Bayesian Empirical Likelihood

Nicole A. Lazar
Department of Statistics
Carnegie Mellon University
Pittsburgh, PA 15213
nlazar@stat.cmu.edu

Summary

Empirical likelihood has been suggested as a data–based, nonparametric alternative to the usual likelihood function. Research has shown that empirical likelihood tests have many of the same asymptotic properties as those derived from parametric likelihoods. This leads naturally to the possibility of using empirical likelihood as the basis for Bayesian inference. Different ways in which this goal might be accomplished are considered. The validity of the resultant posterior inferences is examined, as are frequentist properties of the Bayesian empirical likelihood intervals.

Key words: Alternative Likelihoods, Bayesian Inference, Nonparametric Inference

1 Introduction

The likelihood function is a key concept in statistics. Much of traditional statistical theory is based around ideas arising from the use of a likelihood to model the state of nature. It has long been known, however, that properties of standard likelihood–based estimation and testing procedures, such as consistency and asymptotic normality, can break down, for instance, in problems with many nuisance parameters; in such situations, an alternative to the classical likelihood approach is desirable. One example is the profile likelihood (see for example Barndorff–Nielsen and Cox, 1994, for a definition; also Fraser and Reid, 1989; McCullagh and Tibshirani, 1990 for adjustments to the basic profile likelihood). In other cases, one might not be willing to specify a fully parametric model; perhaps not enough is known about the data generating mechanism, or one is reluctant to make assumptions about it. The quasi likelihood (Wedderburn, 1974) allows the modelling of data in a likelihood–type way, but requires the specification of only the first two moments, instead of a full likelihood function. Such constructs are useful alternatives to the standard likelihood and have been the focus of research over the last few decades.

A relatively recent development is empirical likelihood, (Owen, 1988, 1990) which acts as a non–parametric analog of the usual likelihood. It has been demonstrated that in some circumstances, likelihood ratio tests based on the empirical likelihood inherit properties of parametric likelihoods, for instance, an asymptotic χ^2 distribution for the test statistic and Bartlett correctability (DiCiccio *et al.*, 1991; DiCiccio and Romano, 1989). For other types of models, the higher–order properties don't necessarily carry over

(Lazar and Mykland, 1999). However, in some ways, it can be claimed that empirical likelihood behaves like a likelihood function, or can be treated like one.

The likelihood function is also a major factor for Bayesian data analysis, which combines a likelihood to model the data with prior information about the parameters of the likelihood. In this framework, the likelihood function may be thought of as the means of updating prior knowledge, in light of the data, to yield posterior inference.

Given these ideas, a natural question is the following: Can likelihoods other than purely parametric ones, or, even more specifically, the density from which the data are assumed to be generated, be used as the likelihood portion in a Bayesian analysis? On the face of it, this is a relatively straightforward question. From a purely operational point of view, it is surely possible to write down an "artificial likelihood," multiply it by a prior and get some result. What is less clear, however, are the properties of the posterior quantities that would result. Can the posterior densities obtained from applying Bayes' rule with an artificial likelihood be interpreted as posterior densities in the same way as those built around model–based likelihoods? Asymptotically, will they coincide with some parametric based inference? Will they combine prior and data in ways that are familiar from Bayes' theorem? An initial attempt at answering some of these questions was provided by Monahan and Boos (1992), who developed a criterion for evaluating whether or not an alternative likelihood is proper for Bayesian inference.

Their idea was the following. Consider an alternative to the usual likelihood, call it L_a, built for a functional θ of the underlying unspecified distribution. Denote by $p_a(t|y)$ the "posterior" density at $\theta = t$ which is the result of using L_a as the likelihood component in Bayes' rule. Is L_a a valid likelihood with which to perform Bayesian inference? To answer this question, it is necessary to define what is meant by a valid or proper likelihood. Monahan and Boos (1992) propose a notion of validity that centres on the posterior probability statements based on $p_a(t|y)$; these should obey the rules of probability in a fashion that is consistent with statements derived from Bayes' rule. In practice, it may be difficult to verify the requisite probability calculus directly. A definition of validity based on the coverage properties of posterior sets is therefore suggested instead, along with a numerical technique that may be used to invalidate certain likelihoods.

At the core of the procedure is the concept that sets that are built to include a certain amount of posterior probability should have the correct probability of containing the parameter, assuming that the prior $p(t)$ and model $f(y|t)$ are correctly specified. For standard applications of Bayes' theorem the model and the likelihood would coincide; posterior probability statements are then valid by the logic of Bayes' rule as a law of probability. A posterior density based on L_a is defined to be valid by coverage for the model $f(y|t)$ if and only if $pr\{\theta \in S_\alpha(y)\} = \alpha$ for every $S_\alpha(y)$, a posterior coverage set function of level α, under the measure $p(t)f(y|t)$. L_a itself is defined to be proper in terms of coverage if and only if the posterior $p_a(t|y)$ is valid by coverage for every absolutely continuous prior. We return to this point in Section 3 below.

The simulation method proposed is to calculate, in the one dimensional case,

$$H = \int_{-\infty}^{\theta} p_a(t|y)dt.$$

This corresponds to posterior coverage set functions of the form $(-\infty, t_a^\alpha]$, where t_a^α is the α^{th} percentile point of the posterior density $p_a(t|y)$. If $p_a(t|y)$ is valid by coverage, then H is distributed as Uniform(0,1). Furthermore, if a prior can be found for which the distribution of H is not uniform, then $L_a(y|t)$ is not a proper likelihood for Bayesian inference.

Posterior intervals could also be evaluated in terms of their frequentist properties, in particular length and coverage. Such an assessment should give information on how empirical likelihood, or any other likelihood alternative, combines with a prior and can yield specific answers to the questions asked above.

In the next section, a review of empirical likelihood is given, which leads into a brief discussion of the different ways to set down a prior in this case. Section 3 will explore the method of Monahan and Boos (1992) for determining whether or not empirical likelihood gives valid posterior inference. Section 4 takes up the question of the frequentist properties of posterior intervals obtained using empirical likelihood instead of the data generating density in the inferential process.

