Re-examining informative prior elicitation through the lens of MCMC

Eugene D. Hahn

Salisbury University, Salisbury, MD, USA

Summary. In recent years, advances in Markov chain Monte Carlo (MCMC) techniques have had a major impact on the practice of Bayesian statistics. An interesting but hitherto largely underexplored corollary of this fact is that MCMC techniques make it practical to consider broader classes of informative priors than have been used previously. Conjugate priors, long the workhorse of classic methods for eliciting informative priors, have their roots in a time when modern computational methods were unavailable. In the current environment more attractive alternatives are practicable. A re-appraisal of these classic approaches is undertaken, and principles for generating modern elicitation methods are described. A new prior elicitation methodology in accord with these principles is then presented.

Keywords: Bayesian inference, Markov chain Monte Carlo, nonconjugate, Kullback-Liebler divergence, pentagonal distribution

1. Introduction

The Bayesian paradigm allows one to incorporate prior information into statistical models for decision making. This prior information is combined with information from the data using the axioms of probability to yield posterior distributions for parameters of interest. Bayes' rule can be written as $p(\theta|y) \propto \ell(y|\theta)p(\theta)$, which is to say that the posterior is proportional to the likelihood times the prior. In recent years, advances in Markov chain Monte Carlo (MCMC) techniques have had a major impact on the practice of Bayesian statistics (Berger, 2000). A great deal of this development has resulted because MCMC techniques have made it practical to consider more complex models than were employed previously. To date, researchers have focused extensively on how MCMC may be used for problems in which the likelihood function is more complex. An interesting but hitherto largely unexamined corollary is that MCMC techniques can make it practical to consider broader classes of informative priors than have been used previously. That is, the same techniques that have made the use of more complex likelihoods, $\ell(y|\theta)$, tractable for inference can be used to make more complex priors, $p(\theta)$, tractable for inference. Since modern MCMC techniques allow for the development of new classes of flexible informative priors, ample room exists for an echo of the MCMC explosion in the area of research on informative prior assessment, specification and development.

In the following section, we review past research on methods of prior assessment. In Section 3, we identify four principles that may be used to guide research on methodologies for specifying and eliciting informative priors. In Section 4, we present a method which satisfies these principles. We then illustrate the method with two examples in Section 5 and draw conclusions in Section 6.

2. Current practices for prior specification

In the practice of Bayesian statistics, one may use a non-informative or reference prior obtained by formal rules such as a Jeffreys's prior (Kass and Wasserman, 1996). This approach is useful for

cases in which expert judgment is unavailable or not of interest. We may also wish to develop an informative prior that incorporates historical information and/or expert judgment. Although there are some earlier exceptions (e.g., Phillips and Edwards, 1966; Smith, 1967; Winkler, 1967), many of the more recent works in this area have been designed to elicit what is known as a conjugate prior (Raiffa and Schlaifer, 1961). In this approach, for the prior one selects a distribution that has the same functional form as that of the likelihood. Then the decision task involves a determination of which specific instance of this distribution appropriately represents the expert's beliefs. The advantage of this approach is that the work required of the statistician to obtain the posterior is usually substantially reduced because the prior and the likelihood have the same form.

Nonetheless, it is well known that psychological phenomena may intrude into the process of capturing an expert's subjective probability distributions (Wallsten and Budescu, 1983), which may potentially contaminate the results of the elicitation. So, in developing strategies for specifying conjugate priors, researchers have recognized the importance of carefully eliciting an expert's judgments so that the translation from belief to a probability distribution is as accurate as possible. As a result, a wide variety of procedures for eliciting conjugate priors have been developed. Procedures are often designed for a particular setting or model, or reflect a particular philosophy. A brief review of methods for eliciting conjugate priors provides an indication of the breadth in this area of research and serves to identify key themes in the literature.

Staël von Holstein (1971) described a method for assessing a conjugate prior for a Bernoulli process. His technique for assessing beta distributions made use of estimates of the median and the first and third quartiles of experts' subjective probability distributions. Chaloner and Duncan (1983) presented a method called predictive modal estimation in the context of assessing beta priors. Here, the expert first provides an assessment of the mode of the distribution. Then he or she assesses the likelihood of other points along the distribution relative to the likelihood of the mode. Chaloner and Duncan argued that such a method is in accordance with the anchoring and adjustment processes that people are known to use in their assessments of uncertainty. Inconsistencies in judgments were resolved via a feedback mechanism in which judgments were respecified if they were inconsistent with previously obtained ones. Kadane *et al.* (1980) presented a method for estimating conjugate priors for a linear regression model. Their method made use of the multivariate t predictive distribution which involved the assessment of belief about measures of central tendency for the regression coefficients, as well as assessment of belief about variation/covariation and the appropriate value of a degrees of freedom parameter. The necessary judgments were obtained by eliciting fractiles as was done by Staël von Holstein (1971). Winkler *et al.* (1978) have also considered the elicitation of conjugate priors for the linear regression model using predictive distributions. Garthwaite and Dickey (1988) likewise examined the linear regression problem but introduced a technique based on the concept of points of constrained minimum variance. In this technique, certain values of the independent variables are given to the expert. The expert is then asked to select values for the remaining independent variables so that his or her uncertainty regarding the dependent variable is minimized. A small number of fractile assessments are also required of the expert in this approach to complete the elicitation of the subjective probability distributions. Carlin *et al.* (1992) presented an approach for eliciting normal and inverse gamma priors in the context of random-effects logistic regression models based on a series of problem-specific interrelated probabilistic considerations. The case of modeling with ellipsoidal distributions is examined in Dickey and Chen (1985). Winkler *et al.* (2002) advanced the use of beta priors assessed from fractiles in the context of the zero-numerator problem. Finally, a software program that aids in the elicitation of the parameters of the v-Poisson distribution is described in Shmueli *et al.* (2001). The software makes use of interactive graphical

