The Present and Future of Bayesian Multivariate Analysis*

by

James O. Berger
Purdue University

Technical Report #92-33C

July, 1992

*This research was supported by National Science Foundation grants DMS–8717799 and DMS–8923071 at Purdue University.

THE PRESENT AND FUTURE OF BAYESIAN MULTIVARIATE ANALYSIS

by

James O. Berger
Department of Statistics
Purdue University
1399 Mathematical Sciences Bldg.
West Lafayette, IN 47907-1399

Abstract

The current activity in multivariate Bayesian analysis is reviewed, with an eye towards those areas likely to be especially prominent in the future. Extra attention is devoted (somewhat arbitrarily) to the areas of computation, development of prior distributions, robustness and sensitivity, model selection and testing, and interaction and comparison of Bayesian and classical methods.

1. Introduction

1.1. Goals

This is too big a topic. In recent years Bayesian methods have proliferated through virtually all theoretical and applied domains of multivariate analysis. No article could review even a significant fraction of this recent work, much less predict where the field is going.

The purpose of this article is thus considerably more modest. First, an attempt will be made to describe why this proliferation of Bayesian methods is occurring. The chief factor, a freedom from the need to employ "standard" models – a freedom arising out of powerful computational developments – will be highlighted in Section 2.

The second goal of the paper is to provide a brief summary or catalogue of current subjects in Bayesian multivariate analysis, along with recent references that allow an accessing of the literature. The subjects are organized as Development of Prior Distributions (Section 3), Bayesian Robustness, Sensitivity, and Diagnostics (Section 4), Model Choice and Testing (Section 5), Other Methodological Developments (Section 6), and Areas of Classical and Bayesian Interaction and Comparison (Section 7). This last section includes discussions of design, sequential analysis, asymptotics, consistency, and nonparametrics.

The final goal of the paper is to provide some speculation about the future, or at least projection of current trends. These speculations or projections will be intermixed with the various subject summaries.

1.2. Notation

Notation will be kept basic. The (entire) data will be denoted by X, assumed to have a conditional density $f(x|\theta)$, given an unknown parameter $\theta \in \Theta$, the parameter space. A prior density for θ will be denoted by $\pi(\theta)$, with the posterior density being

$$\pi(\theta|x) = f(x|\theta)\pi(\theta)/m(x);$$

here $m(x) = \int f(x|\theta)\pi(\theta)d\theta$ is the marginal or predictive density. The focus of Bayesian analysis is typically computation of various posterior expectations

$$E^*[g(\theta)] = \int g(\theta)\pi(\theta|x)d\theta;$$

the "$*$" as a superscript to E will always denote posterior expectation, and if $g(\theta)$ is a vector or matrix, the expectation is to be taken componentwise. Common choices of g include

$g(\theta) = \theta$, since then $E^*[\theta] =$ *posterior mean*;

$g(\theta) = (\theta - E^*[\theta])(\theta - E^*[\theta])^t$, since then $E^*[g(\theta)] =$ *posterior covariance matrix*;

$g(\theta) = 1_C(\theta) = \begin{cases} 1 & \text{if } \theta \in C \\ 0 & \text{otherwise,} \end{cases}$ since then $E^*[g(\theta)] =$ *posterior probability of* C.

2. The Upsurge in Use of Bayesian Methods and Computation

2.1. Introduction

Foundational and intuitive arguments in support of the Bayesian approach are legion (cf., Berger, 1985, and Sections 5 and 7.2 here), and certainly have played a strong role in popularizing the Bayesian approach. Recent work has significantly clarified and strengthened many of these arguments, but most of the arguments themselves are old.

The "new" development that has greatly contributed to the upsurge in use of Bayesian methods is the development of computational tools that allow analysis of highly complex and nonstandard models. Indeed, for complicated models, Bayesian analysis has now arguably become the *simplest* (and often only possible) method of analysis. We illustrate this point by presenting an example in Section 2.2, followed by a brief review in Section 2.3 of recent computational advances. Section 2.4 discusses software issues.

2.2. An Illustration

The best way to illustrate the power of current Bayesian methods is to present an example. The following example is based on a problem studied in Andrews, Berger, and Smith (1992), though certain features of the problem are here simplified or enriched for purposes of exposition.

The problem was to determine the effect of certain automotive technologies, such as fuel injection, on the fuel efficiency of automobiles. After certain transformations of the data and variables, the base model became

$$Y_{ijk} = \beta^t X_{(ijk)} + \alpha^t X^*_{(ij)} + \varepsilon_{ijk};$$

here

Y_{ijk} = *log*(fuel efficiency in MPG) of a vehicle;
$X_{(ijk)}$ = a vector of the vehicle characteristics, including indicators of presence of technologies of interest;
β = a vector of unknown "fixed effects";
$X^*_{(ij)}$ = an indicator vector specifying the vehicle model and manufacturer;
α = a vector of unknown "random effects";
i = $1, \ldots, I$, denoting the manufacturer;
j = $1, \ldots, J_i$, denoting the vehicle model for manufacturer i;
k = $1, \ldots, N_{ij}$, denoting a particular vehicle of model j from manufacturer i.

Error distributions deemed possible are $\varepsilon_{ijk} \overset{\text{i.i.d.}}{\sim}$ Normal $(0, \sigma^2)$, with σ^2 unknown, and $\varepsilon_{ijk} \overset{\text{i.i.d.}}{\sim} t$-distribution with median 0, unknown scale σ, and unknown degrees of freedom ν. The model is complicated by being unbalanced (the J_i and the N_{ij} are highly variable) and there is considerable missing data.

For the fixed effects, β, certain sign and order restrictions are known; indeed, it is known that

$$\beta \in \Omega = \{\beta: \beta_{10} > 0, \beta_{15} > 0, \beta_{18} > 0, \beta_4 \leq \beta_5 \leq \beta_6\}.$$

The car model effects, α, are modeled as

$$\alpha_{ij} \overset{\text{i.i.d.}}{\sim} \text{Normal } (\mu_i, V_i), \quad j = 1, \ldots, J_i,$$

where μ_i and V_i are the overall mean and variance for manufacturer i. It is believed, however, that there is a time trend to the overall manufacturer means; this is modeled by the AR (1) process

$$\mu_i(t) = \rho_i \mu_i(t-1) + \gamma_{it}, \quad i = 1, \ldots, I,$$

where t denotes the year of vehicle manufacture, and the γ_{it} are i.i.d. Normal $(0, \gamma)$ errors. Finally, the unknown ρ_i and V_i are also modeled as random effects from the population of all manufacturers, with the ρ_i being i.i.d. Beta (λ, τ) and the V_i being i.i.d. Inverse Gamma (ξ, η).

The desired goal of the analysis is to predict fuel efficiencies, Y, but at uncertain (i.e., random) future vehicle configurations (X, X^*). Estimates, standard errors, and confidence (credible) sets for Y are desired, as well as tests for certain of the $\beta_i = 0$.

There are numerous features of this problem that would virtually preclude the possibility of a classical analysis. Having both fixed and random effects in an unbalanced situation, even with normal errors, is by itself enough to almost require a Bayesian analysis (to produce reasonable standard errors and credible sets). Adding the complications of t-errors, restrictions on the parameters, time series structures for some of the random effects, and the desire to predict Y at random future X creates a problem of almost unapproachable complexity from a classical perspective.

Solving this problem from the Bayesian perspective is comparatively straightforward. One must first place a prior distribution on all unknown parameters that do not already have a "random effects" distribution. The simplest possibility is the constant density on these parameters (restricted to Ω, of course), i.e.,

$$\pi(\beta, \gamma, \lambda, \tau, \xi, \eta, \sigma^2, \nu) = 1_\Omega.$$

(Choosing improper "noninformative" densities such as this can be justified from a number of perspectives; see Section 3.3 for further discussion.) Using the techniques discussed in the next section, one can then compute posterior means and variances (and other desired posterior expectations) for any of the unknown parameters, or for future Y.

There are, of course, the usual variety of concerns with the above analysis, centering around issues of sensitivity to, and plausibility of, assumptions (including choice of the prior density). Also, the computation required is far from trivial. The point to be stressed,

however, is that with the Bayesian approach the statistician has complete freedom to utilize whatever models, structures, or restrictions seem reasonable for a particular problem, while maintaining the capability to compute answers. There is no need to force the problem into a standard mold by oversimplification.

2.3. Bayesian Computation

Although other goals are possible, most Bayesian computation is focused on calculation of posterior expectations $E^*[g(\theta)]$, for various g. Until recently, most Bayesian methodology had been developed for models and priors (conjugate) for which analytic computation of $E^*[g(\theta)]$ was possible. Here, we will instead focus on the exciting vistas arising from utilization of new numerical methods of computation.

2.3.1. Traditional Numerical Methods

The "traditional" numerical methods for computing $E^*[g(\theta)]$ are Numerical Integration, Laplace Approximation, and Monte-Carlo Importance Sampling. Brief introductions to these methods can be found in Berger (1985). Here we say only a few words, to place the methods in context and provide references.

A successful general approach to Numerical Integration in Bayesian problems, using adaptive quadrature methods, was developed in Naylor and Smith (1982). This was very effective in moderate (e.g., 10) dimensional problems.