2 A Review of Empirical Likelihood

Suppose we have data $y_1, \ldots, y_n$ from some unknown distribution F, which has mean μ, a d–dimensional vector, and variance Σ, a non–singular $d \times d$ matrix. We are interested in inference concerning some functional of F, call it $\theta(F)$. A simple example, and one which is used in the ensuing discussion, is the mean, μ, of the underlying distribution.

Define the empirical likelihood function as

$$L(F) = \prod_{i=1}^{n} p_i,$$

where $p_i = pr_F(Y = y_i)$. This is nonzero only for distributions which place an atom of probability on each of the observed data points. $L(F)$ is maximised by the empirical distribution function, F_n, which puts weight $1/n$ on each of $y_1, \ldots, y_n$. To obtain a profile likelihood for μ, maximise $L(F)$ subject to the constraints

$$p_i \geq 0 \qquad \sum_i p_i = 1 \qquad \sum_i p_i y_i = \mu.$$

The maximizing values for p_i can be found by using a Lagrange multiplier argument and are seen to be

$$p_i = n^{-1}\{1 + \lambda^T(y_i - \mu)\}^{-1},$$

where λ is a vector, $d \times 1$, which satisfies

$$\sum_{i=1}^{n} \frac{y_i - \mu}{1 + \lambda^T(y_i - \mu)} = 0.$$

Therefore the profile empirical likelihood is

$$L_E(F) = \prod \frac{1}{n}\{1 + \lambda^T(y_i - \mu)\}^{-1}.$$

The empirical likelihood ratio is $R(F) = L_E(F)/L(F_n)$, or $\prod np_i$. We can then define explicitly the profile empirical likelihood ratio function,

$$R_E(\mu) = \prod_{i=1}^{n}\{1 + \lambda^T(y_i - \mu)\}^{-1},$$

as well as the empirical likelihood ratio statistic,

$$W_E(\mu) = 2\sum_{i=1}^{n} \log\{1 + \lambda^T(y_i - \mu)\}.$$

To first order $W_E(\mu)$ has a limiting χ^2 distribution; this may be improved by applying a Bartlett correction, as in DiCiccio *et al.* (1991).

In the rest of this paper, assume for simplicity that $d = 1$.

How might empirical likelihood be incorporated into a Bayesian data analytic approach? The most obvious attack might be to use the profile empirical likelihood function as the likelihood part of Bayes' theorem, which states that the posterior distribution of the parameter of interest, given the observed data, is proportional to the product of the prior distribution of the parameter and the likelihood. Generally speaking,

$$p(\theta|y_1, \ldots, y_n) \propto p(\theta) \times p(y_1, \ldots, y_n|\theta),$$

where $p(\theta)$ is the prior density, possibly improper, on θ, the parameter in the likelihood, $p(y_1, \ldots, y_n|\theta)$. The expression $p(\theta|y_1, \ldots, y_n)$ represents the posterior density of the parameter, after conditioning on the data. In this paper, we consider replacing the parametric likelihood $p(y_1, \ldots, y_n|\theta)$ with an empirical likelihood. What, though, is a sensible way to place the prior? Three possibilities suggest themselves: (a) Place a prior on the p_is; (b) place a prior on θ; (c) place a prior on λ.

The first option, placing a prior on the p_is, carries with it the flavour of the Bayesian bootstrap (Rubin, 1981) and exploits some of the connections between empirical likelihood and the bootstrap. As noted above, the empirical likelihood function, $L(F)$, is maximised when equal weight of $1/n$ is put on each of the observed data points. Therefore, the maximum empirical likelihood estimator of a functional $\theta(F)$ is the bootstrap estimator $\theta(F_n)$ (DiCiccio *et al.*, 1989; Hall and La Scala, 1990). The rationale for a prior specification on the p_i would be similar to that underlying the Bayesian bootstrap – to spread the probability on an individual data point from a single number to a distribution, thereby smoothing, and acknowledging explicitly that any given sample doesn't cover all the possible values that F can take. Assigning the prior in this way would also bring Bayesian empirical likelihood into a family of methods which rely on multinomial sampling based on the observed data, the methods differing in the way in which they weight the data points. In addition to the bootstrap and the Bayesian bootstrap, this family includes ordinary empirical likelihood, weighted likelihood bootstrap (Newton and Raftery, 1994), nonparametric tilts and least favourable families (Efron, 1981; DiCiccio and Romano, 1990), as well as importance sampling. On the other hand, this construction would suffer from the same criticisms as are levelled against the Bayesian bootstrap, namely the extreme sensitivity of the results to the model assumptions, in particular the form of prior, which is usually taken to be Dirichlet for reasons of conjugacy. Furthermore, there is no intuitive way to set the prior, since it is not likely that information will be available *a priori* about the p_is, making any type of subjective Bayesian analysis impossible.

From the standard Bayesian perspective, putting a prior on θ is perhaps the most sensible approach. How would a prior on the functional of interest be incorporated into the process of creating a Bayesian empirical likelihood? For instance, if we thought that, in the case of the mean μ, $a < \mu < b$, say, with perhaps no preference for any of the values in the interval (a, b) over any of the others, it would be reasonable to put a uniform prior on μ. Writing

$$p(\mu|y_1, y_2, \ldots, y_n) = p(\mu) \times \prod_{i=1}^{n} p_i(\mu),$$

it is clear that any other prior can be easily substituted. In particular, assuming that one could carry out the calculations, there is no reason to stay within the conjugate family on the p_i. In practice, it will be difficult to write out closed–form expressions for the posterior distributions, and they will be evaluated by simulation or approximated. This gives more freedom to the choice of prior.

Finally, one could put a prior on the Lagrange multiplier, λ. Following Mykland's thinking in defining dual likelihood, (Mykland, 1995) take the parameter of inferential interest to be λ, instead of θ. Setting priors in this instance might also be difficult, as it would be necessary to translate knowledge about the functional θ into knowledge about λ. On the other hand, since λ and θ are not unrelated, it is conceivable that placing a prior on θ implicitly defines a prior on λ. For the remainder of this paper, we will work within the setting that puts a prior on θ, the functional of interest, focusing on the specific case of the mean, μ.

3 Is Empirical Likelihood Valid for Posterior Inference?

This section explores the validity of empirical likelihood for Bayesian analysis from the perspective of Monahan and Boos (1992). As in that paper, the treatment here is qualitative, based on simulations to calculate the statistic H described in the introduction, and comparing its distribution to a Uniform(0,1).

