Comments About Objective Bayesian Analysis

Jim Berger

Insitute of Statistics and Decision Sciences
Duke University

Outline

- Comments for Non-Bayesians
- Comments for Subjective Bayesians
- Comments for Objective Bayesians

Comments for Non-Bayesians

- There **is** a well-developed 200 year-old Objective Bayesian paradigm. (Bayes $\neq$ subjective.)
  - It is as objective as anything can be in statistics.
  - It can be employed whenever one wants to 'appear objective' or feels unable to find a subjective prior.
  - There are standard objective priors and techniques available for a wide array of problems (Jeffreys priors, reference priors, ...).

- Objective Bayesian analysis is often the easiest way to obtain good frequentist procedures (including incorporation of needed conditioning).
  - For a testing example - see my Fisher lecture.
  - Here is a confidence set example arising in medical diagnosis:
- Objective Bayesian analysis is great for difficult problems.
  - Here is a multiple comparisons example.

A Medical Diagnosis Example (Mossman)

The Medical Problem:

- Within a population, $p_0 = Pr(\text{Disease } D)$.
- A diagnosic test results in either a Positive (P) or Negative (N) reading.
- $p_1 = Pr(P \,|\, \text{patient has } D)$.
- $p_2 = Pr(P \,|\, \text{patient does not have } D)$.

It follows that

$$\theta = Pr(D|P) = \frac{p_0 p_1}{p_0 p_1 + (1 - p_0) p_2}.$$

The Statistical Problem: The p_i are unknown. Based on (independent) data $X_i \sim \text{Binomial}(n_i, p_i)$, find a $100(1-\alpha)\%$ confidence set for θ that has good conditional and frequentist properties.

Suggested Solution: Use the Jeffreys-rule priors $\pi(p_i) \propto p_i^{-1/2}(1-p_i)^{-1/2}$ and compute the $100(1-\alpha)\%$ equal-tailed posterior credible set.

Ease of Use: a simulation algorithm

- draw random p_i from the Beta$(x_i + \frac{1}{2}, n_i - x_i + \frac{1}{2})$ distributions, $i = 0, 1, 2$;
- compute the associated $\theta = p_0 p_1 / [p_0 p_1 + (1 - p_0) p_2]$;
- repeat this process $10,000$ times;
- the $\frac{\alpha}{2}\%$ upper and lower percentiles of these generated θ form the desired confidence limits.

(p_0, p_1, p_2)	O-Bayes	Log Odds	Gart-Nam	Delta
$(\frac{1}{4}, \frac{3}{4}, \frac{1}{4})$	2.86,2.71	1.53,1.55	2.77,2.57	2.68,2.45
$(\frac{1}{10}, \frac{9}{10}, \frac{1}{10})$	2.23,2.47	0.17,0.03	1.58,2.14	0.83,0.41
$(\frac{1}{2}, \frac{9}{10}, \frac{1}{10})$	2.81,2.40	0.04,4.40	2.40,2.12	1.25,1.91

Table 1: The probability that the nominal 95% interval misses the true θ on the left and on the right, for the indicated parameter values and when $n_0 = n_1 = n_2 = 20$.

Multiple Comparisons Example

- Suppose $X_i \sim N(\mu_i, \sigma^2)$, $i = 1, \ldots, m$ are observed, with σ^2 known, and it is desired to determine which μ_i are nonzero.
- Let p denote the unknown probability that a given μ_i is zero.
- Assume that the nonzero μ_i follow a $N(0, V)$ distribution, with V unknown.
- Assign p the uniform prior on $(0, 1)$ and V the prior density $\pi(V) = \sigma^2/(\sigma^2 + V)^2$.

- Then the posterior probability that μ_j is zero can be shown to be

$$p_j = \frac{\int_0^1 \int_0^1 p \prod_{i \neq j} \left(p + (1-p)\sqrt{1-w}\, e^{w{X_i}^2/(2\sigma^2)} \right) dp dw}{\int_0^1 \int_0^1 \prod_{i=1}^m \left(p + (1-p)\sqrt{1-w}\, e^{w{X_i}^2/(2\sigma^2)} \right) dp dw}.$$

- $(p_1, p_2, \ldots, p_m)$ can be computed numerically if m is moderate. For large m, it is most efficient to do the computation via importance sampling, with a common importance sample for all p_j. (Even a uniform importance density usually works fine.)