Extension of the Laplace Approximation method of analytically approximating $E^*[g(\theta)]$, leading to a reasonably accurate general technique, was carried out in Tierney, Kass, and Kadane (1989). The chief limitations of the method are the need for analytic derivatives, the need to redo parts of the analysis for each different $g(\theta)$, and the lack of an estimate of the error of the approximation. For many problems, however, the technique is remarkably successful.

Monte Carlo Importance Sampling (see Geweke (1988, 1989) and Wolpert (1991) for discussion) has been the most commonly used method of computing $E^*[g(\theta)]$. The method can work in very large dimensions, and carries with it a fairly reliable accuracy measure. Although one of the oldest computational devices, it is still one of the best, being nearly "optimal" in many problems. It does require determination of a good "importance function," however, and this can be a difficult task. Current research continues to address the problem of choosing a good importance function; for instance, Oh and Berger (1992) develop a method of selecting an importance function for a multimodal posterior.

2.3.2. Markov Chain Simulation Techniques

The newest techniques to be extensively utilized for numerical Bayesian computations are Markov Chain Simulation Techniques, including the popular Gibbs Sampling. (Certain of these techniques are actually quite old – see, e.g., Hastings (1970); it is their application and

adaption to Bayesian problems that is new.) A brief generic description of these methods is as follows:

Step 1. Select a "suitable" Markov chain on Θ, with $p(\cdot,\cdot)$ being the transition probability density (i.e., $p(\theta,\theta^*)$ gives the transition density for movement of the chain from θ to θ^*). Here "suitable" means primarily that $\pi(\theta|x)$ is a stationary distribution of the Markov chain, which can be assured in a number of ways.

Step 2. Starting at a point $\theta^{(0)} \in \Theta$, generate a sequence of points $\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(m)}$ from the chain.

Step 3. Then, for large m, $\theta^{(m)}$ is (approximately) distributed as $\pi(\theta|x)$ and

$$\frac{1}{m}\sum_{i=1}^{m} g(\theta^{(i)}) \cong E^*[g(\theta)].$$

The main strengths of Markov chain methods for computing $E^*[g(\theta)]$ are:

(i) Many different g can simultaneously be handled via Step 3, once the sequence $\theta^{(1)}, \ldots, \theta^{(m)}$ has been generated.

(ii) Programming tends to be comparatively simple.

(iii) Methods of assessing convergence and accuracy exist and/or are being developed.

The main weaknesses of the Markov chain methods are:

(i) They can be quite slow. It is not uncommon in complicated problems to need m to be in the hundreds of thousands, requiring millions of random variable generations if the dimension of θ is appreciable.

(ii) One can be misled into prematurely judging that convergence has obtained.

The more common Markov chain methods, corresponding to different choices of $p(\cdot,\cdot)$, will briefly be discussed.

Metropolis-Hastings Algorithm: One generates a new θ^* based on a "probing" distribution, and then moves to the new θ^* or stays at the old θ according to certain "accept-reject" probabilities. See Hastings (1970).

Gibbs Sampling: The Markov chain moves from $\theta^{(i)}$ to $\theta^{(i+1)}$ one coordinate at a time (or one group of coordinates at a time), the transition density being the conditional posterior density of the coordinate(s) being moved given the other coordinates. This is a particularly attractive procedure in many Bayesian scenarios, such as analysis of hierarchical models, because the conditional posterior density of one parameter given the others is often relatively simple (or can be made so with the introduction of auxiliary variables). Extensive discussion and illustration of Gibbs sampling can be found in Geman and Geman (1984), Gelfand and Smith (1990), Gelman and Rubin (1992), Raftery (1992), and Smith and Gelfand (1992).

Hit and Run Sampling: The idea here is roughly that one moves from $\theta^{(i)}$ to $\theta^{(i+1)}$ by choosing a random direction and then moving in that direction according to the appropriate conditional posterior distribution. This method is particularly useful when Θ is a sharply constrained parameter space. Extensive discussion and illustration can be found in Belisle, Romeijn, and Smith (1990), and Chen and Schmeiser (1992).

Hybrid Methods: Complex problems will typically require a mixture of the above (and other) methods. Here is an example, from Müller (1991), the purpose of which is to do Gibbs sampling when the posterior conditionals (e.g., $\pi(\theta_i|x$, other $\theta_k)$) are not "nice":

Step 1. Each step of the Markov chain will either

- generate $\theta_j^{(i)}$ from $\pi(\theta_j|x$, other $\theta_k^{(i)})$ if the conditional posterior is "nice" or
- generate $\theta_j^{(i)}$ by employing one or several steps of the Metropolis-Hastings algorithm if the conditional is not nice.

Step 2. For the probing function in the Metropolis-Hastings algorithm, use the relevant conditional distribution from a global multivariate normal (or t) importance function, as typically developed in Monte Carlo importance sampling.

Step 3. Adaptively update the importance function periodically, using estimated posterior means and covariance matrices.

Other discussions or instances of use of hybrid methods include Tanner and Wong (1987), Geyer (1991), Gilks and Wild (1992), Tanner (1991), Tierney (1991), Geweke (1992), Neal (1992), and Berger and Chen (1992).

2.4. Software Existence and Development

Availability of general user-friendly Bayesian software would rapidly advance use of Bayesian methods. A number of software packages do exist, and are very useful for particular scenarios. An example is BATS (cf., West and Harrison (1989)), which is designed for Bayesian time series analysis. A listing and description of pre-1990 Bayesian software can be found in Goel (1988) and Press (1989).

Four recent software developments are BAIES, a Bayesian expert system (see Cowell, 1992); [B/D], an "expectation based" subjective Bayesian system (see Goldstein (1988), Wooff (1992)); BUGS, designed to analyze general hierarchical models via Gibbs sampling (see Thomas, et. al., 1992); and XLISP-STAT, a general system complete with excellent interactive and graphics facilities (see Tierney, 1990).

Two of the major strengths of the Bayesian approach create certain difficulties in developing generic software. One is the extreme flexibility of Bayesian analysis, with virtually any constructed model being amenable to analysis. Classical packages need contend with only a few well-defined models or scenarios for which a classical procedure has been determined. Another strength of Bayesian analysis is the possibility of extensive utilization of subjective

prior information, and Bayesians tend to feel that software should include an elaborate expert system for prior elicitation. This is hard, in part because much remains to be done empirically to determine optimal ways to elicit priors. Note that such an expert system is not, by any means, a strict need for Bayesian software; it is possible to base a system on use of noninformative priors.

3. Development of Prior Distributions

3.1. Introduction

Selection of the prior distribution, $\pi(\theta)$, is inherently a subject of importance to Bayesians. The centrality of prior selection is not as great as most non-Bayesians believe, however; many practical Bayesians operate mainly with the constant prior $\pi(\theta) = 1$, feeling it to be much more important to concentrate their time on modeling. Indeed, this attitude has caused a great deal of attention to be paid to development of "noninformative priors," a topic that will be discussed in Section 3.3. First, however, development of subjective priors will be discussed.

3.2. Subjective Multivariate Prior Elicitation

Eliciting multivariate priors is extremely challenging, especially when coordinates of θ are not independent. Complicating the matter further is the fact that, in many situations, elicitation needs to be done on observables, X^*, rather than model parameters, θ. One then works backwards to the model parameters, working from specifications concerning $m(x^*)$ to implied specifications for $\pi(\theta)$. See Dickey, Dawid, and Kadane (1986) and Goldstein (1988) for discussion.

As an example of multivariate assessment methodology, consider the following situation encountered by Andrews, Berger, and Smith (1992) concerning the automotive fuel efficiency example of Section 2.2. It was mentioned there that the goal of the analysis was to predict Y (fuel efficiency) at future vehicle configurations (X, X^*). To be a bit more precise, it was desired to predict what would happen if a technology were added to a vehicle; this would correspond to one of the X_i switching from a 0 to a 1. The difficulty is that adding a technology alters other vehicle characteristics X_j; for instance, adding overhead cams to an engine would alter the horsepower. Since the goal of the analysis was to predict the *fuel efficiency* benefit of adding the technology, not benefits in, say, horsepower, it was necessary to determine how the vehicle could be reconfigured (i.e., the other X_j altered) to restore the vehicle to previous performance levels in all categories except fuel efficiency.

Data concerning such reconfiguration was essentially unavailable, and so it was necessary to elicit the reconfiguration of variables from automotive engineers. Each technology affected up to four other variables, so it was necessary to elicit up to four-dimensional distributions. The elicitation proceeded as follows:

(i) The 10th, 50th, and 90th percentiles of the marginals of relevant X_j were elicited.

(ii) Correlations among the X_j were obtained, elicited as either "none" (i.e., 0), "moderate" (i.e., -0.5 or 0.5), "high" (i.e., -0.9 or 0.9) or "complete" (i.e., -1 or 1). Note that we were able to elicit this rather high level of detail because the engineers involved were familiar with statistics.

The next step in the construction of the prior was to model these inputs via a probability distribution. The major difficulty was that several of the elicited quantiles could only be modeled by highly skewed distributions yet, for later purposes, a computationally simple prior was required.

A very useful distribution for this type of situation is the multivariate split-normal distribution. Each of the marginal distributions is a split normal distribution

$$g_i(x_i|\mu_i,\tau_i,\rho_i) = \begin{cases} \frac{1}{\sqrt{2\pi\tau_i}}\exp\{-\frac{1}{2\tau_i}(x_i-\mu_i)^2\} & \text{if } x_i \le \mu_i \\ \frac{1}{\sqrt{2\pi\rho_i}}\exp\{-\frac{1}{2\rho_i}(x_i-\mu_i)^2\} & \text{if } x_i > \mu_i. \end{cases}$$

Here, μ_i is the specified median of X_i, while τ_i and ρ_i can be chosen so that X_i has the specified 10th and 90th percentiles, respectively.