In a simulation experiment, values of the mean, μ, were generated from a Uniform(1,5) prior. Then, observations $y_1, y_2, \ldots, y_n$ were generated from $N(\mu, 1)$ for $n = 2, 5, 10, 50$. For $n = 2$, this procedure was repeated 10000 times; for the other sample sizes, there were only 1000 repetitions, due to the heavy computational burden involved. Distributions of H are shown for these four cases in Figure 1, with the line at 1 drawn in to show where the distribution should lie if it were in fact uniform. Values of the Kolmogorov–Smirnov tests for uniformity are also given; for a sample of size 10000, the critical (95%) value is 0.0136, while for samples of size 1000, it is 0.043. Based on both the plots and the values of the Kolmogorov–Smirnov statistics, it appears that by $n = 50$, empirical likelihood with a uniform prior gives valid Bayesian inference.

What does this statement mean? Monahan and Boos use the Kolmogorov–Smirnov criterion as a way of deciding that a particular non–standard likelihood is invalid for posterior inference. If any prior can be found for which the distribution of H varies from Uniform(0,1), the alternative likelihood is declared to be invalid. However, a more useful

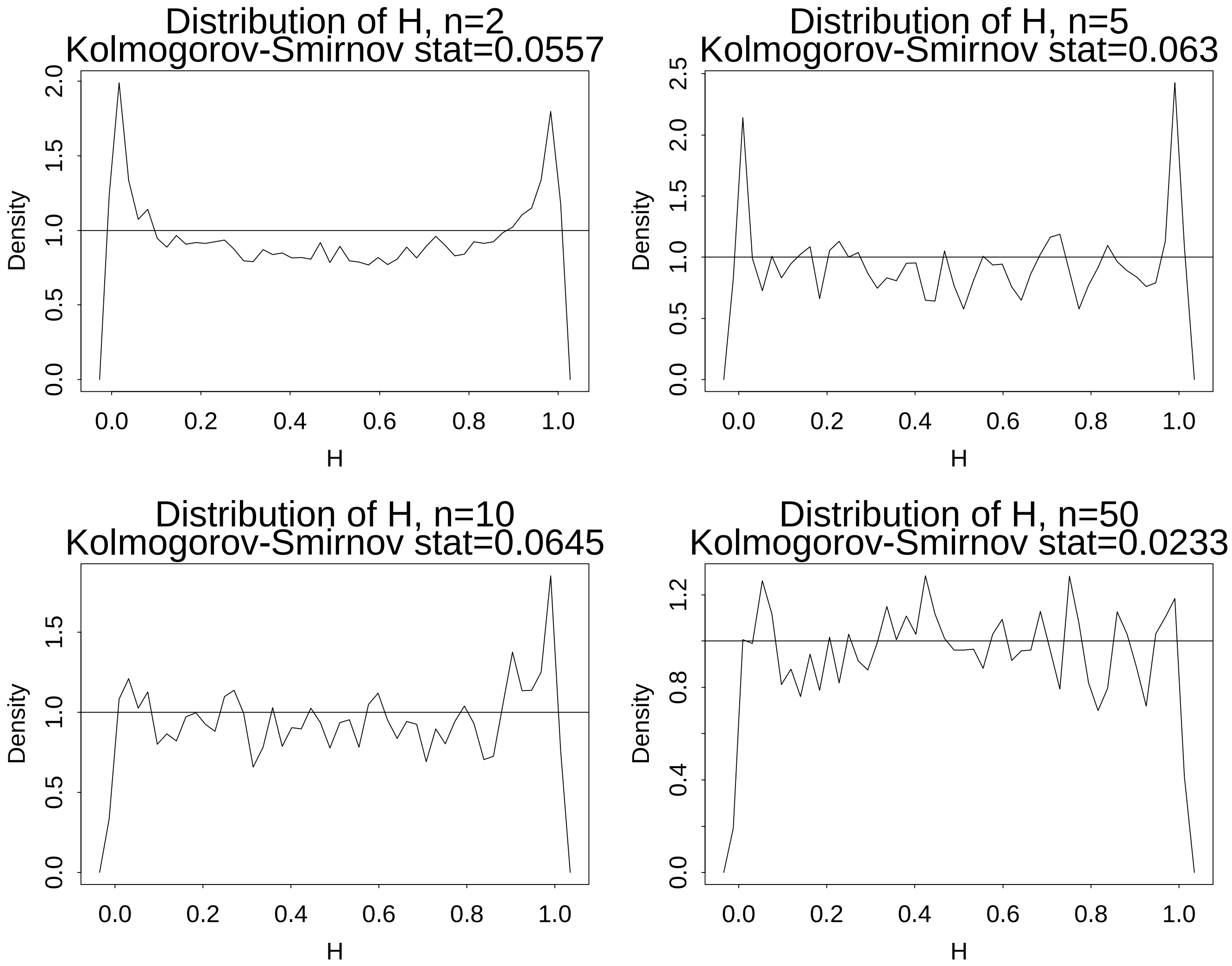

Figure 1: Distribution of the statistic H for empirical likelihood with a uniform prior, for $n = 2, 5, 10, 50$. For $n = 2$, 10000 samples were generated; for $n = 5, 10, 50$, the distribution is based on 1000 samples. By $n = 50$, the statistic appears to be distributed Uniform(0,1).

approach would be to use the method as a diagnostic tool for how close a particular combination of prior and likelihood is to giving valid posterior inference. This conclusion is drawn from a more in–depth study conducted by the author, not reported in detail here, on Jeffreys' nonparametric likelihood for the median, $L_J(y|t) = \binom{n}{nF_n(t)}2^{-n}$, (Jeffreys, 1967, pp. 211–212), one of the examples used in Monahan and Boos (1992). Looking at different sample sizes and priors of gradually increasing diffuseness, it is evident that the distribution of H is affected by the sample size, with larger n resulting in distributions that look more uniform, as might be expected; and by the prior, with the distribution of H becoming more uniform as the prior variance decreases. This was found for Jeffreys' likelihood using both a normal prior and an exponential prior. Instead of concluding that Jeffreys' likelihood is not a valid alternative, information about the circumstances under which proper posterior inferences are obtained may be gleaned by looking carefully at how the distribution of H is affected by prior assumptions.

