BAYESIAN MODEL CHOICE: ASYMPTOTICS AND EXACT CALCULATIONS

by

Alan E. Gelfand

D.K. Dey

TECHNICAL REPORT No. 470

JUNE 15, 1993

Prepared Under Contract

N00014-92-J-1264 (NR-042-267)

FOR THE OFFICE OF NAVAL RESEARCH

Professor Herbert Solomon, Project Director

DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA 94305-4065

Bayesian Model Choice: Asymptotics and Exact Calculations

A.E. Gelfand and D.K. Dey*

SUMMARY

Model determination is a fundamental data analytic task. Here we consider the problem of choosing amongst a finite (with loss of generality we assume two) set of models. After briefly reviewing classical and Bayesian model choice strategies we present a general predictive density which includes all proposed Bayesian approaches we are aware of. Using Laplace approximations we can conveniently assess and compare asymptotic behavior of these approaches. Concern regarding the accuracy of these approximation for small to moderate sample sizes encourages the use of Monte Carlo techniques to carry out exact calculations. A data set fit with nested non linear models enables comparison between proposals and between exact and asymptotic values.

Key words: Bayes factor, Laplace approximations, Likelihood ratio statistics, Monte Carlo methods.

1. INTRODUCTION

Model determination is a fundamental data analytic task. It is also a complex matter influenced by intended use for the model as well as perspective regarding the class of models being entertained. Thus, it is not surprising that no widely accepted strategy for building models has emerged.

Here we consider a much less ambitious problem, that of choosing, within a Bayesian modeling framework, amongst a finite specified set of models. Such selection is based upon predictive distributions and we provide a general predictive formulation which includes all proposed solutions we are aware of. By employing Laplace approximations (see, e.g., Lindley 1980, Tierney & Kadane 1986) we can carry out asymptotic calculations to conveniently assess and compare asymptotic behavior of these proposals and develop tie–ins with classical likelihood ratio based strategies. We shall then describe how simulation based techniques can be employed to perform exact calculations. In this regard, we wish to use the output of recently discussed Markov Chain Monte Carlo methods for Bayesian computation which supply samples essentially from the posterior distribution. We shall also present some comparison between proposals, and between exact and asymptotic values through the fitting of nested non linear models to a data set consisting of 57 points given in Bates and Watts (1988).

The literature on Bayesian model choice is considerable by now. It begins with the formal Bayes approach which, in the case of two models, results in the Bayes factor. Subsequent work has proposed modified Bayes factors. A current, reasonably thorough review appears in Gelfand, Dey and Chang (1992) and its attendant discussion. Our hope in this paper is to achieve some unification of this substantial literature.

We note that all of this work presumes that choice is made by reducing each model to a single summary number and then comparing these numbers. Such severe data reduction only permits model comparison in aggregate; observation or case–level diagnostics enable a clearer comparison of model performance. For work of this sort in the Bayesian context see Geisser 1988, Pettit and Young 1990, Gelfand, Dey & Chang (1992).

The outline of the paper is thus the following. In Section 2 we review the behavior of the likelihood ratio statistic and well known adjustments to it. In Section 3 we summarize the Bayesian view of model choice through predictive distributions and the Bayes factor. This motivates a general definition of predictive densities in Section 4 which includes known cases in the literature and yields a variety of alternative Bayes factors. Laplace method asymptotics for general predictive densities are discussed in Section 5 and illustrated for various Bayes factors in Section 6. Concern for the accuracy of the Laplace

approximations as well as the limitations of these approximations leads us, in Section 7, to seek arbitrarily accurate exact calculations using Monte Carlo methods. Approximate and exact calculations are applied to a nested pair of nonlinear models in Section 8. We conclude with a few summary remarks in Section 9.

2. CLASSICAL APPROACHES

In what follows, we shall assume a choice between two parametric models denoted interchangeably by joint density $f(\mathbf{y}|\boldsymbol{\theta}_i; M_i)$ or likelihood $L(\boldsymbol{\theta}_i; \mathbf{y}, M_i)$, $i = 1, 2$ where $\mathbf{y}$ is $n \times 1$ and $\boldsymbol{\theta}_i$ is $p_i \times 1$. Since the model choice techniques we consider reduce models to single summary numbers overall selection can be made through such pairwise comparision. Also, in practice models are typically considered in pairs, i.e., model exploration is often evolutionary, modifying a current model to see if improvement ensues. Moreover, classical Neyman–Pearson theory for testing of models requires pairwise processing.

Indeed, a few remarks on classical approaches for model choice seems an appropriate starting point. Informal procedures are generally based upon predictive performance in the form of comparison, in some fashion, of distances between observed values and values predicted under a given model. Occasionally certain optimalities can be ascribed to such procedures. Implementation requires "fitting" the model. Following Neyman–Pearson theory suppose we create the hypotheses H_i: data $\mathbf{y}$ arise from model M_i, $i = 1, 2$ and set, say H_1 as the null hypothesis. If the M_i are completely general there is no optimal test of H_1 vs H_2 unless both models are fully specified. The formulation of a likelihood ratio test requires an unambiguous specification of a null and alternative hypothesis such as in the nested models case where M_1 is the reduced model and M_2 is the full model. The likelihood ratio test then takes the form: reject H_1 if $\lambda_n < c < 1$ where

$$\lambda_n = \frac{L(\hat{\boldsymbol{\theta}}_1,\ \mathbf{y},\ M_1)}{L(\hat{\boldsymbol{\theta}}_2;\ \mathbf{y},\ M_2)}. \tag{1}$$

We assume here and in the sequel a regular case, i.e., the p_i remain fixed as $n \to \infty$ whence, under mild conditions, $-2\log \lambda_n$ is approximately distributed as $\chi^2_{p_2-p_1}$ under H_1. With this approximation, inconsistency of the likelihood ratio test arises, that is,

$$\lim_{n\to\infty} P(\text{choose } M_2 | M_1 \text{ true}) = \lim_{n\to\infty} P(\lambda_n < c | M_1 \text{ true})$$

$$= \lim_{n \to \infty} P(-2\log \lambda_n > -2\log c) = P(\chi^2_{p_2 - p_1} > -2\log c) > 0.$$