- Note that, in the Bayesian approach, the 'penalty' for adding additional (mean zero) comparisons is that p will concentrate on smaller values. This is automatic, and one need not do any adjustments (e.g. Bonferoni).

Example: Draw ten $N(0, 4^2)$ 'signal' observations:
-8.48, -5.43, -4.81, -2.64, -2.40, 3.32, 4.07, 4.81, 5.81, 6.24
Generate $n = 10, 50, 500, \text{and}\, 5000$ $N(0, 1)$ 'noise' observations.

	Central seven 'signal' observations							
n	-5.4	-4.8	-2.6	-2.4	3.3	4.1	4.81	noise> .6
10	1	1	.94	.89	.99	1	1	1
50	1	1	.71	.59	.94	1	1	0
500	1	1	.26	.17	.67	.96	1	2
5000	1.0	.98	.03	.02	.16	.67	.98	1

Table 2: The posterior probabilities of the 'signal' means being nonzero, as the number, n, of noise observations increases.

Comments for Subjective Bayesians

- An objective Bayesian analysis can provide a reference for the effect of the prior in a subjective analysis.
- It is rare that subjective elicitation can be thoroughly done for all unknowns in a problem, so that some utilization of objective priors is almost inevitable.
- An objective Bayesian analysis can be run initially, to assess if subjective priors are even needed. (Perhaps the data will 'swamp the prior'.)

- Through study of objective priors, one can obtain insight into possibly bad behavior of standard (e.g., conjugate) subjective priors.

Example: Covariance matrices

Consider the multivariate normal model $\boldsymbol{X} \sim N_k(\boldsymbol{\theta}, \boldsymbol{\Sigma})$.

A commonly used subjective prior for $\boldsymbol{\Sigma}$ is the inverse Wishart prior (for subjectively specified a and b)

$$\pi(\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-a/2} \exp(-\frac{1}{2}\mathrm{tr}[b\boldsymbol{\Sigma}^{-1}]).$$

Write $\boldsymbol{\Sigma} = \boldsymbol{H}^t\boldsymbol{D}\boldsymbol{H}$, where $\boldsymbol{H}$ is orthogonal and $\boldsymbol{D}$ is diagonal with entries $d_1 > d_2 > \cdots > d_k$. Change of variables yields

$$\pi(\boldsymbol{\Sigma})\, d\boldsymbol{\Sigma} = \frac{1}{|\boldsymbol{D}|^{a/2}} e^{(-\frac{1}{2}\mathrm{tr}[b\boldsymbol{D}^{-1}])} \prod_{i<j}(d_i - d_j) \cdot I_{[d_1 > \cdots > d_k]}\, d\boldsymbol{D} d\boldsymbol{H}.$$

Thus, unexpectedly, one is using a prior distribution that very strongly forces the eigenvalues of the covariance matrix **apart**, contrary to typical prior beliefs.

Objective Bayes: The reference prior (Yang and Berger, 1993) is

$$\pi^R(\boldsymbol{D}, \boldsymbol{H}) = |\boldsymbol{D}|^{-1} d\boldsymbol{D} d\boldsymbol{H},$$

which has no such problem (and gives much better statistical inferences).

The Point: Proper priors that are commonly used by subjectivists can have hidden undesirable features; objective Bayesian tools can expose these features and allow for development of better subjective priors.

Comments for Objective Bayesians

- Use of improper objective priors is better than use of 'vague proper priors.'
  - Improprieties in the posterior can be detected. (In contrast, a vague proper prior will appear okay in such a situation, but will actually yield an arbitrary answer.)
  - If model selection or mixtures are involved, vague proper priors are typically terrible.
  - (Improper) objective priors (e.g., *reference priors*) often have excellent properties.

- Model selection must be treated very carefully.
- Use of data dependent priors needs formal study, as a part of objective Bayesian analysis.
- Posterior propriety is a serious concern, but is mitigated if reference priors are used.