In light of this, it also seems reasonable to look more closely at the distribution of H itself, to see where it deviates from uniformity. This information is lacking in the Kolmogorov–Smirnov statistic, and is hard to pick out from the density plots as well. Figure 2 shows quantile–quantile plots of H against the uniform distribution, for empirical likelihood and a series of uniform priors on the mean, increasing in diffuseness. For comparison, similar plots are shown for Jeffreys' likelihood, with normal priors of increasing variance (Figure 3). As can be seen in Figure 2, even for the most diffuse of the priors, the distribution of H lies very close to the quantiles of the uniform distribution,

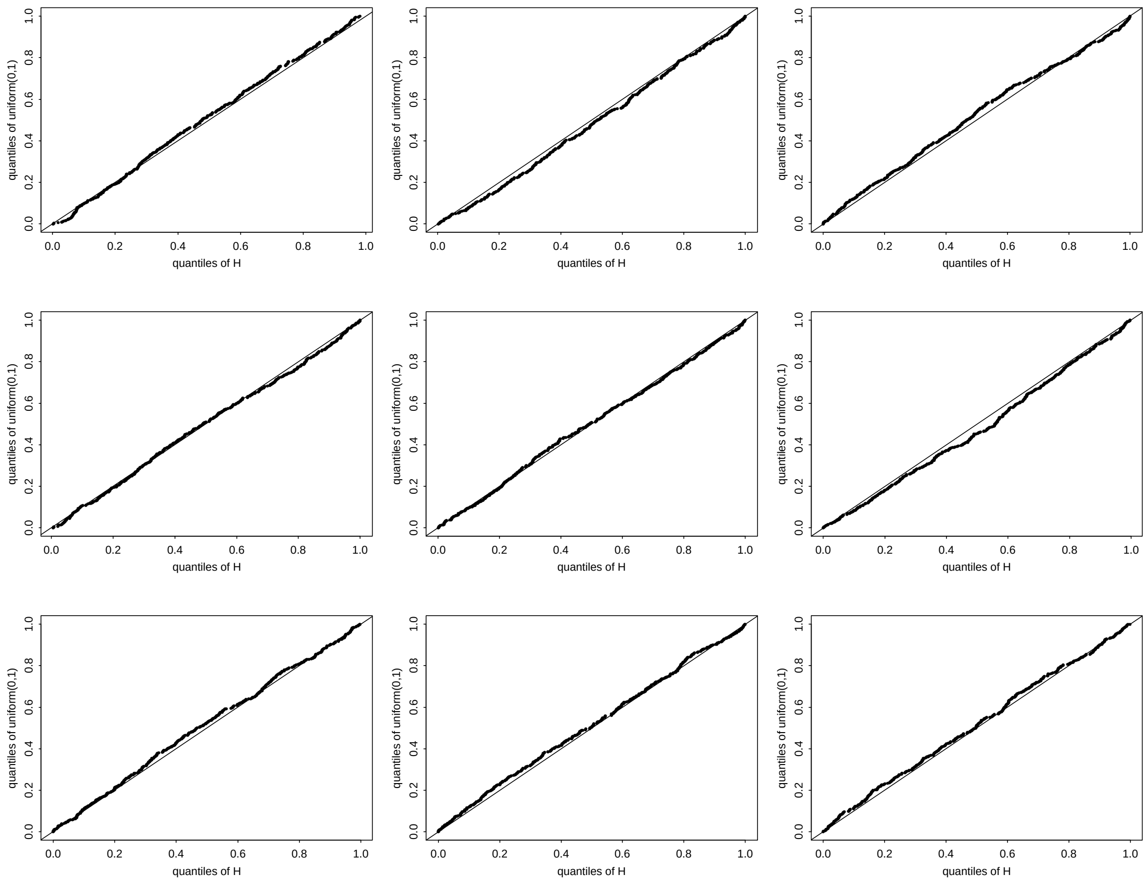

Figure 2: Quantile–quantile plots for the distribution of H against Uniform$(0, 1)$, for empirical likelihood with uniform priors, a sample size of $n = 20$ and simulation size of 500. The uniform priors are all centered at 3, with variance ranging from 0.003 to 12.

bolstering the impression that empirical likelihood gives valid posterior intervals. In contrast, Figure 3 indicates that Jeffreys' likelihood quickly moves away from the uniform distribution; the disparities seem to be more prominent in the lower tail. Empirical likelihood is, to a certain extent, robust to the choice of prior, as will also be seen in the next section.

Based on the criterion suggested by Monahan and Boos, empirical likelihood with a uniform prior set on the functional of interest can provide proper posterior inference. This certainly seems to be true when considering the mean of the underlying distribution.