In other words, λ_n tends to be too small; the likelihood ratio test gives too much weight to the full model. As a result numerous authors (see below) have proposed penalizing the likelihood in the form $\log L(\hat{\theta}_i; y, M_i) - k(n, p_i)$ where $k(n, p) > 0$, and increasing in n and p. Hence, the full model will be penalized more than the reduced model. We replace $\log \lambda_n$ by the larger quantity $\log \lambda_n + k(n, p_2) - k(n, p_1)$. For the above inconsistency to vanish we need $k(n, p_2) - k(n, p_1) \to \infty$ as $n \to \infty$. The form $k(n, p) = \alpha p$ is most common in the literature (though it does not eliminate inconsistency). Values for α in the interval $1 \leq \alpha \leq 2.5$ appear in e.g. Akaike (1973) and in Bhansali and Downham (1977). Nelder and Wedderburn (1972) suggest $\alpha = \frac{1}{2}$. Aitkin (1991) suggests $\alpha = \log 2$. Choices of k which depend upon n and do eliminate inconsistency include $k(n, p) = \frac{p}{2} \log n$ (Schwarz, 1978), $k(n, p) = p\,\beta \log(\log n)$, $\beta > 2$ (Hannan and Quinn, 1979) and $k(n, p) = n \log(n + 2p)$ (Shibata, 1980).

3. THE BAYESIAN FORMULATION

The Bayesian model adds a prior specification $\pi(\theta)$ to the likelihood specification. Inference is based upon the posterior distribution $\pi(\theta|y) \propto L(\theta; y) \cdot \pi(\theta)$. For Bayesian model choice the two model components may be varied. The case where L is held fixed and π is varied to assess the sensitivity of the posterior to such prior variation is referred to as Bayesian robustness (Berger 1984, 1985). Our intent here is to parallel Section 2, hence to vary L. As such we will assume "noninformative" priors.

The formal Bayesian model choice procedure goes as follows. Let w_i be the prior probability of M_i, $i = 1, 2$ and $f(y|M_i)$ the *predictive distribution* for model M_i, i.e.,

$$f(y|M_i) = \int f(y|\theta_i, M_i) \cdot \pi(\theta_i|M_i)\, d\theta_i.$$

If y_{obs} denotes the observed data then we choose the model yielding the larger $w_i\, f(y_{obs}|M_i)$. If $w_i = \frac{1}{2}$ we use the Bayes factor (of M_1 with respect to M_2)

$$BF = \frac{f(y_{obs}|M_1)}{f(y_{obs}|M_2)}. \tag{2}$$

Jeffreys (1961), (see also Pettit and Young, 1990) suggests interpretive ranges for

the Bayes factor. Foundational arguments (see, e.g., DeGroot, 1970) insist that the only way to compare models is through the "probabilities of these models" and hence, with two models, through the ratio. It is noteworthy that the Bayes factor employs no presumption of nesting and does not require "fitting".

In fact, presuming (2) can be calculated, why would one seek alternatives? One criticism is that if $\pi(\theta)$ is improper (as it usually will be under noninformative specification) then $f(y)$ is as well. Hence, we can not interpret the $f(y|M_i)$ as the "probabilities of these models" nor can we interpret the ratio. Several authors have attempted, primarily in the context of normal data models, to develop a multiplier for BF to overcome this problem (Smith and Spiegelhalter, 1980; Spiegelhalter and Smith, 1982; Pericchi 1984). Pericchi suggests the essence of the problem is that, for a given experiment, the expected increase in information about model parameters varies with the specification of the model and that the multiplier should neutralize this differential.

Closely related to this is the fact that even under proper priors with arbitrarily large sample sizes the Bayes factor tends to attach too little weight to the correct model. An illustration is the well known Lindley paradox dating at least to Bartlett (1957). In the nested model case, under usual regularity conditions, we shall show in Section 6 that $BF \to \infty$ as $n \to \infty$. In other words, regardless of the data, as n grows large, model M_1 will be chosen. The behavior of BF contrasts strikingly with that of λ_n which provides too much support for M_2; BF is qualitatively larger than λ_n. This comparison is quantified in an asymptotic sense in Section 6.

4. GENERAL PREDICTIVE DENSITIES

The use of predictive distributions in some form has long been recognized as the correct Bayesian approach to model determination. In particular, Box (1980) notes the complementary roles of the posterior and predictive distributions arguing that the posterior is used for "estimation of parameters conditional on the adequacy of the model" while the predictive distribution is used for "criticism of the entertained model in light of the current data". In examining two models, it is clear that the predictive distributions will be comparable while the posteriors will not.

Box and others have encouraged a less formal view with regard to Bayesian model choice resulting in alternative predictionist criteria to the Bayes factor. Using cross validation ideas (Stone, 1974; Geisser, 1975) the pseudo Bayes factor (PsBF) arises (Geisser and Eddy, 1979). Aitkin (1991) proposed the posterior Bayes factor (PoBF) while recent work of Berger and Pericchi (1992) introduces the intrinsic Bayes factor (IBF). All

are intended to address the aforementioned problems associated with the Bayes factor.

The underlying suggestion is that we adopt a broader notion of predictive distributions and densities. In fact, we shall say that a predictive density arises by averaging a density defined over some portion of the sample space (arising from the likelihood) with respect to a distribution on the parameter space (arising from a data–based updating of the prior). We assume that the data, $\mathbf{y}$, consists of a set of conditionally independent univariate observations, y_j given the parameter θ. However, minor modifications (replacing marginal densities for the y_j given θ with appropriate conditional densities) permit the handling of more general models.

We introduce some notation. Let y_j, $j = 1,\ldots,n$ be a sequence of independent observations which, under model M_i have density $f(y_j|\theta_i, M_i)$, $i = 1, 2$. Let J_n denote the set $\{1, 2,\ldots,n\}$ and let S be an arbitrary subset of J_n. Define $L(\theta_i; y_S, M_i) = \prod_{j=1}^{n} \{f(y_j|\theta_i, M_i)\}^{d_j}$ where $d_j = 1$ if $j \in S$, $= 0$ if $j \notin S$. Finally, let $\pi_i(\theta_i)$, $i = 1, 2$ be the prior density for θ_i under model M_i with respect to Lebesgue measure.