The multivariate split-normal distribution can now easily be constructed by breaking up the space into orthants, with center at the specified marginal medians. In each orthant, one constructs the multivariate normal distribution made from the appropriate "halves" of the split-normal marginals, together with the elicited correlations. The resulting distribution does not necessarily have exactly the correct correlations, but usually they are close enough (and adjustments can be made to induce the correct correlations if desired). See Andrews, Berger, and Smith (1992) for further details.

Computations involving the multivariate split-normal distribution are typically almost as easy as computations for the multivariate normal distribution. Combined with its capability of handling skewness and correlation, this makes the multivariate split-normal distribution an excellent tool for modeling multivariate elicitations.

3.3. Development of Noninformative Priors

As mentioned earlier, it is quite common to simply use $\pi(\theta) = 1$ as the prior distribution in multivariate Bayesian analysis. In the great majority of problems this provides perfectly satisfactory results, but sometimes difficulties are encountered such as nonintegrability of the posterior $\pi(\theta|x)$ (see Ye and Berger (1991), Ibrahim and Laud (1991), and Liseo (1992)). This has led to a long search for a "better" noninformative prior.

The first serious alternative to $\pi(\theta) = 1$ was developed by Jeffreys (1961), and is the now famous Jeffreys prior

$$\pi_J(\theta) = \sqrt{\det\, I(\theta)},$$

where $I(\theta)$ is the expected Fisher information matrix and "det" stands for determinant. Not only does π_J virtually always yield an integrable posterior, but it is also invariant in the sense that it transforms properly under reparameterization. (Note that $\pi(\theta) = 1$ does not;

one cannot, say, be simultaneously uniform in θ and θ^2.) It has come to be recognized, however, that the Jeffreys prior can have serious inadequacies in multivariate problems. (Even Jeffreys was aware of the potential problem.) Hence there is extensive activity in pursuing two alternatives.

The first is based on an asymptotic frequentist argument. The idea is to consider $100(1-\alpha)\%$ credible sets, derived from a given prior, as frequentist confidence sets, and to choose that prior for which the (asymptotic) frequentist confidence is $(1-\alpha)$, up to second order. Tibshirani (1989) and Ghosh and Mukerjee (1992) develop this approach in the multivariate setting. The approach is highly attractive theoretically, but unfortunately is very hard to implement. Indeed, its application is currently limited to the two-dimensional case.

The most successful alternative approach to development of multivariate noninformative priors has been the reference prior approach, reviewed in Berger and Bernardo (1992). This approach can actually be thought of as an extension of the Jeffreys method, the idea being to write $\pi(\theta)$ as $\pi(\theta_1)\pi(\theta_2|\theta_1)\ldots\pi(\theta_p|\theta_1,\ldots,\theta_{p-1})$, and apply a modification of the Jeffreys method to the sequence of conditional problems so defined. This method has successfully handled all "counterexamples" to the Jeffreys method, and seems to produce priors with excellent asymptotic frequentist properties. Here is one of the simplest examples.

Example 1. Berger and Bernardo (1989) consider the problem of inference for $\xi = \theta_1\theta_2$ (note that the reference prior typically depends on what is defined to be the quantity of interest) based on $X_1 \sim \mathcal{N}(\theta_1, 1)$ and (independently) $X_2 \sim \mathcal{N}(\theta_2, 1)$. The reference prior in this example turns out to be $\pi_R(\theta_1, \theta_2) = \sqrt{\theta_1^2 + \theta_2^2}$, in contrast to the Jeffreys prior $\pi_J(\theta_1, \theta_2) = 1$. In Berger and Bernardo (1989), the inferences obtained from use of these priors are compared in a variety of ways, with those from π_R being shown to be clearly superior.

3.4. Partial Information Priors

There is obviously a middle ground between completely subjective and completely noninformative priors. One can have partial prior information (e.g., knowledge of some moments of the prior), yet desire to be otherwise noninformative.

There is a huge and rapidly growing literature on dealing with this situation by entropy methods: the idea is to choose the maximum entropy prior subject to the given constraints. Applications exist to physics, astronomy, image processing, and numerous other domains. Discussion and references can be found in Jaynes (1983), Fougere (1990), and Caselton, Kan, and Zidek (1991).

4. Bayesian Robustness, Sensitivity, and Diagnostics

Statistics is increasingly acknowledging the importance of investigating sensitivity to assumptions. Bayesian statistics is no exception, with robustness being one of the most rapidly growing areas.

Bayesian Diagnostics are currently being developed for a variety of purposes. Most diagnostics are based on utilization of the predictive density of the data or some subset thereof, since

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

can be thought of as a "likelihood" for both the model f and prior π. Recent references include Box (1985), Carlin and Polson (1991), Guttman (1991), Geisser (1992), and Kass and Slate (1992).

An increasingly popular Bayesian approach to robustness can be termed Constructive Robustness. The idea is to choose models and priors that are inherently robust in ways that are deemed to be important. One common example is to use t-distributions, instead of normal, for either the likelihood or the prior whenever it is desired to protect against outliers or the "unexpected." It is well known, for instance, that modeling i.i.d. data as being $\mathcal{T}_\nu(\mu, \sigma^2)$, instead of $\mathcal{N}(\mu, \sigma^2)$, will yield Bayesian estimates of μ for which outliers have vanishing influence. Note that, with the new computational methods such as Gibbs sampling, generic use of t-distributions, instead of normal, is quite feasible. Recent references to constructive robustness include Angers and Berger (1991) and Geweke (1992).

Sensitivity and the related Global Robustness are concerned with directly determining how changes in assumptions change the answer. The most common approach is to simply try a few plausible models and priors, and see if the answer changes much. Note the great advantage here of the Bayesian being able to produce the answer for virtually any model or prior that is of interest.

A more formal approach to sensitivity, that is currently under development, is based on functional differentiation. Since typical Bayesian inferences can be written as posterior expectations,

$$E^*[g(\theta)] = \frac{\int g(\theta)f(x|\theta)\pi(\theta)d\theta}{\int f(x|\theta)\pi(\theta)d\theta} \equiv \Psi(f, \pi),$$

sensitivity to f and π, at an elicited model f_0 and prior π_0, can be investigated by looking at functional derivatives of Ψ at f_0 and/or π_0. Such derivatives are actually directional, and this can even be turned into an elicitation tool: finding the "direction" in which the derivative is largest (in absolute value) might indicate a particularly important direction in which to concentrate elicitation efforts. Study of sensitivity through functional differentiation can be found in Diaconis and Freedman (1986), Srinivasan and Truszczynska (1990), and Ruggeri and Wasserman (1991).

In the global robustness approach to sensitivity, one selects classes of (plausible) models and/or priors, and then computes the range of the Bayesian action or inference of interest as the model and/or prior vary over all elements of the classes. If this range is small, one can confidently assert that the conclusion is robust. If not, further refinement of the classes is required (through additional subjective input) or more data is needed. This is not merely a convenient way to investigate robustness; it can be argued (cf., Walley (1991)) to be the axiomatically valid foundation for statistics, assuming the axioms include the (reasonable) assumption that infinitely fine subjective discrimination between alternatives is not possible.

Recent multivariate references concerning this approach include Leamer (1978), Berger and Berliner (1986), Polasek and Pötzelberger (1988), Lavine (1989), Berger (1990), Bose (1990), Lavine, Wasserman, and Wolpert (1991), Moreno and Cano (1991), Bayarri and Berger (1992), and Sivaganesan and Berger (1992).

Global Bayesian robustness has also been used to define a quantitative Occam's razor (cf., Berger and Jeffreys, 1992), and to demonstrate the serious conflict between Bayesian and classical testing of precise hypotheses (see also Section 5). Here is a multivariate example of the latter, taken from Delampady (1989).

Example 2. Suppose X is p-variate normal with mean $\theta = (\theta_1, \ldots, \theta_p)$ and identity covariance matrix. It is desired to test

$$H_0: \theta = \theta^0 \quad \text{versus} \quad H_1: \theta \neq \theta^0,$$

where θ^0 is a specified vector. The classical P-value of observed data, x, is

$$\alpha = P(\chi_p^2 \geq |x - \theta^0|^2),$$

where χ_p^2 is a chi-squared random variable with p degrees of freedom.

A Bayesian would typically measure the evidence against H_0 by the Bayes factor

$$B = f(x|\theta^0) / \int_{\{\theta \neq \theta_0\}} f(x|\theta) g(\theta) d\theta,$$

where g is the (conditional) prior density on the alternative. (The Likelihood school of statistics calls B the weighted likelihood ratio, and g the weight function.)

It is of considerable interest to find the lower bound on B over a broad class of reasonable, objective g. A class of g that is attractive is

$$\mathcal{G} = \{g(\theta) = h(|\theta - \theta^0|): h(\cdot) \text{ is nonincreasing.}\}$$

This class arises from imposing the reasonable constraints that $g(\theta)$ be spherically symmetric about θ^0 (equivalent to the classical reduction to the statistic $|X - \theta^0|$) and that values of θ far from θ^0 be no more plausible than closer values.