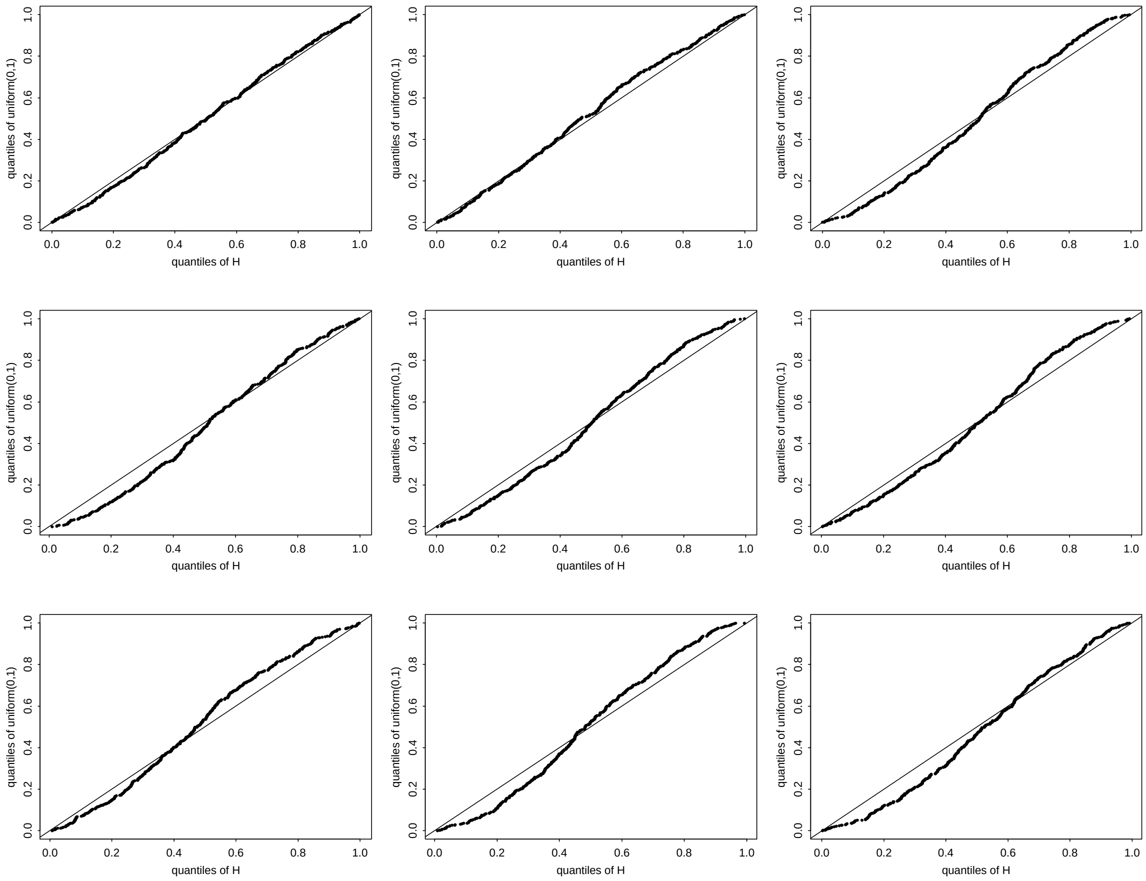

Figure 3: Quantile–quantile plots for the distribution of H against Uniform$(0, 1)$, for Jeffreys' likelihood with normal priors, a sample size of $n = 20$ and simulation size of 500. The normal priors are all centered at 0, with variance ranging from 0.01 to 10.

Even for relatively vague priors, the distribution of the random variable H is very close to uniform, which is not the case, for example, for Jeffreys' likelihood. It should be noted, however, that while this conclusion is relevant for the mean, it might not hold for other functionals. The empirical likelihood for the median is related to Jeffreys' likelihood, and here the two could be expected to exhibit similar behaviours.

4 Frequentist Properties of Posterior Empirical Likelihood Intervals

We now move on to consider frequentist properties of the posterior intervals obtained from using empirical likelihood as the likelihood part of Bayes' theorem, namely length and coverage. Two hundred samples of size $n = 10$ were drawn from a $N(2, 1)$ distribution. Instead of using a normal likelihood, empirical likelihood was used, with a normal prior on the mean, μ_0, which in this case was known to be 2. Twenty different normal priors were investigated, with five possible values of the mean and four of the variance, as given in the tables below. For each of the 200 samples, length and coverage of the standard normal–theory intervals, of the ordinary empirical likelihood intervals, and of the posterior credible intervals based on empirical likelihood, were calculated, for nominal 95% levels. Posterior distributions were approximated on a grid of values. The results of the simulation are given in Tables 1 and 2. For purposes of comparison, the normal–theory intervals had an average length of 1.391, with coverage 95%, while the empirical likelihood intervals based on the χ^2 approximation had average length 1.149 and coverage 91.5%.

Inspection of Table 1 reveals, as would be expected from standard Bayesian theory, that the Bayesian intervals are shorter, on average, than the non–Bayesian intervals. In particular, the intervals for all of the priors with variance $1/n$ are much shorter than the others. They have correspondingly poor coverage when the prior mean is far from the true mean, although they do better than the non–Bayesian intervals when the prior mean is close to the true μ_0. The ordinary empirical likelihood intervals undercover, as established in an unpublished Oxford University technical report by Corcoran, Davison and Spady, 1995; while use of a Bayesian analysis improves the situation for some priors, the improvement is not consistent, and, as can be seen in Table 2, if the prior is very far from the truth and very concentrated, corresponding to a strong prior belief in a wrong value, the posterior credible intervals can be almost arbitrarily bad in their coverage properties. Most of the sensible priors, however, give coverage that is similar to the non–Bayesian empirical likelihood intervals, and offer the advantage that prior information can be incorporated in a relatively straightforward way. It is also easy to check priors of different forms; in this experiment, for example, an exponential prior with mean 5 on μ had average length 1.142 and coverage 90%. Looking at both coverage and length, the posterior empirical likelihood intervals are similar for a range of priors; that is, there is a certain robustness against priors of differing spreads and locations.

Plots of representative posterior distributions reveal that these look more or less

	mean				
var.	2	2+1/n	2+1/$n^{1/2}$	2+3/$n^{1/2}$	5
1/n	0.825	0.825	0.822	0.795	0.518
1/$n^{1/2}$	1.002	1.002	1.000	0.994	0.891
1	1.091	1.090	1.090	1.090	1.075
n	1.138	1.137	1.137	1.137	1.137

Table 1: Average length of posterior intervals, based on 200 simulated data sets.

	mean				
var.	2	2+1/n	2+1/$n^{1/2}$	2+3/$n^{1/2}$	5
1/n	95.5	94.5	93.0	44.25	3.0
1/$n^{1/2}$	91.5	92.0	91.5	88.5	31.5
1	90.75	90.75	90.5	90.25	85.5
n	90.25	90.0	90.0	89.75	89.5

Table 2: Coverage of posterior intervals, based on 200 simulated data sets.

