

In classical testing H_0 and H_1 are not treated symmetrically while they are in Bayesian analysis.

- **Bayes factor** B in favor of Θ_0 is given by

$$B = \frac{\text{Posterior odds ratio}}{\text{Prior odds ratio}} = \frac{\alpha_0 \pi_1}{\alpha_1 \pi_0}$$

  - In the simple vs. simple hypothesis testing it is the likelihood ratio.
  - $\alpha_i = P(\Theta_i | x)$, posterior and π_i are prior probabilities, $i = 0, 1$.

- Bayesian chooses H_0 if B is greater than 1.
- Treats both hypotheses symmetrically.
- The procedure does not provide error probabilities.

Rejection region of Bayesian test by decision theoretic approach under binary loss is

$$C = \left\{ x : P(\Theta_1|x) > \frac{K_1}{K_0 + K_1} \right\}$$

Typically similar to classical test.

K_i are cost factors in the loss function.

P-value against H_0

P-value is the probability, when $\theta = \theta_0$ of observing an X "more extreme" than the actual data x.

- p-value is not a posterior probability of a hypothesis.
- p-value of 0.05 typically means that one can be pretty sure that H_0 is wrong.
- Any reasonably fair Bayesian analysis will show that there is at best very weak evidence against H_0.
- Criticism of p-values ignores the power of the test, which has quite a bit of hidden information.
- Bayesian analysis can be questioned because of the choice of the prior. A Bayesian has no recourse but to attempt subjective specification of the feature.

- Neyman-Pearson testing theory reports the same error probability regardless of the size of the test statistic
- Problematic in the view of many statisticians.
- This led Fisher to push p-values.
- p-values do not have a frequentist justification.

Berger, Brown and Wolpert (1994) proposed a solution for testing by defining the Type I and Type II errors conditioned on the observed values of a statistic measuring the strength of evidence in the data.

Conditional Frequentist Testing

Example

Suppose the data X arises from the simple hypotheses

$$H_0 : f = f_0 \ vs.\ H_1 : f = f_1$$

Conditional error probabilities based on a selected statistic S = S(X) are computed as

$$\begin{aligned}\alpha(s) &= P(\text{Type I error}|S = s)\\ &\equiv P_0(\text{reject } H_0|S(X) = s)\\ \beta(s) &= P(\text{Type II error}|S = s)\\ &\equiv P_1(\text{accept } H_0|S(X) = s)\end{aligned}$$

S and associated test utilize p-values to measure the strength of the evidence in the data.

No connection with Bayesianism till this point.

Conditioning is completely allowed (and encouraged) within the frequentist paradigm.
The Bayesian connection arises as it is known that

$$\alpha(s) = \frac{B(x)}{1 + B(x)} \quad \text{and} \quad \beta(s) = \frac{1}{1 + B(x)}$$

- $B(x)$ is the Bayes factor (or likelihood ratio)
- These expressions are precisely the Bayesian posterior probabilities of H_0 and H_1, assuming the hypotheses have equal prior probabilities of $1/2$.
- A conditional frequentist can simply compute the objective Bayesian posterior probabilities of the hypotheses
- Declare that they are the conditional frequentist error probabilities.
- No need to formally derive the conditioning statistic or perform the conditional frequentist computations.

Bayesian Computation

- Gibbs sampling and other Markov chain Monte Carlo methods have become relatively standard to deal with Hierarchical, Multilevel or Mixed Models.
- These methods are commonly known as Bayesian computation methods.
- The popularity of the methods is not necessarily due to their intrinsic virtues, but rather because the Bayesian computation is now much easier than computation via more classical routes.
- On the other hand, any MCMC method relies fundamentally on frequentist reasoning to do the computation. Diagnostics for MCMC convergence are almost universally based on frequentist tools.

Use of frequentist methods in Prior development

- A subjective Bayesian need not worry about frequentist ideas, if a prior distribution accurately reflects prior beliefs.
- However, it is rare to have a (mathematical) prior distribution that accurately reflects all prior beliefs.

Example

Suppose that the only unknown model parameter is a normal mean μ. Complete assessment of the prior distribution for μ involves an infinite number of judgments. Such as specification of the probability that μ belongs to the interval $(-\infty, r]$ for any rational number r.

In practice, only a few assessments are ever made (e.g., choose prior as a Cauchy density, with median and first quartile specified).

Clearly one should worry about the effect of features of the prior that were not elicited.

Consistency

- The simplest frequentist estimation tool that a Bayesian can usefully employ is consistency.
- Bayes estimates are virtually always consistent if the parameter space is finite-dimensional.
- This need not be true if the parameter space is not finite-dimensional or in irregular cases.

Example (Neyman-Scott Problem)

X_{ij} are normally distributed with mean μ_i and variance σ^2,
$i = 1, \ldots, n;\ j = 1, 2$
Until relatively recently, the most commonly used objective prior was the Jeffreys-rule prior. The resulting Bayesian estimate of σ^2 approximates to half the value, leading to an inconsistent estimate.

The Jeffreys-rule prior is often inappropriate in highdimensional settings, yet it can be difficult or impossible to assess this problem within the Bayesian paradigm itself.

Estimation

- In the standard parametric estimation problems, objective Bayesian and frequentist methods often give similar or even identical answers.
- For the standard normal linear model, frequentist estimates and confidence intervals coincide exactly with the standard objective Bayesian estimates and credible intervals.
- More generally, this occurs in the presence of "invariance structure"
- It is still possible to achieve near-agreement between frequentist and Bayesian estimation procedure, in more complicated problems.

Conclusions

- Both Bayesian and frequentist methodology are here to stay.
- There are many areas of frequentist methodology that should be replaced by (existing) Bayesian methodology that provides superior answers.
- The verdict is still out on those Bayesian methodologies that have been exposed as having potentially serious frequentist problems.
- Philosophical unification of the Bayesian and frequentist positions is not likely, nor desirable, since each illuminates a different aspect of statistical inference.
- These two do different things. Analyze using both and decide.
- Understand the assumptions behind the theory and the mathematical conclusions.
- Properly interpret the results.