In Delampady (1989), it is shown that

$$\underline{B} = \inf_{g \in \mathcal{G}} B = \frac{\exp\{-\frac{1}{2}|x - \theta^0|^2\}}{\sup_k \frac{1}{V(k)} \int_{|\theta - \theta^0| \leq k} \exp\{-\frac{1}{2}|x - \theta|^2\} d\theta},$$

where $V(k)$ is the volume of a ball of radius k. Table 1 gives some selected values of $\underline{B}$, for various p and for x corresponding to certain P-values, α. (Note that the upper bound on B over all $g \in \mathcal{G}$ is infinity, and so is not of interest here.)

Table 1. Lower bounds, $\underline{B}$, on Bayes factors, for various dimensions, p, and P-values, α.

	p								
α	1	2	3	4	5	10	15	20	40
.10	.6437	.5699	.5396	.5232	.5131	.4908	.4826	.4782	.4706
.05	.4092	.3481	.3259	.3141	.3072	.2927	.2875	.2844	.2793
.01	.1227	.0978	.0902	.0850	.0824	.0777	.0752	.0743	.0730
.001	.0182	.0143	.0119	.0114	.0099	.0094	.0093	.0093	.0092

Note the dramatic discrepancy between α and the lower bounds. When $p = 2$ and $\alpha = .01$, for instance, $\underline{B} = .0978$ is almost ten times larger than α. To a Bayesian, $\underline{B} = .0978$ would indicate that the data favors H_1 over H_0 by *at most* a factor of 10 to 1 ("at most" because $\underline{B}$ was a *lower* bound over all plausible g), so that common interpretations of the strength of evidence provided by $\alpha = .01$ are highly misleading.

5. Model Choice and Testing

One of the main areas of conflict between classical and Bayesian statistics is testing of precise hypotheses (as indicated in Example 2 of the previous section) and related issues of model choice. Generally, Bayesian methods are much less likely to reject a more precise hypothesis or model in favor of one that is less precise. Being an area of major disagreement, there is a considerable recent literature on the topic. Discussion and references can be found in Smith and Spiegelhalter (1982), Berger and Delampady (1987), Mitchell and Beauchamp (1988), Delampady (1989), Delampady and Berger (1990), George and McCulloch (1991), Berger and Mortera (1992), and Gelfand, Dey, and Chang (1992).

In Bayesian hypothesis testing and model selection there is a major distinction between comparisons involving hypotheses or models of differing dimensions and those of equal dimension. For instance, there is a fundamental difference in testing $H_0: \theta = 0$ versus $H_1: \theta > 0$ as opposed to testing $H_0: \theta \leq 0$ versus $H_1: \theta \geq 0$. Not only can answers vary by orders of magnitude, but the methodology needed is very different. For testing $H_0: \theta \leq 0$ versus $H_1: \theta > 0$, it is possible to use a noninformative prior (e.g., $\pi(\theta) = 1$), but for testing $H_0: \theta = 0$ versus $H_1: \theta > 0$ this cannot be done; a proper prior must be utilized on $(0, \infty)$. (Use of the Bayes factor can avoid the need for specification of the prior probability that $\theta = 0$.) In general, a proper prior must be utilized on the "extra" parameters in the higher dimensional hypothesis or model.

How is this proper prior to be chosen? The ideal answer, of course, is to choose it subjectively. For a variety of reasons, however, it is also desirable to have available an "automatic" choice. (Since the prior must be proper, calling it "noninformative" would be just too much of a stretch.) Jeffreys (1961) so argued, and even suggested "automatic" choices for a variety of simple problems. No general procedure for generating these automatic priors existed, however. Smith and Spiegelhalter (1982) did propose a reasonable general

procedure for nested models.

An interesting new approach holds the promise of providing an automatic Bayes factor for any hypothesis testing or model comparison problem. It is called the *intrinsic Bayes factor* (IBF), and is developed in Berger and Pericchi (1992). It utilizes only standard noninformative priors, yet rather remarkably seems to produce answers that are very similar to those of Jeffreys for the problems he studies.

The IBF is easy to describe: suppose it is of interest to compare the evidence in the data, x, for two models (or hypotheses)

$$M_1\colon f_1(x|\theta_1) \quad \text{and} \quad M_2\colon f_2(x|\theta_2),$$

where M_2 is the "more complicated" model. Let $\pi_1(\theta_1)$ and $\pi_2(\theta_2)$ be noninformative priors for M_1 and M_2, respectively. Finally, let $x_{(i)}$, $i = 1, \ldots, L$, be "minimal" subsets of the entire data x, minimal in the sense that $\pi_1(\theta_1|x_{(i)})$ and $\pi_2(\theta_2|x_{(i)})$ are both proper, with subsets of $x_{(i)}$ not yielding proper posteriors. (This is actually only the recommended procedure when the data arises as independent observations from a model. Also, not all minimal subsets are really needed.) Then the intrinsic Bayes factor of M_2 to M_1 is given by

$$\begin{aligned} IBF &= \frac{1}{L}\sum_{i=1}^{L} \frac{\int f_2(x|\theta_2, x_{(i)})\pi(\theta_2|x_{(i)})d\theta_2}{\int f_1(x|\theta_1, x_{(i)})\pi_1(\theta_1|x_{(i)})d\theta_1} \\ &= \frac{m_2(x)}{m_1(x)} \cdot \left(\frac{1}{L}\sum_{i=1}^{L} \frac{m_1(x_{(i)})}{m_2(x_{(i)})}\right). \end{aligned}$$

The first expression for the IBF explains the intuition behind it: one chooses a minimal "training" sample $x_{(i)}$ to obtain proper priors $\pi_1(\theta_1|x_{(i)})$ and $\pi_2(\theta_2|x_{(i)})$, and then uses these proper priors with the rest of the data to compute the Bayes factor. Averaging over all possible training samples completes the process. (If there are more than two models or hypotheses being considered, an additional geometric averaging over IBFs is employed; see Berger and Pericchi (1992).)

Among the interesting properties of this procedure are:

(i) It is superior to the BIC model selection criterion, in the sense that BIC is only "first order" while this is "second order."

(ii) For large L, the IBF seems to closely mimic Jeffreys automatic Bayes factor in the cases he considered.

6. Other Methodological Developments

In this section, a variety of important areas of current and future Bayesian research are mentioned, with a few recent references given. Any of these areas would be deserving of extensive discussion if space allowed.

Several of the exciting new domains of Bayesian analysis involve new probability structures for modelling complex systems. Probabilistic Graphical Structures, that allow Bayesian updating to proceed locally within the structure (often with iterations through the structure), are being extensively developed (cf., Lauritzen and Spiegelhalter (1988), Spiegelhalter and Cowell (1991), and Dawid (1992)). Bayesian Image Processing also employs such structures, and has great promise because of the ability to use the prior distribution on images to tune the image processor to handle the desired type of image. One can create a processor to look for lines, boundaries, or certain shapes by choice of prior distributions concentrating near these shapes. A few of the many references to this work are Geman and Geman (1984), Basag (1986), Geman (1988), Mardia, Kent, and Walder (1991), and Johnson and Bowsher (1992).

Two other "new" probability structures receiving considerable attention in the Bayesian community are Neural Nets (cf., Mackay (1992), and Neal (1992)) and Influence Diagrams (cf., Barlow, 1989). The latter have become a key way of structuring and analyzing complex Bayesian inference and decision problems by clarifying, through graphical means, the correct order in which to carry out the calculations.

Finally, in the domain of "new" structures, the hot area of Chaotic Systems should be mentioned. See Berliner (1992) for Bayesian developments in this area.

Traditional areas of methodology have also seen rapid Bayesian advances. Linear, Mixed, and Hierarchical Models are receiving enormous attention, partly because of the new computational methods such as Gibbs sampling, and partly because they are natural Bayesian domains; complex mixed and hierarchical models, especially, are very difficult to handle in a non-Bayesian way (cf., Section 2.2). The huge list of references in this area includes Lindley and Smith (1972), Box and Tiao (1973), Broemeling (1985), Dawid (1988), Fong and Berger (1992), Schervish (1992), van der Merwe and van der Merwe (1992), Hsu and Leonard (1992), and Ghosh (1992a). Related areas of development are Small Area Estimation (cf., Rubin (1987), and Datta and Ghosh (1991)), and Generalized Linear Models (cf., Albert (1988), Zeger and Karim (1991), Stephens and Dellaportas (1992), and Allenby and Lenk (1992)).

Time Series and Dynamic Models (or time-varying coefficient models) have long been a natural for Bayesian analysis, as definitively demonstrated in West and Harrison (1989). Filtering is another such area, with even the classic Kalman filter being essentially Bayesian (cf., Meinhold and Singpurwalla, 1983). See Bretthorst (1988) for related Bayesian work in Spectrum Analysis.

There are two very active areas of Bayesian research involving combination of evidence. The first is Bayesian Meta-Analysis, the combining of evidence from different experiments. The chief difficulty of meta-analysis is that the different experiments typically involve at least slightly different populations and/or treatments. At the extreme, one may even desire to combine evidence from human experiments and animal experiments (cf., DuMouchel and Harris, 1983) or from laboratory experiments and field experiments (cf., Wolpert and Warren-Hicks, 1992). The Bayesian approach is ideally suited to meta-analysis, since it allows for a variety of ways to account for differences between the populations and/or treatments of the

experiments. For discussion and references, see Morris and Normand (1992).