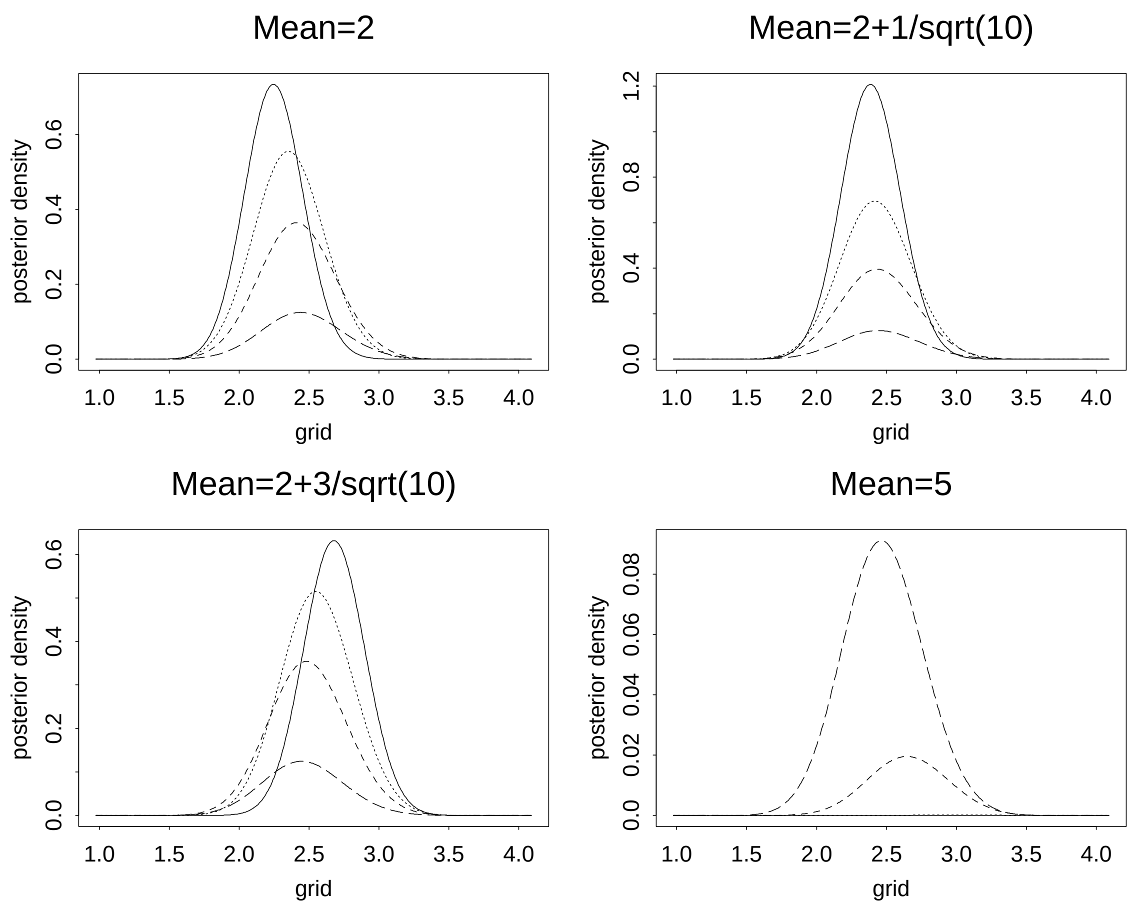

Figure 4: Examples of posterior distributions, with the prior mean fixed but with changing prior variances. The solid line is prior variance $1/10$, the dotted line is prior variance $1/\sqrt{10}$, the dashed line is prior variance 1, and the long dashed line is prior variance 10.

normal. Figures 4 and 5 show the posterior distributions for two of the simulated data sets, with the mean held constant for one and the variance for the other. As can be seen in Figure 4, the prior variance may or may not cause a substantial shift in the posterior distribution, this effect varying from prior mean to prior mean. For fixed prior variance, the prior mean has less of an impact on the posterior density. In all cases, the form of the posterior appears to be very nearly normal. This is true even for a sample drawn from an exponential distribution, with a variety of priors, as in Figure 6. These observations lead to the following theorem.