Consider the formal conditional density

$$f(y_{S_1}|y_{S_2}, M_i) = \int L(\theta_i; y_{S_1}, M_i)\, \pi_i(\theta_i|y_{S_2})\, d\theta_i$$

$$= \frac{\int L(\theta_i; y_{S_1}, M_i) \cdot L(\theta_i; y_{S_2}, M_i)\, \pi_i(\theta_i)\, d\theta_i}{\int L(\theta_i; y_{S_2}, M_i)\, \pi_i(\theta_i)\, d\theta_i} \qquad (3)$$

for S_1, S_2 arbitrary subsets of J_n. The form (3) defines a predictive density which averages the joint density of y_{S_1} with respect to the prior for θ_i updated by y_{S_2}. We take y_{S_1} to be a subset of y since for model choice we want a numerical value for (3).

Examples of (3) in the literature include:

(i) $S_1 = J_n$, $S_2 = \phi$ which yields the standard predictive or marginal density of the data. The denominator integral is ignored in this case.

(ii) $S_1 = \{r\}$, $S_2 = J_n - \{r\}$ which yields the cross–validation density $f(y_r|y_{(r)}, M_i)$ where $y_{(r)} = (y_1, y_2,\ldots,y_{r-1}, y_{r+1},\ldots,y_n)$, as in Stone (1974) or Geisser (1975).

$f(y_r|y_{(r)}, M_i)$ evaluated at the observed data is called the conditional predictive ordinate (CPO), dating to Geisser (1980). The form $\prod_{r=1}^{n} f(y_r|y_{(r)}, M_i)$ has been proposed as a surrogate for $f(y)$ by Geisser and Eddy (1979).

(iii) S_1 a small subset of J_n, usually two or three elements, $S_2 = J_n - S_1$ extending the single point deletion in (ii), as in Peña and Tiao (1992).

(iv) $S_1 = J_n$, $S_2 = J_n$ which yields Aitkin's (1991) posterior predictive density. Aitkin argues that, unlike $f(y)$ which results from averaging the joint density of y with respect to the prior, one should average with respect to the posterior.

(v) $S_1 = J_n - S_2$, $S_2 = \{1, 2,...,[\rho n]\}$ where $[\cdot]$ denotes the greatest integer function. The idea here is that a proportion ρ of the observations be set aside for prior updating with the remainder to be used for model determination. Such an idea was suggested by Atkinson (1978) and by O'Hagan (1991).

(vi) $S_1 = J_n - S_2$, S_2 is a minimal subset (Berger and Pericchi, 1992) i.e., the least number of data points such that $\pi_i(\theta_i|y_{S_2})$ is a proper density. In the regular case, the dimension of S_2 is fixed regardless of n. Then $f(y_{S_1}|y_{S_2}, M_i)$ is proper with, as we will see, with the same asymptotic behavior as $f(y; M_i)$.

Notice that, as $n \to \infty$, (i) and (vi) are qualitatively different from the rest; for (ii) through (v) the cardinality of S_2 approaches ∞ as $n \to \infty$. That is, for (ii) through (v), we are averaging against a distribution over the parameter space which, with increasing sample size, places its mass where θ_i must be. Hence, it is not surprising that (i) and (vi) exhibit different asymptotic behavior from the rest.

The predictive densities (i) , (ii), (iv) and (vi) have been employed in the literature to create model selection criteria. Clearly, (i) produces the Bayes factor, BF, given in (2). From (ii), we obtain the pseudo–Bayes factor, PsBF (Geisser and Eddy, 1979)

$$\text{PsBF} = \frac{\prod_r f(y_r|y_{(r)}, M_1)}{\prod_r f(y_r|y_{(r)}, M_2)}. \tag{4}$$

From (iv), we obtain the posterior Bayes factor (Aitkin, 1991), PoBF,

$$\text{PoBF} = \frac{f(y|y,\ M_1)}{f(y|y,\ M_2)}. \tag{5}$$

Finally, from (vi) we can develop several versions of an intrinsic Bayes factor (IBF) (Berger and Pericchi, 1992). If r is the dimension of the minimal subset and S_l, $l = 1, 2,\ldots,\binom{n}{r}$ indexes the subsets of size r of J_n, then the objects in (vi) are of the form $f(y_{S_l^c} \mid y_{S_l}, M_i)$, where S_l^c denotes the complement of S_l relative to J_n. We might consider the ratio of the averages of such forms,

$$\frac{\sum_l f(y_{S_l^c}|y_{S_l},\ M_1)}{\sum_l f(y_{S_l^c}|y_{S_l},\ M_2)} = \text{BF} \cdot \frac{\sum_l (f(y_{S_l}|M_1))^{-1}}{\sum_l f(y_{S_l}|M_2)^{-1}} \tag{7}$$

or the average of the ratios,

$$\binom{n}{r}^{-1} \sum_l \frac{f(y_{S_l^c}|y_{S_l},\ M_1)}{f(y_{S_l^c}|y_{S_l},\ M_2)} = \text{BF} \cdot \binom{n}{r}^{-1} \sum_l \frac{f(y_{S_l}|M_2)}{f(y_{S_l}|M_1)}. \tag{8}$$

Since $\binom{n}{r}$ can be quite large, Berger and Pericchi suggest instead averaging on the right hand sides of (7) and (8) over a random sample of subsets of size r from J_n.

5. LAPLACE METHOD ASYMPTOTICS

We now investigate the asymptotic behavior of (3). Asymptotics are over the sample space through the sampling distribution of estimates of θ_i as $n \to \infty$. We employ Laplace method approximations as in Tierney and Kadane (1986) whose form depends upon S_1 and S_2. Let $C(S)$ denote the cardinality of the set S. As $n \to \infty$ we have three cases

(a) $C(S_1) \to \infty$, $C(S_2)$ fixed

(b) $C(S_1)$ fixed, $C(S_2) \to \infty$

(c) $C(S_1) \to \infty$, $C(S_2) \to \infty$.

Examples (i) and (vi) of section 4 fall into case (a), (ii) and (iii) into case (b); (iv) and (v) into case (c). For case (a) asymptotics apply only to the numerator in (3). For cases (b) and (c) they apply to both numerator and denominator.