The second active area of work in combination of evidence is Combination of Expert Opinion. Again, the Bayesian approach to this enterprise is "natural," since it shows how to compensate for dependence among experts and for having experts of differing quality. Discussion and references can be found in Genest and Zidek (1986) and West and Crosse (1992).

Biostatistics has seen considerable Bayesian development, much of which is surveyed in Breslow (1990). Methods of Longitudinal Analysis, discussed therein, also apply widely to other areas of social statistics. Calibration is another area in biostatistics that also has much wider applicability; see Brown and Mäkeläinen (1992) for recent Bayesian developments.

Econometrics is today a major source of Bayesian methodology and of Bayesian applications. A few references are Zellner (1971, 1984, 1988), Geweke (1989, 1992), and Phillips (1991).

Several developed statistical methodologies in other fields are based on utilization of mixed or hierarchical models. Typically, the methodologies in these other fields were developed by means of empirical Bayes types of arguments. Today, fully Bayesian analysis of these problems is possible and gives greater flexibility as well as much better estimates of accuracy. Bayesian versions of BLUPs in animal breeding can be found in Gianola and Fernando (1986) and Robinson (1991); Bayesian Kriging in geophysics is discussed in Omre (1987), Myers, Alli, and Edward (1990), Sølna (1992), and Lehn (1992); Bayesian Credibility Theory in insurance is studied in Jewell (1988) and Makov and Smith (1992); and Bayesian hierarchical models in forestry can be found in Green and Strawderman (1992).

Many other fields also have growing Bayesian components of methodology. Examples are forensic science (cf., Berry, 1991), astronomy (cf., Loredo, 1992), anthropology (cf., Buck, Litton, and Stephens, 1992), auditing (cf., van Batenburg, O'Hagan, and Veenstra, 1992), and quality assurance (cf., Irony, Pereira, and Barlow, 1992).

7. Areas of Classical-Bayesian Interaction and Comparison

7.1. Introduction

Debates on classical versus Bayesian statistics have gone on for well over 100 years (although in the early debates it was Bayesian analysis – then called inverse probability – that was the "classical" method). That Bayesian analysis has a major role to play in the future of statistics is no longer in real debate. The powerful methodological advantages of Bayesian analysis mean it will be extensively used, regardless of philosophical or logical issues.

The focus of the debate in the literature has shifted more to discussion of the roles of classical ideas within the Bayesian structure. (Of course, many classical statisticians refuse to participate in the debate, and simply ignore Bayesian analysis.) Roles have been suggested for classical concepts such as consistency (see Section 7.3), admissibility (in the development of "good" noninformative priors), minimaxity and randomization (to increase Bayesian robustness), likelihood and asymptotic methods (which are often an approximation

to Bayesian methods), and sequential and design concepts (before seeing the data, even a Bayesian must average over X). Among the many works discussing some, or all, of these ideas are Morris (1983), Rubin (1984), Berger (1985), Rao (1987), Berger and Robert (1990), Robinson (1991), Green and Strawderman (1992), and Robert (1992).

I view this direction of the current debate as healthy. The case for developing statistical procedures via the Bayesian route is, I feel, compelling, but classical perspectives can be quite useful in fine-tuning the process. The remainder of the section presents a rather arbitrary medly of topics and thoughts on classical – Bayesian interaction and comparison.

7.2. Design and Sequential Analysis

Design and sequential design are naturally combined Bayesian-frequentist enterprises. One has yet to see all or part of the data, so frequentist averages over X are required. At the same time, the optimal design is often highly dependent on the unknown θ, and there is frequently no choice but to invoke prior information about θ in choosing the design. This was historically done rather informally, but today and in the future it will be increasingly common to treat design in a formal Bayesian fashion: elicit a prior (even a crude prior will typically do), compute the overall frequentist Bayes risk for each possible design, and choose the optimal design. Discussion and references can be found in Chaloner (1984), DasGupta and Studden (1991), Pilz (1991), Currin, Mitchell, Morris, and Ylvisaker (1991), Gupta and Miescke (1991), and Verdinelli (1992).

There is one potential difference between Bayesian and classical design that particularly manifests itself in sequential design. This arises from the potential difference in experimental goals. The Bayesian's goal will often be a conclusion such as "the posterior probability of error is 0.05," while the frequentist's goal might be to use a procedure with frequentist error probabilities of 0.05. Different designs might well be optimal for these different goals.

Sequential analysis can also involve profound differences in actual inferences for Bayesians and frequentists, primarily because the Bayesian ignores the stopping rule while the frequentist heavily utilizes it. This has led to considerable controversy in sequential settings such as clinical trials, controversy compounded by issues such as randomization and ethics. See Berry, Wolff, and Sacks (1992) for discussion and references.

The future will hold increasing discussion of these matters, with Bayesian methods becoming ever more prominent in sequential clinical trials. Ethics and logic will both force this change. Interestingly, new logical arguments seem to be arising for the Bayesian position from the growing "conditional frequentist" school of research. Here is an example, from Berger and Wolpert (1992).

Example 3. Suppose $X_1, X_2, \ldots$ are i.i.d. $\mathcal{N}(\theta, 1)$, and that it is desired to test $H_0: \theta = -1$ versus $H_1: \theta = 1$. If the hypotheses have equal prior probability, the Bayesian inference, after stopping experimentation at sample size N, will be to compute the posterior probability of H_0,

$$P(H_0|x_1, \ldots, x_N) = \frac{1}{1 + \exp\{2N\overline{x}_N\}},$$

choose the hypothesis with larger posterior probability (clearly $P(H_1|x_1,\ldots,x_N) = 1 - P(H_0|x_1,\ldots,x_N)$), and report the posterior probability of the rejected hypothesis as the error probability.

The surprise here is that, if the stopping rule is symmetric and one looks at the problem from the conditional frequentist perspective, a result of Brown (1978) shows that the optimal inference is precisely the Bayesian inference, and it has complete (conditional) frequentist justification. But this says that even frequentists should ignore the reason for stopping at N (at least, if the stopping rule is symmetric).

Another interesting feature of this example is that the classical sequential analysis is the SPRT, and the exact error probabilities of the SPRT are somewhat difficult to determine. But the SPRT is suboptimal (to a conditional frequentist), while the optimal $P(H_0|x_1,\ldots,x_N)$ is trivial to use!

7.3. Asymptotics, Consistency, and Nonparametrics

Until recently, the interest in asymptotics and consistency among Bayesians was minimal because, for finite dimensional problems, Bayes rules are virtually always consistent and asymptotically optimal. As Bayesian analysis has moved into nonparametric and infinite parametric domains, however, it has been discovered that consistency is no longer automatic and asymptotics can be strange. See Diaconis and Freedman (1986) and Ghosh (1992). Clarke and Junker (1992) and Sweeting (1992) have other interesting asymptotic results.

Example 4. J.K. Ghosh (personal communication, 1992) has studied an interesting variant of the Neyman-Scott problem. Suppose we observe (all independently) $X_{ij} \sim \mathcal{N}(\mu_i, \sigma^2)$, $i = 1,\ldots,p$ and $j = 1,2$. It is desired to estimate σ^2. A simple consistent estimator, as $p \to \infty$, is $\hat{\sigma}^2 = \sum_{i=1}^{p} (x_{i1} - x_{i2})^2/(2p)$.

Now suppose a Bayesian were to proceed by choosing independent subjective priors for all parameters $\{\sigma^2, \mu_1, \mu_2, \ldots, \mu_p\}$. Then, for "almost all" sequences $\{\mu_1, \mu_2, \ldots\}$, the Bayes estimator of σ^2 seems to be inconsistent. ("Almost all" here is in a topological sense, not probabilistic; the Bayes estimator is consistent for almost all sequences $\{\mu_1, \mu_2, \ldots\}$ in probability under the prior, but the set of such sequences becomes vanishingly small. Conditions on the priors and sequences are needed for the proof of inconsistency, but the result is probably true generally.)

Determining the extent to which such possible inconsistencies are a practical concern for Bayesians will be an important task for the future. At the very least, these concerns should significantly influence the types of priors chosen for these problems (cf., Ghosh (1992b), in regards to the above example).

On the methodological side, there have been a number of significant advances in Bayesian nonparametrics. The vast majority of these have involved use of the Dirichlet process prior on the space of all probability distributions. Recent references include Kuo (1986), Lo and Weng (1989), and the review article Ferguson, Phadia, and Tiwari (1992). An exciting example

of the potential of this approach is Doss (1991), which constructs a prior on nonparametric distributions, but one that concentrates on distributions near a parametric family.

Dirichlet process priors have a number of potentially unappealing features, such as the fact that they give probability one to the set of discrete probability measures. Hence there has been considerable effort expended on priors which are supported on continuous densities, such as Gaussian process priors. An example of such a prior, for the space of continuous densities, $f(t)$, on $[0,T]$, is to let

$$f(t) = \exp\{X(t)\} / \int_0^T \exp\{X(t)\} dt,$$

where $X(t)$ is the sample path of a Gaussian process. These and other such priors are studied in Leonard (1978), Lenk (1988), and Weerahandi and Zidek (1988). Computations with such priors are more difficult than with Dirichlet process priors, but the recent new computational tools should enhance the utilization of these alternative nonparametric priors.