displays and automatically ensures elicited parameters obey consistency considerations.

A few key themes emerge from a consideration of these works. First among these is a desire to simplify the judgment tasks in order to facilitate accuracy. A second theme is the use of feedback mechanisms to address inconsistencies as well as to improve accuracy. One type of feedback mechanism automatically intervenes to prevent experts from making judgments that are formally inconsistent with previously elicited information (e.g., Chaloner and Duncan, 1983; Shmueli *et al.*, 2001). Alternatively, feedback may occur in the context of a side-by-side review involving the expert and the statistician as to the accuracy and sensibility of the results obtained (e.g., O'Hagan, 1998). Another theme that emerges perhaps implicitly is a consideration of the implementation demands that the method places on the statistician. The abovementioned methods were all designed for eliciting informative conjugate priors, which historically have been more amenable to computation. By contrast, the literature on eliciting informative non-conjugate priors is not as extensive (however, see Chaloner, 1996; Gelfand *et al.*, 1995).

A final theme is the acceptance of an underlying assumption that some conjugate distribution reasonably resembles the prior distribution held by the expert. There are undoubtedly many instances where this assumption is not unreasonable and the assumption can be checked via sensitivity analyses (e.g., O'Hagan *et al.*, 1992). Nonetheless, some have challenged the tenability of this assumption (Chaloner, 1996; Gelfand *et al.*, 1995). More importantly, the primary motivation underlying the use of the conjugate approach typically centers on the computational benefits it provides, not whether it is necessarily the best representation of the expert's beliefs (Kadane et al., 1980). It is therefore quite possible in some cases that the task of the expert becomes one of satisficing (Simon, 1957) given the particulars of the method. For example, Staël von Holstein (1971) found some evidence of subjects trying to use distributions that were not in the beta family. While some of these subjects may have been making mistakes, others may have had beliefs that were inconsistent with a beta distribution. Moreover, some consistency-inducing feedback mechanisms described above will prevent experts from making "mistakes"; however, methods that constrain responses to be consistent with what has been articulated previously are predicated on the assumption that previously articulated beliefs are accurate. In such a method, an initial mistake may negatively impact the remaining elicitations, despite an expert's attempts at correction. Mixture priors (Dalal and Hall, 1983; Diaconis and Ylvisaker, 1985) may be used to provide additional flexibility, but the associated cost of decreased simplicity in the elicitation task may not be desirable.

To summarize, researchers have created many useful methods that capture conjugate priors from the beliefs of experts by way of relatively straightforward elicitation processes. Without such methods and their concomitant consideration of the needs of experts, the experts would be burdened with a substantial amount of additional complexity in articulating beliefs in a format suitable for the construction of an informative prior. However, with the advent of MCMC techniques it is no longer necessary to devise elicitation procedures that are confined to the realm of conjugate priors. Instead, we can see that requiring elicited distributions to be conjugate is at best no longer necessary and at worst possibly detrimental to one's objectives. As such, we offer four principles that may be of use in guiding the development of elicitation methodologies in the MCMC era.

3.   Methodologies for informative prior elicitation: Four principles

A review of the research into the development of elicitation methodologies for informative priors suggests the following principles may be useful for guiding these efforts.

Principle 1: It is desirable for elicitation methodologies to produce distributions which are flexible in form. Expert opinion is capable of taking on a wide range of possibilities and as such methodologies should be flexible enough to accurately capture a wide variety of distributions representing expert opinion. For example, distributions representing expert belief may possess skewness, heavy tails, or multiple modes. All other things being equal, elicitation methodologies should be capable of generating distributions which possess properties such as these in order to be faithful to expert opinion.

Principle 2: It is desirable to minimize the cognitive demands that an elicitation methodology places on the expert. Research suggests that simpler probabilistic judgment tasks are more likely to be accurately completed than are complex ones (Hogarth, 1975). Hence, tactics such as breaking a complex judgment down into a series of more straightforward ones, or re-framing a question in terms of more easily envisaged contexts are likely to be beneficial. We also would want to consider addressing concerns about consistency among judgments. Requiring an expert to self-impose perfect consistency among judgments while simultaneously rendering those judgments may be unduly challenging. Rather, a method which handles consistency issues adjunctively to the greater decision task may be preferable.