The basic Laplace approximation is

$$\int e^{mh(\theta)} d\theta = e^{mh(\hat{\theta})} (2\pi)^{p/2} m^{-p/2} |-H^{-1}(\hat{\theta})|^{\frac{1}{2}} + O(m^{-1}) \tag{9}$$

where θ is $p \times 1$ with h having unique mode $\hat{\theta}$ and $H(\theta)$ is a $p \times p$ positive definite matrix such that $(H(\theta))_{jk} = \partial^2 h(\theta)/\partial\theta_j \partial\theta_k$. For a ratio of integrals with $g(\theta) > 0$

$$\frac{\int g(\theta) e^{mh(\theta)} d\theta}{\int e^{mh(\theta)} d\theta} = e^{m(h^*(\theta^*) - h(\hat{\theta}))} \cdot \left[\frac{|-H^{*-1}(\hat{\theta}^*)|}{|-H^{-1}(\hat{\theta})|}\right]^{\frac{1}{2}} + O(m^{-2}) \tag{10}$$

where $mh^*(\theta) = mh(\theta) + \log g(\theta)$ with h^* having unique mode $\hat{\theta}^*$ and $H^*(\theta)$ is $p \times p$ positive definite matrix such that $(H^*(\theta))_{jk} = \partial^2 h^*(\theta)/\partial\theta_j \partial\theta_k$.

We now apply (9) and (10) to (3). In case (a) we use (9) for the numerator of (3) with $m = C(S_1)$ and

$$mh_i(\theta_i) = \log L(\theta_i; y_{S_1}, M_i) + \log L(\theta_i, y_{S_2}, M_i) + \log \pi_i(\theta_i). \tag{11}$$

Let $\bar{\theta}_i(S_1, S_2)$ be the mode of (11). Then for case (a) we have:

$$f(y_{S_1} | y_{S_2}, M_i) \approx L(\bar{\theta}_i(S_1,S_2); y_{S_1}, M_i) \cdot L(\bar{\theta}_i(S_1,S_2); y_{S_2}, M_i) \cdot \pi_i(\bar{\theta}_i(S_1,S_2))$$

$$\cdot (2\pi)^{p_i/2} (C(S_1))^{-p_i/2} |-H_i^{-1}(\bar{\theta}_i(S_1,S_2))|^{\frac{1}{2}} \cdot (f(y_{S_2}; M_i))^{-1}. \tag{12}$$

(12) has $O(n^{-1})$ accuracy. In case (b) we take $g(\theta_i) = L(\theta_i; y_{S_1}, M_i)$ with $m = C(S_2)$ and

$$mh_i(\theta_i) = \log L(\theta_i; y_{S_2}, M_i) + \log \pi_i(\theta_i). \tag{13}$$

Let $\bar{\theta}_i(S_2)$ denote the mode of (13). Applying (10) we have for case (b) $f(y_{S_1} | y_{S_2}, M_i) \approx$

$$\frac{L(\bar{\theta}_i(S_1,S_2); y_{S_1}, M_i)\ L(\bar{\theta}_i(S_1,S_2); y_{S_2}, M_i) \cdot \pi_i(\bar{\theta}_i(S_1,S_2))}{L(\bar{\theta}_i(S_2); y_{S_2}, M_i) \cdot \pi_i(\bar{\theta}_i(S_2))} \left[\frac{|-H_i^{*-1}(\bar{\theta}_i(S_1,S_2))|}{|-H_i^{-1}(\bar{\theta}_i(S_2))|}\right]^{\frac{1}{2}} \cdot \tag{14}$$

(14) has $O(n^{-2})$ accuracy. To handle case (c) assumptions regarding the rates at which the

$C(S_i) \to \infty$ as $n \to \infty$ are required. If, as in examples (iv) and (v), we have $\lim_{n\to\infty} C(S_1)/C(S_2) = k$, then the same approximation as in case (b) arises.

Suppose the usual regularity conditions hold on the likelihoods, $L(\theta_i; y, M_i)$, so that if $\hat{\theta}_{i,n}$ is the maximum likelihood estimator of θ_i based upon a sample of size n we have $\hat{\theta}_{i,n} \xrightarrow{p} \theta_{i,o}$, for some $\theta_{i,o}$ and $n^{-1}(-\frac{\partial^2 \log L(\theta;y,M_i)}{\partial\theta_j\,\partial\theta_k}) \xrightarrow{p} (I(\theta_i))_{jk}$ where $I(\theta)$ denotes Fisher's information matrix. Suppose we assume that $\hat{\theta}_i(S_1,S_2)$ maximizes (11) with $\log \pi_i(\theta_i)$ deleted and that $\hat{\theta}_i(S_2)$ maximizes (13) with $\log \pi_i(\theta_i)$ deleted. Then, provided $C(S_2) = O(n)$, we also have $\cdot = \hat{\theta}_{i,n} + O_p(n^{-1})$ where $\cdot$ can be any of $\hat{\theta}_i(S_2)$, $\hat{\theta}_i(S_1,S_2)$, $\tilde{\theta}_i(S_2)$ or $\tilde{\theta}_i(S_1,S_2)$ whence all of these modes differ by $O_p(n^{-1})$. Moreover, if π_i is continuous $\pi_i(\cdot) = \pi_i(\hat{\theta}_{i,n}) + O_p(n^{-1})$. Finally, $-H(\theta_i) \xrightarrow{p} I(\theta_i)$, and, in fact, $-H(\hat{\theta}_{i,n}) \xrightarrow{p} I(\theta_{i,o})$. As for $H^*(\theta_i)$, in case (b), $L(\cdot; y_{S_1}, M_i)$ is asymptotically negligible so $-H^*(\theta_i) \xrightarrow{p} I(\theta_i)$ and also $-H^*(\hat{\theta}_{i,n}) \xrightarrow{p} I(\theta_{i,o})$. For case (c), if $C(S_1) = O(n)$ then $-H^*(\theta_i) \xrightarrow{p} 2I(\theta_i)$ and also $-H^*(\hat{\theta}_{i,n}) \xrightarrow{p} 2I(\theta_{i,o})$. We can replace $\hat{\theta}_{i,n}$ by $\cdot$ in either H or H^* with the same result holding. In fact, substituting asymptotically equivalent estimators of θ_i in (14), still yields $O(n^{-1})$ accuracy.