In regards to Gaussian process priors, the Bayesian interpretation of smoothing splines should also be mentioned. Smoothing splines can be developed as Bayesian function estimates for certain Gaussian process priors on derivatives of functions. This interpretation has been important in deriving accuracy estimates for smoothing splines (utilizing the associated posterior covariance function). See Nychka (1988), Kohn and Ansley (1988), and Wahba (1990). There is considerable promise in further exploiting this relationship for higher dimensional smoothing splines, especially if structural assumptions on the function are made (e.g., $f(x_1, \ldots, x_p) = \sum_{i=1}^{p} f_i(x_i)$).

References

[1] J.H. Albert (1988). Computational methods using a Bayesian hierarchical generalized linear model. *J. Amer. Statist. Assoc.* **83**, 1037–1044.

[2] G. Allenby and P. Lenk (1992). Modeling household purchase behavior. Technical Report, School of Business Administration, University of Michigan.

[3] R. Andrews, J. Berger, and M. Smith (1992). Bayesian estimation of fuel economy potential due to technology improvements. To appear in *Case Studies in Bayesian Statistics*, J. Hodges, R. Kass, and N. Singpurwalla (Eds.).

[4] J.F. Angers and J. Berger (1991). Robust hierarchical Bayes estimation of exchangeable means. *Canadian J. Statist.* **19**, 39–56.

[5] R.E. Barlow (1989). Influence diagrams. *Encyclopedia of Statist. Sci.*, S. Kotz and N.L. Johnson (Eds.). Wiley, New York.

[6] M.J. Bayarri and J. Berger (1992). Applications and limitations of robust Bayesian bounds and type II maximum likelihood. Technical Report, Department of Statistics, Purdue University.

[7] C. Belisle, H.E. Romeijn, and R. Smith (1990). Hit-and-run algorithms for generating multivariate distributions. Technical Report 90–18, Department of Industrial and Operations Engineering, University of Michigan.

[8] J. Berger (1985). *Statistical Decision Theory and Bayesian Analysis* (2nd edition). Springer-Verlag, N.Y.

[9] J. Berger (1990). Robust Bayesian analysis: sensitivity to the prior. *J. Statist. Plann. Inf.* **25**, 303–328.

[10] J. Berger, and M. Berliner (1986). Robust Bayes and empirical Bayes analysis with ε-contaminated priors. *Ann. Statist.* **14**, 461–486.

[11] J. Berger and J. Bernardo (1989). Estimating a product of means: Bayesian analysis with reference priors. *J. Amer. Statist. Assoc.* **84**, 200–207.

[12] J. Berger and J. Bernardo (1992). On the development of the reference prior method. In *Bayesian Statistics* **4**, J. Bernardo, J. Berger, A. Dawid, and A.F.M. Smith (Eds.). Oxford University Press.

[13] J. Berger and M.H. Chen (1992). Determining retirement patterns: prediction for a multinomial distribution with constrained parameter space. To appear in *The Statistician*, in the Proceedings volume for the Third International Conference on Practical Bayesian Statistics.

[14] J. Berger and M. Delampady (1987). Testing precise hypotheses (with discussion). *Statist. Science* **2**, 317–352.

[15] J. Berger and W. Jefferys (1992). The application of robust Bayesian analysis to hypothesis testing and Occam's razor. *J. Ital. Statist. Society* **1**, 17–32.

[16] J. Berger and J. Mortera (1992). Robust Bayesian hypothesis testing in the presence of nuisance parameters. Technical Report #91–69C, Department of Statistics, Purdue University.

[17] J. Berger and L. Pericchi (1992). The intrinsic Bayes factor: hypothesis testing and model selection. Technical Report, Department of Statistics, Purdue University.

[18] J. Berger and C. Robert (1990). Subjective hierarchical Bayes estimation of a multivariate normal mean: on the frequentist interface. *Ann. Statist.* **18**, 617–651.

[19] J. Berger and R. Wolpert (1992). A unified conditional frequentist and Bayesian test for fixed and sequential simple hypothesis testing. Technical Report, Department of Statistics, Purdue University.

[20] L.M. Berliner (1992). Statistics, probability, and chaos (with Discussion). *Statist. Sci.* **7**, 69–122.

[21] D. Berry (1991). Inferences using DNA profiling in forensic identification and paternity cases. *Statist. Sci.* **6**, 175–205.

[22] D. Berry, M.C. Wolff and D. Sack (1992). Public health decision making: a sequential vaccine trial. In *Bayesian Statistics* **4**, J. Bernardo, J. Berger, A. Dawid, and A.F.M. Smith (Eds.). Oxford University Press.

[23] J. Besag (1986). Statistical analysis of dirty pictures (with discussion). *J. Royal Statist. Soc.* (Ser. B) **48**, 259–302.

[24] S. Bose (1990). Bayesian robustness with shape-constrained priors and mixture priors. Ph.D. Thesis, Purdue University.

[25] G. Box (1985). The *Collected Works of George E.P. Box*, G.C. Tiao (Ed.). Wadsworth, Belmont CA.

[26] G. Box and G. Tiao (1973). *Bayesian Inference in Statistical Analysis.* Addison-Wesley, Reading.

[27] N. Breslow (1990). Biostatistics and Bayes (with discussion). *Statist. Sci.* **5**, 269–298.

[28] G.L. Bretthorst (1988). *Spectrum Analysis and Parameter Estimation.* Springer-Verlag, New York.

[29] L.D. Broemeling (1985). *Bayesian Analysis of Linear Models.* Marcel Dekker, New York.

[30] L. Brown (1978). A contribution to Kiefer's theory of conditional confidence procedures. *Ann. Statist.* **6**, 59–71.

[31] P.J. Brown and T. Mäkeläinen (1992). Regression, sequenced measurements and coherent calibration. In *Bayesian Statistics* **4**, J. Bernardo, J. Berger, A. Dawid, and A.F.M. Smith (Eds.). Oxford University Press.

[32] C.E. Buck, C.D. Litton, and D.A. Stephens (1992). Detecting changes in the shape of prehistoric corbelled tombs. To appear in *The Statistician*, in the Proceedings volume for the Third International Conference on Practical Bayesian Statistics.

[33] B.P. Carlin and N.G. Polson (1991). An expected utility approach to influence diagnostics. *J. Amer. Statist. Assoc.* **86**, 1013–1021.

[34] W.F. Caselton, L. Kan, and J.V. Zidek (1991). Quality data network designs based on entropy. In *Statistics in the Environmental and Earth Sciences*, P. Guttorp and A. Walden (Eds.). Griffin, London.

[35] K. Chaloner (1984). Optimal Bayesian experimental design for linear models. *Ann. Statist.* **12**, 283–300.

[36] M.H. Chen and B. Schmeiser (1992). Performance of the Gibbs, hit-and-run, and Metropolis samplers. Technical Report SMS92–2, School of Industrial Engineering, Purdue University.

[37] B. Clarke and B. Junker (1992). Inference from the product of marginals of a dependent likelihood. Technical Report, Department of Statistics, Purdue University.

[38] R.G. Cowell (1992). BAIES: A probabilistic expert system shell with qualitative and quantitative learning. In *Bayesian Statistics* **4**, J. Bernardo, J. Berger, A. Dawid, and A.F.M. Smith (Eds.). Oxford University Press.

[39] C. Currin, T. Mitchell, M. Morris, and D. Ylvisaker (1991). Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments. *J. Amer. Statist. Assoc.* **86**, 953–963.

[40] A. DasGupta and W.J. Studden (1991). Robust Bayesian analysis of experimental designs in normal linear models. *Ann. Statist.* **19**, 1244–1256.

[41] Dawid, A.P. (1988). Symmetry models and hypotheses for structured data layouts (with Discussion). *J. Roy. Statist. Soc. B* **50**, 1–34.

[42] Dawid, A.P. (1992). Applications of a general propogation algorithm for probabilistic expert systems. To appear in *Statistics and Computing 2.*

[43] G.S. Datta and M. Ghosh (1991). Bayes prediction in linear models: applications to small area estimation. *Ann. Statist.* **19**, 1748–1770.

[44] M. Delampady (1989). Lower bounds on Bayes factors for invariant testing situations. *J. Multivariate Anal.* **28**, 227–246.

[45] M. Delampady and J. Berger (1990). Lower bounds on Bayes factors for multinomial and chi-squared tests of fit. *Ann. Statist.* **18**, 1295–1316.

[46] P. Diaconis and D. Freedman (1986). On the consistency of Bayes estimates. *Ann. Statist.* **14**, 1–26.

[47] J.M. Dickey, A.P. Dawid and J.B. Kadane (1986). Subjective-probability assessment methods for multivariate-t and matrix-t models. In *Bayesian Inference and Decision Techniques*, P.K. Goel and A. Zellner (Eds.). Elsevier, Amsterdam.

[48] H. Doss (1991). Bayesian nonparametric estimation for incomplete data via successive substitution sampling. Technical Report, Department of Statistics, Florida State University.

[49] W. DuMouchel and J. Harris (1983). Bayes methods for combining the results of cancer studies in human and other species (with discussion). *J. Amer. Statist. Assoc.* **78**, 293–315.

[50] T.S. Ferguson, E.G. Phadia, and R.C. Tiwari (1992). Bayesian nonparametric inference. In *Current Issues in Statistical Inference: Essays in Honor of D. Basu*, M. Ghosh and P.K. Pathak (Eds.). Institute of Mathematical Statistics, Hayward CA.

[51] D.K.H. Fong and J. Berger (1992). Ranking, estimation, and hypothesis testing in unbalanced two-way additive models – a Bayesian approach. To appear in *Statistics and Decisions.*

[52] P. Fougere (Ed.) (1990). *Maximum Entropy and Bayesian Methods.* Klower Academic Publ., Dordrecht.