A Spatial Example

$Z(\mathbf{s})$ is a random process observed at $\mathbf{s}_i \in \Re^k, i = 1, \ldots, n,$

$$Z(\mathbf{s}) = \sum_{j=1}^{p} \beta_j f_j(\mathbf{s}) + \varepsilon(\mathbf{s})$$

where $f_j(\mathbf{s})$ are known regression functions, β_j are unknown coefficients and ε is a Gaussian process with zero mean and

$$cov(\varepsilon(\mathbf{s}), \varepsilon(\mathbf{u})) = cov(Z(\mathbf{s}), Z(\mathbf{u})) = \sigma^2 K_\vartheta(\|\mathbf{s} - \mathbf{u}\|)$$

Example: $K_\vartheta(\|\mathbf{s} - \mathbf{u}\|) = e^{-\|\mathbf{s}-\mathbf{u}\|/\vartheta}, \;\; \vartheta > 0.$

Common Objective Priors

Natural objective priors are of the form

$$\pi(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\vartheta}) = \frac{\pi(\boldsymbol{\vartheta})}{\sigma^2}.$$

How should $\pi(\boldsymbol{\vartheta})$ be chosen?

- Subjective information is rarely available.
- Uniform prior: $\pi(\boldsymbol{\vartheta}) = 1$.

  Problem: The resulting posterior is improper.
- Other common objective priors can also fail to give a proper posterior.

The Reference Prior

$$\pi^R(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\vartheta}) = \frac{1}{\sigma^2} [\sum_{i=1}^{n-p} (w_i(\boldsymbol{\vartheta}) - \overline{w}(\boldsymbol{\vartheta}))^2]^{\frac{1}{2}},$$

where

- $w_1, \ldots, w_{n-p}$ are eigenvalues of W_ϑ, with $\overline{w}$ their average;
- $W_\vartheta = \frac{\partial}{\partial \vartheta} \Sigma_\vartheta^{-1} \Sigma_\vartheta^{-1} (I - X(X' \Sigma_\vartheta^{-1} X)^{-1} X' \Sigma_\vartheta^{-1})$;
- X $(n \times p)$ is defined by $X_{ij} = f_j(\mathbf{s}_i)$;
- Σ_ϑ $(n \times n)$ is defined by $\Sigma_{\vartheta,ij} = K_\vartheta(\|\mathbf{s}_i - \mathbf{s}_j\|)$.

Properties and comments:

- The reference prior always yields a proper posterior for suitable $K_\vartheta(\|\mathbf{s} - \mathbf{u}\|)$.
- We are currently studying the frequentist performance of posterior credible sets from this prior.

Concluding Comment

Objective Bayesian analysis does not have as attractive an axiomatic foundation as does subjective Bayesian analysis, but it has long been an indispensable part of Bayesian analysis and, I suspect, its role will only be growing.