6. ASYMPTOTICS FOR VARIOUS BAYES FACTORS

Applying (12) to M_1 and M_2 and using the asymptotics at the end of the previous section, we obtain

$$BF \approx \frac{L(\hat{\theta}_{1,n};y,\ M_1)}{L(\hat{\theta}_{2,n};y,\ M_2)} \frac{\pi_1(\hat{\theta}_{1,n})}{\pi_2(\hat{\theta}_{2,n})} \left[\frac{|-H_1^{-1}(\hat{\theta}_{1,n})|}{|-H_2^{-1}(\hat{\theta}_{2,n})|}\right]^{\frac{1}{2}} (\frac{n}{2\pi})^{(p_2-p_1)/2} \tag{15}$$

whence, with obvious definition of $K(\hat{\theta}_{1,n}, \hat{\theta}_{2,n})$,

$$\log BF \approx \log \lambda_n + \frac{p_2-p_1}{2} \log n + K(\hat{\theta}_{1,n}, \hat{\theta}_{2,n}) \tag{16}$$

Expression (15) appears in Kass and Vaidyanathan (1992). Expression (16) precisely reveals the difference in asymptotic behavior between λ_n and BF in the nested

case; K = O(1) so the Bayes factor will be consistent but will exhibit Lindley's paradox. Also, apart from K (16) yields Schwarz's (1978) BIC adjustment of λ_n. If we ignore K in choosing a model, we can be misled since K need not be negligible (see section 8). Such concerns are pertinent to the ensuing approximations for the other variants of the Bayes factor and encourage exact calculation as discussed in section 7.

The pseudo Bayes factor provides an interesting case for asymptotic approximation. It is built from the cross validation distributions $f(y_r|y_{(r)}, M_i)$ which lend themselves to a variety of approximations based upon (14) and the different but asymptotically equivalent estimators of θ_i discussed above. Suppose, for instance, in the numerator of (14) we replace $\hat{\theta}_i(S_1,S_2)$ by $\hat{\theta}_{i,n}$ and in the denominator we replace $\hat{\theta}_i(S_2)$ by $\hat{\theta}_{i,n}^{(r)}$, the MLE of θ_i based upon the data with y_r removed. We obtain

$$f(y_r|y_{(r)},M_i) \approx \frac{\prod_{j=1}^{n} f(y_j|\hat{\theta}_{i,n},M_i)}{\prod_{j\neq r} f(y_r|\hat{\theta}_{i,n}^{(r)},M_i)} \frac{\pi_i(\hat{\theta}_{i,n})}{\pi_i(\hat{\theta}_{i,n}^{(r)})} \left[\frac{|-H_i^{*-1}(\hat{\theta}_{i,n})|}{|-H_i^{-1}(\hat{\theta}_{i,n}^{(r)})|}\right]^{\frac{1}{2}}. \tag{17}$$

Since the second and third ratios on the right hand side of (17) tend to 1, we also have

$$f(y_r|y_{(r)}, M_i) \approx \frac{\prod_j f(y_j|\hat{\theta}_{i,n}, M_i)}{\prod_{j\neq r} f(y_j|\hat{\theta}_{i,n}^{(r)}, M_i)} \tag{18}$$

and

$$\prod_r f(y_r|y_{(r)}, M_i) \approx \prod_r \frac{\prod_j f(y_j|\hat{\theta}_{i,n}, M_i)}{\prod_{j\neq r} f(y_j|\hat{\theta}_{i,n}^{(r)}, M_i)}. \tag{19a}$$

Two obvious simplifying approximations of (19a) are

$$\prod_r f(y_r|y_{(r)}, M_i) \approx \prod_r f(y_r|\hat{\theta}_{i,n}, M_i) = L(\hat{\theta}_{i,n}; y, M_i) \tag{19b}$$

and

$$\prod_r f(y_r|y_{(r)}, M_i) \approx \prod_r f(y_r|\hat{\theta}_{i,n}^{(r)}, M_i). \tag{19c}$$

Stone (1977) and Geisser and Eddy (1979) discuss these approximations calling (19b) the predictive likelihood and (19c) the quasi–predictive likelihood. Let us compare all three

approximations. Stone looks at (19b) and (19c), and argues that

$$\lim_{n\to\infty} (\log \prod_r f(y_r|\hat{\theta}_{i,n}^{(r)}, M_i) - \log \prod_r f(y_r|\hat{\theta}_{i,n}, M_i)) = -p_i.$$

We can readily compare (19a) and (19c). After some manipulation

$$\log \prod_r \frac{\prod_j f(y_j|\hat{\theta}_{i,n}, M_i)}{\prod_{j\neq r} f(y_j|\hat{\theta}_{i,n}^{(r)}, M_i)} - \log \prod_r f(y_r|\hat{\theta}_{i,n}^{(r)}, M_i)$$

$$= \sum_r (\sum_j \log f(y_j|\hat{\theta}_{i,n}, M_i) - \sum_j \log f(y_r|\hat{\theta}_{i,n}^{(r)}, M_i))$$

$$= -\tfrac{1}{2} \sum_{r=1} (\hat{\theta}_{i,n}^{(r)} - \hat{\theta}_{i,n})^T H_i(\theta_{i,n}^{*(r)}) (\hat{\theta}_{i,n}^{(r)} - \hat{\theta}_{i,n}) \qquad (20)$$

where H_i is the Hessian matrix of $L(\theta_i; y, M_i)$ and $\theta_{i,n}^{*(r)}$ lies between $\theta_{i,n}^{(r)}$ and $\theta_{i,n}$. By standard argumentation for quadratic forms we may show that, as $n \to \infty$, $(20) \to p_i/2$.

Finally, using the definition (4), the approximation (19a), and the above limiting relationships between (19a), (19b) and (19c) we obtain

$$\log \text{PsBF} \approx \log \lambda_n + \frac{p_2 - p_1}{2}. \qquad (21)$$

Hence, using the approximation (19a) we obtain the Nelder and Wedderburn (1972) adjustment of λ_n as in Section 2. Stone observes that using the approximation (19c) we obtain the customary AIC adjustment ($\alpha = 1$) of λ_n (Akaike, 1973).