[53] S. Geisser (1992). Bayesian perturbation diagnostics and robustness. In *Bayesian Analysis in Statistics and Econometrics*, P.K. Goel and N.S. Iyengar (Eds.). Springer-Verlag, New York.

[54] A.E. Gelfand, D.K. Dey and H. Chang (1992). Model determination using predictive distributions with implementation via sampling-based methods. In *Bayesian Statistics* **4**, J. Bernardo, J. Berger, A. Dawid, and A.F.M. Smith (Eds.). Oxford University Press.

[55] A.E. Gelfand, S. Hills, A. Racine-Poon and A.F.M. Smith (1990). Illustration of Bayesian inference in normal models using Gibbs sampling. *J. Amer. Statist. Assoc.* **85**, 972–985.

[56] A.E. Gelfand and A.F.M. Smith (1990). Sampling based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.* **85**, 398–409.

[57] A. Gelman and D.B. Rubin (1992). On the routine use of Markov Chains for simulation. In *Bayesian Statistics* **4**, J. Bernardo, J. Berger, A. Dawid, and A.F.M. Smith (Eds.). Oxford University Press.

[58] S. Geman (1988). Experiments in Bayesian Image Analysis. In *Bayesian Statistics 3*, J. Bernardo, M. DeGroot, D. Lindley and A. Smith (Eds.). Oxford University Press.

[59] S. Geman and D. Geman (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. Patt. Anal. Mach. Intel.* **6**, 721–740.

[60] C. Genest and J. Zidek (1986). Combining probability distributions: A critique and an annotated bibliography. *Statist. Science* **1**, 114–135.

[61] E.I. George and R.E. McCulloch (1991). Variable selection via Gibbs sampling. Technical report No. 99, Graduate School of Business, University of Chicago.

[62] J. Geweke (1989). Bayesian inference in econometrics models using Monte Carlo integration. *Econometrica* **57**, 1317–1340.

[63] J. Geweke (1992). Priors for macroeconomic time series and their application. To appear in the *Proceedings of the Conference on Bayes Methods and Unit Roots*, a special issue of *Econometric Theory.*

[64] C.J. Geyer (1991). Reweighting Monte-Carlo mixtures. Technical Report No. 568, School of Statistics, University of Minnesota.

[65] J.K. Ghosh and R. Mukerjee (1992). Non-informative priors. In *Bayesian Statistics* **4**, J. Bernardo, J. Berger, A. Dawid, and A.F.M. Smith (Eds.). Oxford University Press.

[66] M. Ghosh (1992a). Hierarchical and empirical Bayes multivariate estimation. In *Current Issues in Statistical Inference: Essays in Honor of D. Basu*, M. Ghosh and P.K. Pathak (Eds.). Institute of Mathematical Statistics, Hayward CA.

[67] M. Ghosh (1992b). On some Bayesian solutions of the Neyman-Scott problem. To appear in *Statistical Decision Theory and Related Topics V*, S.S. Gupta and J. Berger (Eds.). Springer-Verlag, N.Y.

[68] D. Gianola and R.L. Fernando (1986). Bayesian methods in animal breeding theory. *J. Animal Sci.* **63**, 217–244.

[69] W.R. Gilks and P. Wild (1992). Adaptive rejection sampling for Gibbs sampling. In *Bayesian Statistics* **4**, J. Bernardo, J. Berger, A. Dawid, and A.F.M. Smith (Eds.). Oxford University Press.

[70] P. Goel (1988). Software for Bayesian analysis: current status and additional needs. In *Bayesian Statistics 3*, J.M. Bernardo, M. DeGroot, D. Lindley, and A. Smith (Eds.). Oxford University Press.

[71] M. Goldstein (1988). Adjusting belief structures. *J. Roy. Statist. Soc. B* **50**, 133–154.

[72] E.J. Green and W.E. Strawderman (1992). A comparison of hierarchical Bayes and empirical Bayes methods for simultaneous linear models. Paper of the Journal Series, NJ Agricultural Experiment Station, Rutgers University.

[73] S.S. Gupta and K.J. Miescke (1991). Bayesian look ahead one stage sampling allocations for selecting the largest normal mean. Technical Report #91–60C, Department of Statistics, Purdue University.

[74] I. Guttman (1991). A Bayesian look at the question of diagnostics. Technical Report No. 9104, Department of Statistics, University of Toronto.

[75] W.K. Hastings (1970). Monte-Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109.

[76] J. Hsu and T. Leonard (1992). Bayesian inference for a covariance matrix. To appear in *Ann. Statist.*

[77] J. Ibrahim and P.W. Laud (1991). On Bayesian analysis of generalized linear models using Jeffreys's prior. *J. Amer. Statist. Assoc.* **88**, 981–986.

[78] T.Z. Irony, C.A.B. Pereira, and R.E. Barlow (1992). Bayesian models for quality assurance. In *Bayesian Statistics* **4**, J. Bernardo, J. Berger, A. Dawid, and A.F.M. Smith (Eds.). Oxford University Press.

[79] E. Jaynes (1983). *Papers on Probability, Statistics and Statistical Physics*, R.D. Rosencrantz (Ed.). D. Reidel Publishing Company.

[80] H. Jeffreys (1961). *Theory of Probability* (3rd edition). Oxford University Press, London.

[81] W.S. Jewell (1988). A heteroscedastic hierarchical model. In *Bayesian Statistics 3*, J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith (Eds.). Oxford University Press.

[82] V. Johnson and J. Bowsher (1992). Inference in medical images: a model for segmentation and analysis of noisy images. To appear in *The Statistician*, in the Proceedings volume for the Third International Conference on Practical Bayesian Statistics.

[83] R.E. Kass and E.H. Slate (1992). Reparametrization and diagnostics of posterior nonnormality. In *Bayesian Statistics* **4**, J. Bernardo, J. Berger, A. Dawid, and A.F.M. Smith (Eds.). Oxford University Press.

[84] R. Kohn and C.F. Ansley (1988). The equivalence between Bayesian smoothness priors and optimal smoothing for function estimation. In *Bayesian Analysis of Time Series and Dynamic Models*, J.C. Spall (Ed.). Marcel Dekker, New York.

[85] L. Kuo (1986). Computations of mixtures of Dirichlet processes. *SIAM J. Of Scient. and Statist. Comput.* **7**, 60–71.

[86] S.L. Lauritzen and D.J. Spiegelhalter (1988). Local computations with probabilities on graphical structures and their application to expert systems (with discussion). *J. Roy. Statist. Soc. B* **50**, 157–224.

[87] M. Lavine (1989). An approach to robust Bayesian analysis for multidimensional parameter spaces. *J. Amer. Statist. Assoc.* **86**, 400–403.

[88] M. Lavine, L. Wasserman, and R.L. Wolpert (1991). Bayesian inference with specified marginals. *J. Amer. Statist. Assoc.* **86**, 964–971.

[89] E.E. Leamer (1978). *Specification Searches: AdHoc Inferences with Nonexperimental Data.* John Wiley, New York.

[90] P.J. Lenk (1988). The logistic normal distribution for Bayesian, nonparametric, predictive densities. *J. Amer. Statist. Assoc.* **83**, 509–516.

[91] T. Leonard (1978). Density estimation, stochastic processes, and prior information. *J. Roy. Statist. Soc. B* **40**, 113–146.

[92] D. Lindley and A.F.M. Smith (1972). Bayes estimates for the linear model. *J. Royal Statist. Soc.* (Ser. B) **34**, 1–41.

[93] B. Liseo (1991). Elimination of nuisance parameters with reference noninformative priors. Technical Report, Department of Statistics, Universitá di Roma, Rome.

[94] A.Y. Lo and C.S. Weng (1989). On a class of Bayesian nonparametric estimates: II. Hazard rate estimates. *Ann. Inst. Statist. Math.* **41**, 227–245.

[95] T.J. Loredo (1992). The promise of Bayesian inference for astrophysics. To appear in *Statistical Challenges in Modern Astronomy*, E.D. Feigelson and G.J. Babu (Eds.).

[96] D.J.C. Mackay (1992). A practical Bayesian framework for backprop networks. To appear in *Neural Computation.*

[97] U. Makov and A.F.M. Smith (1992). Bayesian methods in insurance. To appear in *The Statistician*, in the proceedings volume for the Third International Conference on Practical Bayesian Statistics.

[98] K.V. Mardia, J.T. Kent, and A.N. Walder (1992). Statistical shape models in image analysis. In *Proceedings of the 23rd Symposium on the Interface*, E.M. Keramidas (Ed.).

[99] R. Meinhold and N. Singpurwalla (1983). Understanding the Kaman filter. *Amer. Statist.* **37**, 123–127.

[100] T.J. Mitchell and J.J. Beauchamp (1988). Bayesian variable selection in linear regression. *J. Amer. Statist. Assoc.* **83**, 1023–1036.

[101] E. Moreno and J.A. Cano (1991). Robust Bayesian analysis with ε-contaminations partially known. *J. Roy. Statist. Soc. B* **53**, 143–155.

[102] C.N. Morris (1983). Parametric empirical Bayes inference: theory and applications. *J. Amer. Statist. Assoc.* **78**, 47–65.