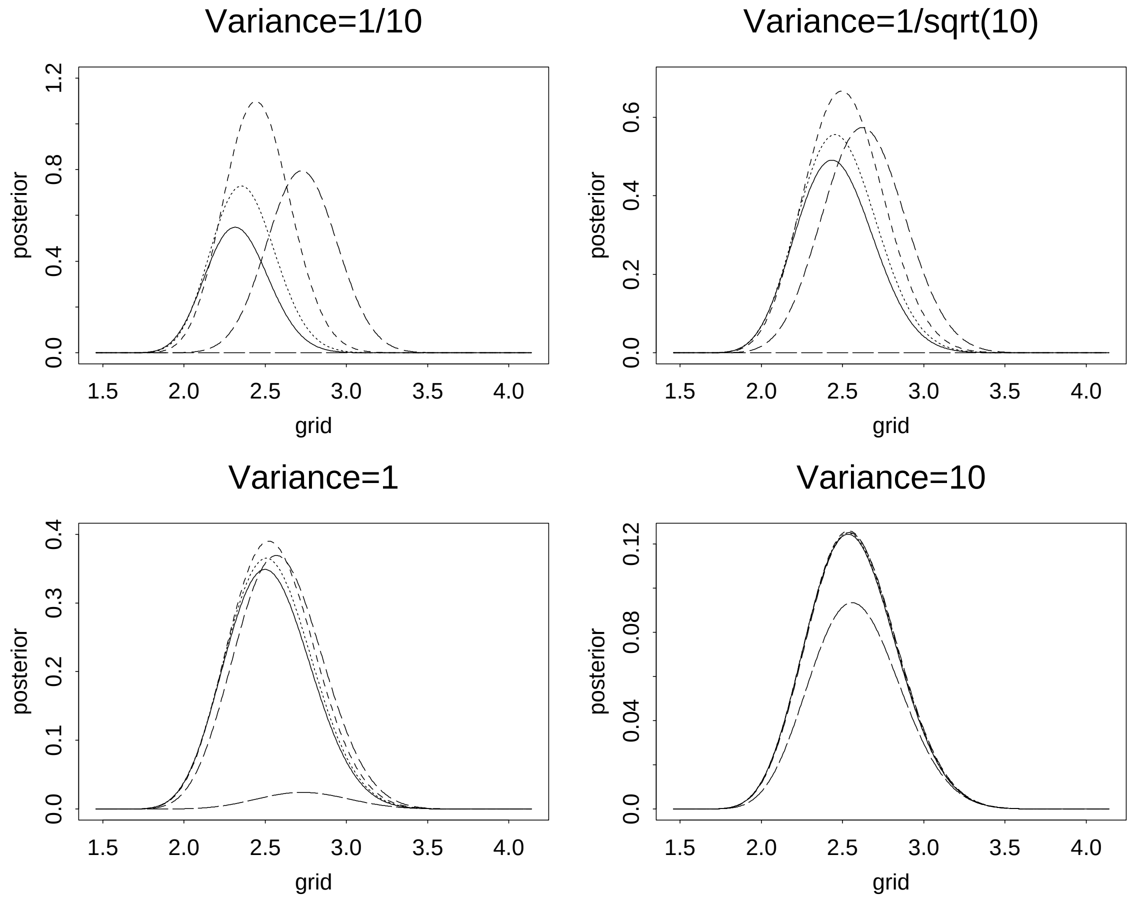

Figure 5: Examples of posterior distributions, fixing the prior variance and letting the prior mean change. The solid curve has prior mean 2, the dotted curve has prior mean 2.1, the dashed curve has prior mean $2 + 1/\sqrt{10}$, the long dashed curve has prior mean $2 + 3/\sqrt{10}$, and the very long dashed curve, falling below the others, has prior mean 5.

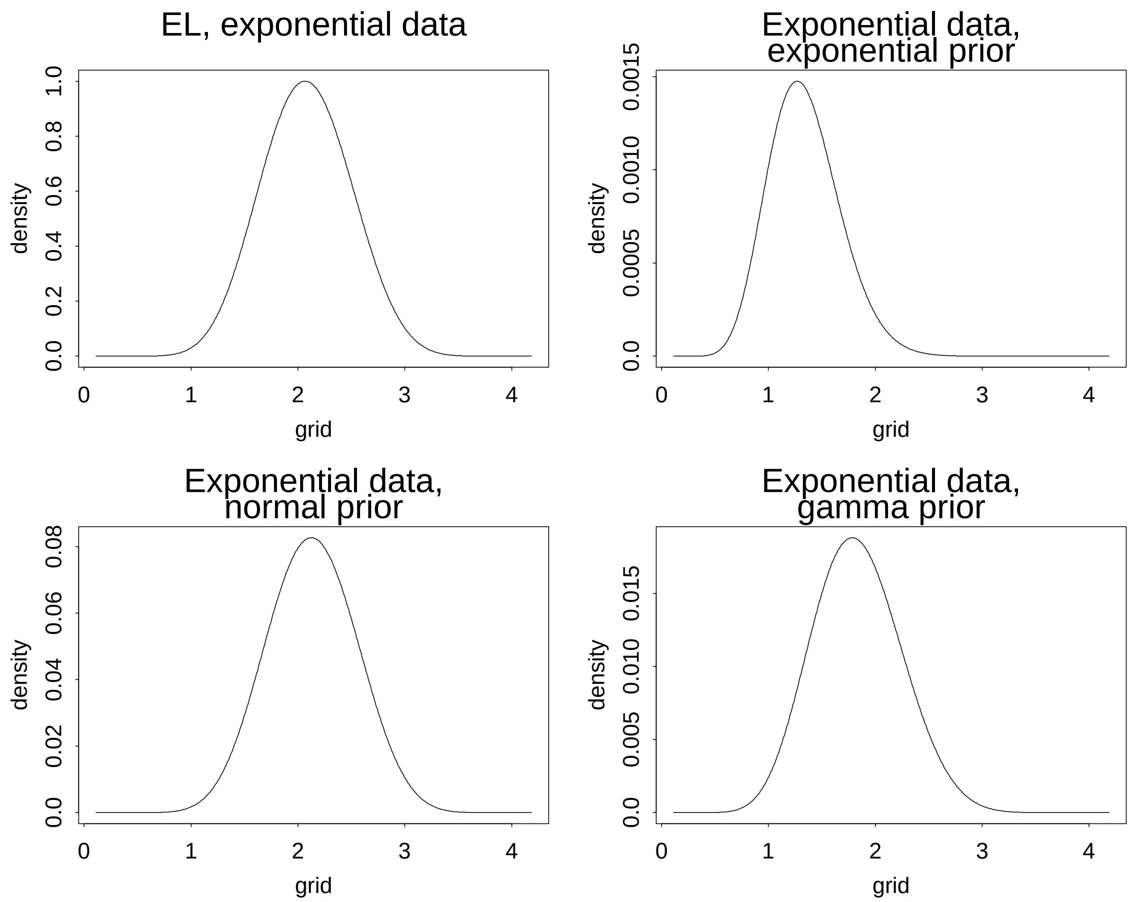

Figure 6: Posterior distributions for a sample drawn from an exponential distribution with mean 2. Priors are exponential with mean 0.2, N(5,10), and Gamma(0.2,1).

Theorem. Under standard regularity conditions, as $n \to \infty$, the posterior distribution of μ converges to normal, with mean m_n and variance J_n, where

$$J_n = J_0 + J(\hat{\mu}_n)$$

$$m_n = J_n^{-1}\{J_0\mu_0 + J(\hat{\mu}_n)\hat{\mu}_n\}$$

with μ_0 being the prior mode, $\hat{\mu}_n$ being the profile maximum empirical likelihood estimate of the mean, J_0 being minus the second derivative of the log prior distribution evaluated at μ_0, and $J(\hat{\mu}_n)$ minus the second derivative of the log empirical likelihood evaluated at $\hat{\mu}_n$.