Turning to the PoBF and recalling earlier discussion about the case (c) asymptotics, suppose in (14) we replace $\tilde{\theta}_i(S_1,S_2)$ and $\tilde{\theta}_i(S_2)$ with $\hat{\theta}_{i,n}$. Then we obtain

$$\text{PoBF} \approx \lambda_n \, 2^{(p_2-p_1)/2}$$

and

$$\log \text{PoBF} \approx \log \lambda_n + \frac{(p_2-p_1)}{2} \log 2. \qquad (22)$$

Result (22) along with some additional asymptotic calculations are provided in Aitkin (1991). Hence, the correcton of the likelihood associated with the PoBF falls below that of Nelder and Wedderburn which, in turn, falls below the customary AIC adjustment. Neither the PsBF nor PoBF will suffer the Lindley paradox.

We next consider the IBF. From expressions (7) and (8), given (15) or (16) the asymptotic behavior of the IBF depends upon that of the right hand side summations. Though the components of the sums are not necessarily independent, in many cases a "law of large numbers" argument will produce a limiting constant whence, asymptotically the IBF behaves like a multiple of the Bayes factor.

7. EXACT CALCULATIONS USING MONTE CARLO METHODS

The accuracy of the previous analytic approximations is unknown in practice. Additionally, these approximations do not produce functional forms since required modes can rarely be obtained as explicit functions of the data. Rather, for a given observed sample, $\hat{\theta}$'s or $\bar{\theta}$'s would be calculated numerically yielding a numerical value for (3). Therefore, we can not study the behavior of or features of such predictive distributions. A sampling–based approach is attractive in avoiding the above difficulties. Such simulation approaches might be noniterative as in standard Monte Carlo (see e.g. Geweke, 1989) or iterative as for example using the Gibbs sampler or other Markov chain Monte Carlo techniques (see e.g. Gelfand and Smith, 1990; Tierney, 1991). We present no detail regarding these techniques. Rather we just describe how they enable arbitrarily accurate estimates of (3) as a function of y_{S_1} or of expectations associated with (3).

Suppose $g(\theta_i)$ is taken as an importance sampling density for $L(\theta_i; y_{S_2}; M_i)\pi_i(\theta_i)$. If θ^*_{ij}, $j = 1,...,B_i$ is a sample from g and we define $w_{ij} = L(\theta^*_{ij}; y_{S_2}; M_i)\ \pi_i(\theta^*_{ij})/g(\theta^*_{ij})$ then a Monte Carlo integration for (3) is

$$\hat{f}(y_{S_1} | y_{S_2}, M_i) = \sum_j L(\theta^*_{ij}, y_{S_1}; M_i) \cdot w_{ij}/\sum_j w_{ij}. \tag{23}$$

If a Markov chain Monte Carlo technique has been used, the output is usually taken to be a sample θ^*_{ij}, $j = 1,...,B_i$ from the posterior $\pi_i(\theta_i | y)$. But then we can take the posterior as the importance sampling density in (23) resulting in the approximant

$$\hat{f}(y_{S_1} | y_{S_2}, M_i)\left[\sum_j \frac{1}{L(\theta^*_{ij}; y_{S_2^c}, M_i)}\right]^{-1} = \sum_j \frac{L(\theta^*_{ij}; y_{S_1}, M_i)}{L(\theta^*_{ij}; y_{S_2^c}, M_i)} \tag{24}$$

where $S_2^c = J_n - S_2$. The estimator (24) is routine to calculate and simulation consistent.

However, its precision depends upon the stability of the weights $w_{ij} = (L(\theta^*_{ij}, y^c_{S_2}, M_i)^{-1}$, that is, upon how good an importance sampling density $\pi_i(\theta_i|y)$ is for $\pi_i(\theta_i|y_{S_2})$. In the case of the PoBF it is perfect and the resulting estimate takes the simple form

$$(\sum_{j=1}^{B_2} L(\theta^*_{2j}; y_{S_1}, M_2)/B_2)^{-1} \sum_{j=1}^{B_1} L(\theta^*_{1j}; y_{S_1}, M_1)/B_1. \tag{25}$$

In the context of cross validation we would expect $\pi_i(\theta_i|y)$ to be a good importance sampling density for each $\pi_i(\theta_i|y_{(r)})$ and (24) becomes a harmonic mean

$$\hat{f}(y_r|y_{(r)}, M_i) = B_i(\sum_j \frac{1}{f_r(y_r|\theta^*_{ij}, M_i)})^{-1} \tag{26}$$

from which the PsBF can be straightforwardly calculated.

Consider the special case of $f(y, M_i)$. Here, if $\tau(\theta_i)$ is a proper density,

$$(f(y, M_i))^{-1} = \int \frac{\tau(\theta_i)}{L(\theta_i; y, M_i)\pi_i(\theta_i)} \cdot \pi_i(\theta_i|y)\, d\theta_i.$$

Thus, our estimator becomes

$$\hat{f}(y, M_i) = (\frac{1}{B}\sum \frac{\tau(\theta^*_{ij})}{L(\theta^*_{ij}; y, M_i)\pi_i(\theta^*_{ij})})^{-1}. \tag{27}$$

In (27) τ plays the role of an importance sampling density and natural choices to "match" the posterior would be multivariate normal or t densities with mean and covariance computed from the θ^*_{ij}'s. If π_i is proper we could take it as τ obtaining an estimator of $f(y; M_i)$ proposed by Newton and Raftery (1991).