[103] C.N. Morris and S.L. Normand (1992). Hierarchical models for combining information and for meta-analyses. In *Bayesian Statistics* **4**, J. Bernardo, J. Berger, A. Dawid, and A.F.M. Smith (Eds.). Oxford University Press.

[104] P. Müller (1991). A generic approach to posterior integration and Gibbs sampling. Technical Report 91–09, Department of Statistics, Purdue University.

[105] D.E. Myers, M.M. Alli, and A. Edward (1990). Probabilistic analysis of collapsing soil by indicator kriging. *Math. Geology* **90**, 15–38.

[106] J. Naylor and A.F.M. Smith (1982). Application of a method for the efficient computation of posterior distributions. *Applied Statistics* **31**, 214–225.

[107] R.M. Neal (1992). Bayesian training of backpropagation networks by the hybrid Monte-Carlo method. Technical Report CRG–TR–92–1, Department of Computer Science, University of Toronto.

[108] D. Nychka (1988). Bayesian confidence intervals for smoothing splines. *J. Amer. Statist. Assoc.* **83**, 1134–1143.

[109] M.S. Oh and J. Berger (1991). Integration of multimodal functions by Monte-Carlo importance sampling. Technical Report #91–30C, Purdue University.

[110] H. Omre (1987). Bayesian Kriging-merging observations with qualified guesses in Kriging. *Math. Geology* **19**, 25–39.

[111] P. Phillips (1991). To criticize the critics: an objective Bayesian analysis of stochastic trends. *J. Applied Econometrics* **6**, 333–364.

[112] J. Pilz (1991). *Bayesian Estimation and Experimental Design in Linear Regression Models* (2nd edition). Wiley, NY.

[113] W. Polasek and K. Pötzelberger (1988). Robust Bayesian analysis in hierarchical models. In *Bayesian Statistics* **3**, J.M. Bernardo, M.H. DeGroot, D.V. Lindley, and A.F.M. Smith (Eds.). Oxford University Press.

[114] J. Press (1989). *Bayesian Statistics.* Wiley, NY.

[115] A. Raftery (1992). How many iterations in the Gibbs sampler? In *Bayesian Statistics* **4**, J. Bernardo, J. Berger, A. Dawid, and A.F.M. Smith (Eds.). Oxford University Press.

[116] C.R. Rao (1987). Prediction of future observations in growth curve models (with Discussion). *Statist. Sci.* **2**, 434–471.

[117] C. Robert (1992). *L'Analyse Statistique Bayesienne.* Economica, Paris.

[118] G.K. Robinson (1991). That BLUP is a good thing: The estimation of random effects (with Discussion). *Statist. Sci.* **6**, 15–51.

[119] D.B. Rubin (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Ann. Statist.* **12**, 1151–1172.

[120] D.B. Rubin (1987). *Multiple Imputation for Nonresponse in Surveys.* Wiley, New York.

[121] F. Ruggeri and L. Wasserman (1991). Density based classes of priors: infinitesimal properties and approximations. Technical Report, Carnegie-Mellon University.

[122] M. Schervish (1992). Bayesian analysis of linear models. In *Bayesian Statistics* **4**, J. Bernardo, J. Berger, A. Dawid, and A.F.M. Smith (Eds.). Oxford University Press.

[123] S. Sivaganesan and J. Berger (1992). Robust Bayesian analysis of the binomial empirical Bayes problem. Technical Report, Statistics Department, Purdue University.

[124] A.F.M. Smith and A.E. Gelfand (1992). Bayesian statistics without tears: a sampling–resampling perspective. *American Statistician* **46**, 84–88.

[125] A.F.M. Smith and D. Spiegelhalter (1982). Bayes factors for linear and log-linear models with vague prior information. *J. Royal Statist. Soc.* (Ser. B) **44**, 377–387.

[126] K. Sølna (1992). Including prior information into a marked point process model. In *Bayesian Statistics* **4**, J. Bernardo, J. Berger, A. Dawid, and A.F.M. Smith (Eds.). Oxford University Press.

[127] D. Spiegelhalter and R. Cowell (1991). Learning in probabilistic expert systems. In *Bayesian Statistics 4*, J. Berger, J. Bernardo, A.Dawid and A.F. Smith (Eds.). Oxford University Press.

[128] S. Srinivasan and H. Truszczynska (1990). Approximation to the range of a ratio-linear posterior quantity based on the Frechet derivative. Technical report No. 289, Department of Statistics, University of Kentucky.

[129] D.A. Stephens and P. Dellaportas (1992). Bayesian analysis of generalized linear models with covariate measurement error. In *Bayesian Statistics* **4**, J. Bernardo, J. Berger, A. Dawid, and A.F.M. Smith (Eds.). Oxford University Press.

[130] T.J. Sweeting (1992). On the asymptotic normality of posterior distributions in the multiparameter case. In *Bayesian Statistics* **4**, J. Bernardo, J. Berger, A. Dawid, and A.F.M. Smith (Eds.). Oxford University Press.

[131] M.A. Tanner (1991). *Tools for Statistical Inference: Observed Data and Data Augmentation Methods.* Lecture Notes in Statistics **67**, Springer Verlag, NY.

[132] M.A. Tanner and W.H. Wong (1987). The calculation of posterior distributions by data augmentation. *J. Amer. Statist. Assoc.* **82**, 528–550.

[133] A. Thomas, D.J. Spiegelhalter, and W. Gilks (1992). BUGS: A program to perform Bayesian inference using Gibbs sampling. In *Bayesian Statistics* **4**, J. Bernardo, J. Berger, A. Dawid, and A.F.M. Smith (Eds.). Oxford University Press.

[134] R. Tibshirani (1989). Noninformative priors for one parameter or many. *Biometrika* **76**, 604–608.

[135] L. Tierney (1990). *Lisp-Stat, an Object-Oriented Environment for Statistical Computing and Dynamic Graphics.* Wiley, New York.

[136] L. Tierney (1991). Markov chains for exploring posterior distributions. Technical Report No. 560, School of Statistics, University of Minnesota.

[137] L. Tierney, R. Kass and J. Kadane (1989). Fully exponential Laplace approximations to expectations and variances of non-positive functions. *J. Amer. Statist. Assoc.* **84**, 710–716

[138] P. van Batenburg, A. O'Hagan, and R. Veenstra (1992). Bayesian discovery sampling in financial auditing. To appear in *The Statistician*, in the Proceedings volume for the Third International Conference on Practical Bayesian Statistics.

[139] A.J. van der Merwe and C.A. van der Merwe (1992). Empirical and hierarchical Bayes estimation in multivariate regression models. In *Bayesian Statistics* **4**, J. Bernardo, J. Berger, A. Dawid, and A.F.M. Smith (Eds.). Oxford University Press.

[140] I. Verdinelli (1992). Advances in Bayesian experimental design. In *Bayesian Statistics* **4**, J. Bernardo, J. Berger, A. Dawid, and A.F.M. Smith (Eds.). Oxford University Press.

[141] G. Wahba (1990). *Spline Models for Observational Data. SIAM*, Philadelphia.

[142] P. Walley (1991). *Statistical Reasoning with Imprecise Probabilities.* Chapman and Hall, London.

[143] S. Weerahandi and J.V. Zidek (1988). Bayesian nonparametric smoothers for regular processes. *Canadian J. Statist.* **16**, 61–74.

[144] M. West (1990). Bayesian computations: Monte-Carlo density estimation. To appear in *J. Roy. Statist. Soc. B.*

[145] M. West and J. Crosse (1992). Modelling probabilistic agent opinion. *J. Roy. Statist. Soc. B* **54**, 285–299.

[146] M. West and J. Harrison (1989). *Bayesian Forecasting and Dynamic Models.* Springer-Verlag, New York.

[147] R.L. Wolpert (1991). Monte-Carlo importance sampling in Bayesian statistics. In *Statistical Multiple Integration*, N. Flournoy and R. Tsutakawa (Eds.). *Contemporary Mathematics*, Vol. 115.

[148] R.L. Wolpert and W.J. Warren Hicks (1992). Bayesian hierarchical logistic models for combining field and laboratory survival data. In *Bayesian Statistics* **4**, J. Bernardo, J. Berger, A. Dawid, and A.F.M. Smith (Eds.). Oxford University Press.

[149] D.A. Wooff (1992). [B/D] works. In *Bayesian Statistics* **4**, J. Bernardo, J. Berger, A. Dawid, and A.F.M. Smith (Eds.). Oxford University Press.

[150] K.Y. Ye and J. Berger (1991). Noninformative priors for inferences in exponential regression models. *Biometrika* **78**, 645–656.

[151] S.L. Zeger and M.R. Karim (1991). Generalized linear models with random effects; a Gibbs sampling approach. *J. Amer. Statist. Assoc.* **86**, 79–86.

[152] A. Zellner (1971). *An Introduction to Bayesian Inference in Econometrics.* Wiley, New York.

[153] A. Zellner (1984). *Basic Issues In Econometrics.* University of Chicago Press, Chicago.

[154] A. Zellner (1988). A Bayesian era. In *Bayesian Statistics* **3**, J. Bernardo, M.H. DeGroot, D. Lindley, and A.F.M. Smith (Eds.). Oxford University Press, Oxford.

[155] J.V. Zidek and S. Weerahandi (1992). Bayesian predictive inference for samples from smooth processes. In *Bayesian Statistics* **4**, J. Bernardo, J. Berger, A. Dawid, and A.F.M. Smith (Eds.). Oxford University Press.