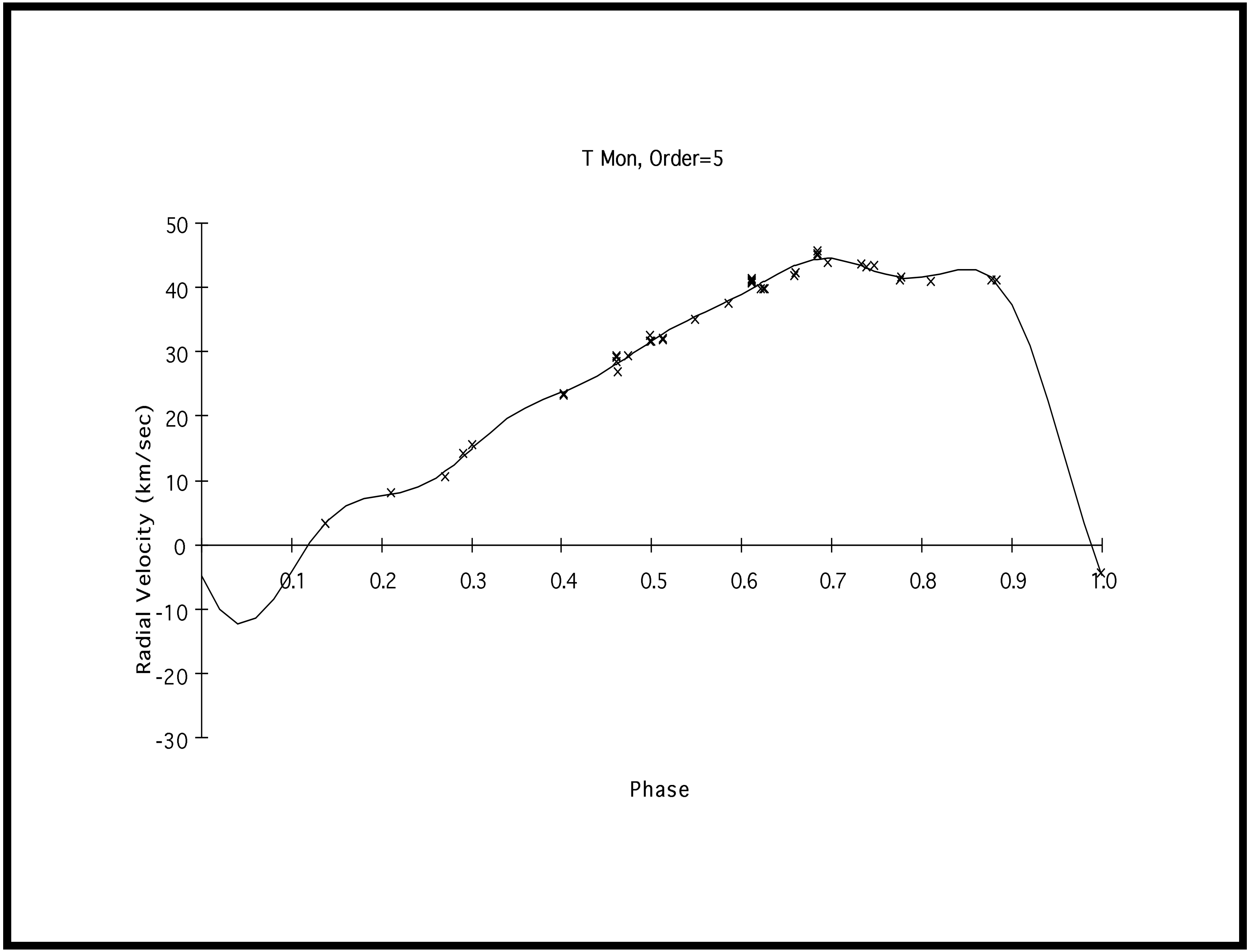
Postscript: A Typical Application – Inference for Cepheid Distances (with Jefferys and Müller)

- A Cepheid star pulsates, regularly varying its luminosity (light output) and size.

- From the Doppler shift as the star surface pulsates, one can compute surface velocities at certain phases of the star's period.

- From the luminosity and 'color' of the star, one can learn about the angular radius of the star (the angle from Earth to opposite edges of the star).

- Combining, allows estimation of s, a star's distance.



T Mon, Order=5
Radial Velocity (km/sec)
50
40
30
20
10
0
-10
-20
-30
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Phase

Choice of Cepheid Prior Distributions

- The orders, (M, N), of the trigonometric polynomials used to fit the velocity curves, are given a uniform distribution up to some cut-off (e.g., $(10, 10)$).

- τ_u, τ_v, τ_c, which adjust the measurement standard errors, are given the standard objective priors for 'scale parameters,' namely the Jeffreys-rule priors $p(\tau_u) = \frac{1}{\tau_u}$, $p(\tau_v) = \frac{1}{\tau_v}$, and $p(\tau_c) = \frac{1}{\tau_c}$.

- The mean velocity and luminosity, u_0 and v_0, are 'location parameters' and so can be assigned the standard objective priors $p(u_0) = 1$ and $p(v_0) = 1$.

- The angular diameter ϕ_0 and the unknown phase shift $\Delta\phi$ are also assigned the objective priors $p(\Delta\phi) = 1$ and $p(\phi_0) = 1$. It is unclear if these are 'optimal' objective priors but the choice was found to have neglible impact on the answers.

- The Fourier coefficients arising from the curve fitting (which is done by model selection from among trigonmetric polynomials) occur in linear models, so the Zellner-Siow conventional model selection priors were utilized.

The prior for distance s of the star should account for

- *Lutz-Kelker bias*: a uniform spatial distribution of Cepheid stars would yield a prior proportional to s^2.
- The distribution of Cepheids is flattened wrt the galactic plane; we use an exponential distribution.
- So, we use $p(s) \propto s^2 \exp\left(-|s \sin\beta|/z_0\right)$,
  - β being the known galactic latitude of the star (its angle above the galactic plane),
  - z_0 being the 'scale height,' assigned a uniform prior over the range $z_0 = 97 \pm 7$ parsecs.