Finally, suppose $f(y_{S_1}|y_{S_2}; M_i)$ is proper and we seek the expectation of $h(y_{S_1})$ with respect to this density. Suppose the conditional expectation $a_1(\theta_i) \equiv E(h(y_{S_1})|\theta_i, M_i)$ with respect to $L(\theta_i; y_{S_1}, M_i)$ is available explicitly. Then since $Eh(y_{S_1}|y_{S_2}; M_i) = \int a_1(\theta_i)\, \pi_i(\theta_i|y_{S_2})\, d\theta_i =$

$$\int a_1(\theta_i) \frac{\pi_i(\theta_i|y_{S_2})}{\pi_i(\theta_i|y)} \pi_i(\theta_i|y)\, d\theta_i = f(y_{S_2^c}|y_{S_2}; M_i) \cdot \int a_1(\theta_i) \frac{\pi_i(\theta_i|y)}{L(\theta_i; y_{S_2^c}, M_i)} d\theta_i$$

a Monte Carlo integration for this expectation using (24) takes the form

$$\hat{E}(h(y_{S_1})|y_{S_2}; M_i) = [\sum_j a_1(\theta^*_{ij})/L(\theta^*_{ij}; y_{S_2^c}, M_i)]/ \sum_j 1/L(\theta^*_{ij}; y_{S_2^c}, M_i) \qquad (30)$$

If $a_1(\theta_i)$ is unavailable explicitly, but $y'_{S_{1,j}}$, $j = 1,...,B_i$ is a sample from $f(y_{S_1}|y_{S_2}; M_i)$ then $\hat{E}\, h(y_{S_1}|y_{S_2}; M_i) = B_i^{-1} \sum_j h(y'_{S_{1,j}})$. To draw $y'_{S_{1,j}} \sim f(y_{S_1}|y_{S_2}, M_i)$ it suffices to draw $y'_{S_{1,j}} \sim L(\theta'_{ij}; y_{S_1}, M_i)$ where θ'_{ij}, are a sample from $\pi_i(\theta_i|y_{S_2})$. Smith and Gelfand (1992) discuss converting a sample from one posterior to that from another. In the present case, we must convert the θ^*_{ij} from $\pi_i(\theta_i|y)$ to θ'_{ij} from $\pi_i(\theta_i|y_{S_2})$.

8. A NUMERICAL EXAMPLE

We consider a data example where the objective is to choose between nested nonlinear models. We employ both asymptotic results (Section 6) and exact calculation (Section 7) to numerically compare λ_n and the various Bayes factors. The data concerns the steady state adsorption of o–xylene as a function of oxygen concentration, inlet o–xylene concentration and temperature. The sample of 57 points is presented along with the full model in Bates and Watts (1988, p. 306–309).

The full model M_2 is, in fact, $y_j = b_f(x_j, \theta|M_2) + \epsilon_j$ where x_j is 3×1 and the error ϵ_j is assumed independent $N(0, \sigma^2)$. Here $b_f(x, \theta) = b_1 b_2/(b_1 + .22788\, b_2)$ with $b_1(x, \theta) = \theta_1 x_1 \exp(-\theta_3/x_3)$ and $b_2(x, \theta) = \theta_2 x_2 \exp(-\theta_4/x_3)$. A convenient reduced model, M_1, sets $\theta_3 = \theta_4$ yielding $y_j = b_r(x_j, \theta|M_1) + \epsilon_j$ where $b_r(x, \theta|M_1) = \theta_1\theta_2 x_1 x_2 \exp(-\theta_3/x_3)/(\theta_1 x_1 + 2.2788\, \theta_2 x_2)$. The modeling implicitly assumes that all $\theta_i > 0$. Hence, we take a flat prior on $\beta_i = \log \theta_i$ independent of the prior $\pi(\sigma^2) = \sigma^{-2}$ on σ^2.

A nonlinear regression fitting package (SAS PROC NLIN) was used for M_1 and for M_2 to obtain the maximum likelihood estimates of all the parameters and thus to compute the likelihood ratio statistic $\lambda_n = (\hat{\sigma}_2^2/\hat{\sigma}_1^2)^{n/2} \exp\{p_1 - p_2)/2\}$ which turns out to be .004.

The maximum likelihood estimates and their asymptotic covariance were used to obtain, for each model, a multivariate t–distribution which served as an importance sampling density for a noniterative Monte Carlo approach. "Exact" calculations based on 10,000 simulations along with asymptotic approximations are given in Table 1.

	$\log \lambda_n$	log BF	log PSBF	log PoBF
"Exact" Calculation	–5.52	–4.94	–5.13	–5.27
Asymptotic Approximation	–	–3.50	–5.02	–5.17

Table 1. Monte Carlo estimates and asymptotic approximations for model selection criteria

Table 1 indicates that regardless of the criterion, the full model is overwhelmingly selected. The asymptotic approximation for BF is poor suggesting that for such very nonlinear models the sample size 57 is not large enough; K in (16) is not negligible. More precise approximation using (15) is nontrivial in the present case and seems hard to justify given the relative ease with which the simulation approach can be implemented.

9. CONCLUDING REMARKS

Our effort here has focused on modeling situations where $p_i << n$, so–called regular models. Though this encompasses a broad range of classical problems, much of contemporary Bayesian modeling considers hierarchical or structured random effects models. In such cases p_i or $p_2 - p_1$ can tend to ∞ as $n \to \infty$. Among the disasters which befall us in such nonregular problems are: i) all of the asymptotics presented here break down ii) parameters may not be consistently estimated (a matter which was recognized as early as Neyman and Scott, 1948) iii) under improper priors the posterior need not be proper (a form of Bayesian nonidentifiability). Attention to the prior specification can remedy (ii) and (iii) but not (i). Hence, the model choice criteria discussed here require exact calculation through the approaches of Section 7. In this regard a valuable supplement to these experiment–level criteria is investigation of model performance at the observation or case level.

REFERENCES

Aitkin, M. (1991). Posterior Bayes factors. J.R. Statist. Soc. B, 1, 111–142.

Atkinson, A.C. (1978). Posterior probabilities for choosing a regression model. Biometrika, 65, 39–48.

Akaike, H. (1973). Information theory and the extension of the maximum likelihood principle. In: Proc. 2nd Int. Symp. Information Theory (eds. B.N. Petrior and F. Csaki), 267–281. Budapest: Akademiai Kiado.

Bartlett, M. (1957). A comment on D.V. Lindley's statistical paradox, Biometrika 44, 533–534.

Bates, D.M. and Watts, D.G. (1988). Nonlinear Regression Analysis and Its Applications. John Wiley and Sons.