Proof. The heuristic proof in Bernardo and Smith (1994) applies equally to the situation where the likelihood portion is empirical likelihood, since in the simple case under examination here, the latter inherits the relevant properties, such as behaviour of the low–order cumulants (Mykland, 1994). For more detail on the technical conditions required for the proof, see Bernardo and Smith (1994, Sec. 5.3). Write the posterior distribution as

$$\begin{aligned} p(\mu|y_1, \ldots, y_n) &\propto p(\mu) \times \prod_{i=1}^{n} p_i(\mu) \\ &\propto \exp\{\log p(\mu) + \log L(\mu)\}. \end{aligned}$$

Here, $L(\mu)$ denotes the product of the individual $p_i(\mu)$. Expanding the logarithms of the prior and of the empirical likelihood around their maxima, μ_0 and $\hat{\mu}_n$ respectively, obtained by setting the first derivatives of the logarithms to zero, gives

$$\log p(\mu) = \log p(\mu_0) - \frac{1}{2}(\mu - \mu_0)^2 J_0 + R_0$$

and

$$\log L(\mu) = \log L(\hat{\mu}_n) - \frac{1}{2}(\mu - \hat{\mu}_n)^2 J(\hat{\mu}_n) + R_n.$$

R_0 and R_n are remainder terms, which are assumed small for large n. Therefore,

$$\begin{aligned} p(\mu|y_1, \ldots, y_n) &\propto \exp\{-\frac{1}{2}(\mu - \mu_0)^2 J_0 - \frac{1}{2}(\mu - \hat{\mu}_n)^2 J(\hat{\mu}_n)\} \\ &\propto \exp\{-\frac{1}{2}(\mu - m_n)^2 J_n\}, \end{aligned}$$

m_n and J_n as defined above.

For the small samples of $n = 10$ used in the simulation, there are already strong indications of the normality of the posterior distribution, although comparison of the posteriors resulting from different prior means and variances shows that the posterior mode doesn't yet attain the true value. Rather, for the normal/empirical likelihood combination, the posterior mode combines the prior mean and the data mean, as expected by the theorem, with the prior playing a considerable role, since n is small. The shape of the posterior distribution reflects this as well. An advantage of the empirical likelihood intervals is that they take account of the configuration of the data, becoming more or less symmetrical as the sample does as well as reflecting the orientation of the data. This quality is not lost in the Bayesian analysis – the form of the posterior distribution is a compromise between the prior and the data, through the likelihood. Asymmetries in the data affect the posterior distribution, and hence also the credible intervals, but are mitigated by the symmetry of the normal prior used in the simulations. Interestingly, when the data are themselves drawn from an exponential distribution, the posterior distributions based on a number of different priors may still look quite normal, if the empirical likelihood is symmetric or nearly so.

5 Conclusion

Based on both the Monahan and Boos (1992) heuristic and an examination of the frequentist properties of Bayesian intervals, it appears reasonable to use empirical likelihood within the Bayesian paradigm. This paper explores a simple way of doing so, via a straightforward application of the standard theory with a prior on the functional of interest and the empirical likelihood taking the place of a model–based likelihood. While we have only considered the case of the mean, other results from the literature on empirical likelihood indicate that similar findings should hold for smooth functions of means and for functionals defined via a set of estimating equations when there are no nuisance parameters.

At the outset, it was hoped that the incorporation of good prior information might help to offset the problem with standard empirical likelihood intervals of undercovering. Based on the simulation study reported here this is true in a limited sense only. For many of the priors that were considered, posterior empirical likelihood intervals had coverage that was similar to that of the frequentist intervals based on the χ^2 approximation. For a few priors, in particular those that were centered near the true value of μ under the simulation, there was an improvement in coverage. In spite of the lack of consistent improvement, there are still advantages to using the posterior intervals. One advantage is that they are on average shorter. A second advantage is that it is easy to include prior information, and if we have this knowledge, it is desirable as a matter of principle to use it. The Bayesian approach lends further flexibility to users of empirical likelihood, allowing knowledge about the problem to be allocated between the data component, *i.e.* the likelihood, and the prior on the functional of interest.

Acknowledgements: The author would like to thank Chin–Shan Chuang, Valérie Ventura and Larry Wasserman for helpful conversations and comments on this work.

REFERENCES

Barndorff–Nielsen, O.E. and Cox, D.R. (1994) *Inference and Asymptotics*, London: Chapman & Hall.

Bernardo, J.M. and Smith, A.F.M. (1994) *Bayesian Theory*, Chichester: John Wiley & Sons.

DiCiccio, T.J., Hall, P. and Romano, J.P. (1989) Comparison of parametric and empirical likelihood functions. *Biometrika*, **76**, 465–476.

DiCiccio, T.J., Hall, P. and Romano, J.P. (1991) Empirical likelihood is Bartlett–correctable. *Annals of Statistics*, **19**, 1053–1061.

DiCiccio, T.J. and Romano, J.P. (1989) On adjustments to the signed root of the empirical likelihood ratio statistic. *Biometrika*, **76**, 447–456.

DiCiccio, T.J. and Romano, J.P. (1990) Nonparametric confidence limits by resampling methods and least favorable families. *International Statistical Review*, **58**, 59–76.

Efron, B. (1981) Nonparametric standard errors and confidence intervals (with discussion). *Canadian Journal of Statistics*, **9**, 139–172.

Fraser, D.A.S. and Reid, N. (1989) Adjustments to profile likelihood. *Biometrika*, **76**, 477–488.

Hall, P. and La Scala, B. (1990) Methodology and algorithms of empirical likelihood. *International Statistical Review*, **58**, 109–127.

Jeffreys, H. (1967) *Theory of Probability*, 3rd. ed. Oxford University Press.

Lazar, N.A. and Mykland, P.A. (1999) Empirical likelihood in the presence of nuisance parameters. *Biometrika*, **86**, 203–213.

McCullagh, P. and Tibshirani, R. (1990) A simple method for the adjustment of profile likelihoods. *Journal of the Royal Statistical Society, Series B*, **52**, 325–344.

Monahan, J.F. and Boos, D.D. (1992) Proper likelihoods for Bayesian analysis. *Biometrika*, **79**, 271–278.

Mykland, P.A. (1994) Bartlett type identities for martingales. *Annals of Statistics*, **22**, 21–38.

Mykland, P.A. (1995) Dual likelihood. *Annals of Statistics*, **23**, 396–421.

Newton, M.A. and Raftery, A.E. (1994) Approximate Bayesian inference by the weighted likelihood bootstrap (with discussion). *Journal of the Royal Statistical Society, Series B*, **56**, 3–48.

Owen, A.B. (1988) Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, **75**, 237–249.

Owen, A.B. (1990) Empirical likelihood confidence regions. *Annals of Statistics*, **18**, 90–120.

Rubin, D.B. (1981) The Bayesian bootstrap. *Annals of Statistics*, **9**, 130–134.

Wedderburn, R.W.M. (1974) Quasi–likelihood functions, generalised linear models, and the Gauss–Newton method. *Biometrika*, **61**, 439–447.