Berger, J. (1984). The robust Bayesian viewpoint. In: Robustness of Bayesian Analysis, J. Kadane, ed., p. 63–124, North Holland, Amsterdam.

Berger, J. (1985). Statistical Decision Theory and Bayesian Analysis, Springer–Verlag, New York.

Berger, J.O. and Pericchi, L.R. (1992). The intrinsic Bayes factor. Technical Report, Department of Statistics, Purdue University.

Bhansali, R.J. and Downham, D.Y. (1977). Some properties of the order of an autoregressive model selected by a generalization of Akaike's FPE criterion. Biometrika 64, 547–551.

Box, G. (1980). Sampling and Bayes' inference in scientific modeling and robustness (with discussion). J.R. Statist. Soc. A, 143, 382–430.

DeGroot, M.H. (1970). Optimal Statistical Decisions. McGraw–Hill, New York.

Geisser, S. (1975). The predictive sample reuse method with application. J. Amer. Statist. Assoc. 70, 320–328, 350.

Geisser, S. (1980). Discussion to paper of G.E. P. Box. J.R. Statist. Soc. A, 143, 416–417.

Geisser, S. (1988). The future of statistics in retrospect. In: Bayesian Statistics, 3, (J. Bernardo, et. al., eds.). Oxford University Press, Oxford, 147–158.

Geisser, S. and Eddy, W. (1979). A predictive approach to model selection. J. Amer. Statist. Assoc., 74, 153–160.

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

Gelfand, A.E., Dey, D.K. and Chang, H. (1992). Model determination using predictive distributions with implementation via sampling–based methods. In: Bayesian Statistics, 4 (J. Bernardo, et. al., eds.). Oxford University Press. Oxford, 147–167.

Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration. Econometrica, 57, 1317–1339.

Hannan, E.J. and Quinn, B.G. (1979). The determination of the order of an autoregression. J.R. Statist. Soc. B, 41, 190–195.

Jeffreys, H. (1961). Theory of Probability (3rd Edition). Oxford University Press, London.

Kass, R. and Vaidyanathan, S.K. (1992). Approximate Bayes factors and orthogonal parameters, with application to testing equality of two binomial proportions. J.R. Stat. Soc. B, 54, 129–144.

Nelder, J. and Wederburn, R.W.M. (1972). Generalized linear models. J.R. Statist. Soc. A, 135, 370–384.

Newton, M.A. and Raftery, A.E. (1991). Approximate Bayesian inference by the weighted likelihood bootstrap. Technical Report 199, University of Washington.

Neyman, J. and Scott, E. (1948). Consistent estimates based on partially consistent observations. Econometrica, 16, 1–32.

O'Hagan, A. (1991). Discussion to paper of M. Aitkin. J.R. Statist. Soc. B, 136.

Peña, D. and Tiao, G.C. (1992). Bayesian robustness functions for linear models. Bayesian Statistics 4 (J. Bernardo, et. al. eds.) Oxford University Press, 365–389.

Pericchi, L. (1984). An alternative to the standard Bayesian procedure for discrimination between normal linear models. Biometrika, 71, 576–586.

Pettit, L.I. and Young, K.D.S. (1990). Measuring the effect of observation on Bayes factors. Biometrika, 77, 3, 455–466.

Schwarz, G. (1978). Estimating the dimension of a model. Ann. Statist. 6, 461–464.

Shibata, R. (1980). Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. Ann. Statist. 8, 147–164.

Smith, A.F.M. and Gelfand, A.E. (1992). Bayesian statistics without tears: a sampling–resampling perspective. Amer. Statist. 46, 2, 84–88.

Smith, A.F.M. and Spiegelhalter, D. (1980). Bayes factors and choice criteria for linear models. J.R. Statist. Soc. B, 42, 213–220.

Spiegelhalter, D. and Smith, A.F.M. (1982). Bayes factors for linear and log–linear models with vague prior information. J.R. Statist. Soc. B, 377–387.

Stone, M. (1974). Cross–validatory choice and assessment of statistical predictions. J.R. Statist. Soc. B, 36, 111–147.

Stone, M. (1977). An asymptotic equivalence of choice of model by cross–validation and Akaike's criterion. J.R. Statist. Soc. B, 39, 44–47.

Tierney, L. and Kadane, J.B. (1986). Accurate approximations for posterior moments and marginal densities. J. Amer. Statist. Assoc., 81, 82–86.

Tierney, L. (1991). Markov chains for exploring posterior distributions. Technical Report 560, University of Minnesota

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Bayesian Model Choice: Asymptotics & Exact Calculations		5. TYPE OF REPORT & PERIOD COVERED Technical 6. PERFORMING ORG. REPORT NUMBER 470
7. AUTHOR(s) A. E. Gelfand and D.K. Dey		8. CONTRACT OR GRANT NUMBER(s) N0025-92-J-1264
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Stanford University Stanford, CA 94305-4065		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-042-267
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics & Probability Program Code 111		12. REPORT DATE June 15, 1993 13. NUMBER OF PAGES 20
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

THE VIEW, OPINIONS, AND/OR FINDINGS CONTAINED IN THIS REPORT ARE THOSE OF THE AUTHOR(S) AND SHOULD NOT BE CONSTRUED AS AN OFFICIAL DEPARTMENT OF THE ARMY POSITION, POLICY, OR DECISION, UNLESS SO DESIGNATED BY OTHER DOCUMENTATION.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Key words: Bayes factor, Laplace approximations, Likelihood ratio statistics, Monte Carlo methods.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

See Reverse Side

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Bayesian Model Choice: Asymptotics and Exact Calculations

A.E. Gelfand and D.K. Dey*

SUMMARY

Model determination is a fundamental data analytic task. Here we consider the problem of choosing amongst a finite (with loss of generality we assume two) set of models. After briefly reviewing classical and Bayesian model choice strategies we present a general predictive density which includes all proposed Bayesian approaches we are aware of. Using Laplace approximations we can conveniently assess and compare asymptotic behavior of these approaches. Concern regarding the accuracy of these approximation for small to moderate sample sizes encourages the use of Monte Carlo techniques to carry out exact calculations. A data set fit with nested non linear models enables comparison between proposals and between exact and asymptotic values.