Principle 3: It is desirable to minimize the demands that an elicitation methodology places on the statistician. The widespread use of conjugate techniques suggests that this is an important concern. Methodologies which are cumbersome for the statistician to implement will be less attractive than those which are not. Hence, developed methodologies should lend themselves readily to computational efforts.

Principle 4: All other things being equal, methodologies for prior elicitation which can be easily applied to a wide range of models or scenarios may have some added desirability. The vast majority of the methods mentioned in the above literature review were specifically tailored for a certain class of model, or even a certain application. While it is desirable to have methodologies that are highly focused on certain problems, it would also be attractive to have more general methodologies that could be used in a variety of settings. Such approaches would minimize the need to invent and validate elicitation methodologies on a case-by-case basis. For example, an approach that could be applied to real-valued parameters, strictly positive parameters, and parameters existing on the unit interval could save development work on the part of the statistician. Moreover, such a unified approach might be helpful to the expert as only a single elicitation methodology needs to be learned.

4. Elicitation of informative priors: A nonconjugate approach

We describe here a methodology for eliciting a nonconjugate prior that is suitable for use with MCMC methods. The goal is to construct a probability density function (pdf) or, loosely speaking, a histogram for a single parameter of interest. Here division of the parameter space into intervals each having an associated probability is needed to define a proper prior. Thus, consider partitioning the parameter space Θ into k intervals (or "bins"). Denote the ith interval as θ_i. In the first step, the expert is asked to make a series of judgments indicating the relative likelihood or odds of θ_1 as compared to θ_j, where $j = 2$ to k. In this step we obtain $k - 1$ odds ratios of the form $O(\theta_1/\theta_2)$, $O(\theta_1/\theta_3), \ldots, O(\theta_1/\theta_j)$ from the expert. In the second step, the expert is asked to provide a second series of judgments indicating the odds of θ_2 as compared to θ_j, where $j = 3$ to k. This process is repeated by eliciting relative odds ratios for all θ_i and θ_j, $i < j$, until the final judgment $O(\theta_{k-1}/\theta_k)$ is elicited. This requires $k(k-1)/2$ judgments from the expert. We may then construct a matrix

of the relative odds ratios θ_i/θ_j. In constructing this matrix, we note that $O(\theta_i/\theta_i)$ is unity and $O(\theta_j/\theta_i) = O(\theta_i/\theta_j)^{-1}$. Let $\xi_{i,j}$ equal $O(\theta_i/\theta_j)$ and write the matrix $\Xi = (\xi_{i,j})$. Then, we may obtain the underlying pdf by a method involving the Kullback-Leibler divergence. We begin by noting that for a discrete distribution the directed divergence of Kullback and Leibler (1951) is

$$D(P,Q) = \sum_{i=1}^{I} p_i \log \left(\frac{p_i}{q_i} \right).$$

This is a measure of the divergence or distance between the distributions P and Q. In particular, the measure is known as the directed divergence as it is a measure of the information lost when Q is used to replace P. It can also be described as a measure which quantifies the extent to which an observed distribution Q approximates a true generating distribution P (McCulloch, 1988). The Kullback-Leibler divergence has found substantial statistical application (e.g., Kullback, 1959; Soofi, 1994), including the context of non-informative prior distributions (Jaynes, 1963). Here, it is used in the context of forming an explicitly informative prior distribution from sets of judgments as follows.

Consider the ith row of Ξ, which is $(\xi_{i,1},\ldots,\xi_{i,k})$. Note that

$$\phi_i = \frac{1}{\sum_{j=1}^{k} \xi_{i,j}^{-1}} (\xi_{i,1}^{-1},\ldots,\xi_{i,k}^{-1})$$

is a distribution obtained from the expert's judgments. As a small example, if

$$\Xi = \begin{bmatrix} 1 & 2 & 3 \\ 1/2 & 1 & 2 \\ 1/3 & 1/2 & 1 \end{bmatrix},$$

then for example $\phi_1 = (6/11,3/11,2/11)$. Clearly we may obtain k distributions, $\phi_1,\ldots,\phi_k$, from the k rows of judgments in Ξ. We are interested in finding the best approximation to the true generating distribution for the k distributions in Ξ. This is equivalent to finding the distribution p' that minimizes the total directed divergence. In the present context,

$$p' = \arg\min_{p\in\mathscr{P}} \sum_{i=1}^{k} D(P,\phi_i).$$

We see that this distribution has the interpretation of being the best estimate of the expert's underlying distribution that generated the judgments in Ξ. It can easily be shown that the elements of p' are the geometric means of ϕ subject to a normalization constant S. Specifically, $p'_j = S^{-1} \prod_{i=1}^{k} \phi_{i,j}^{1/k}$ where $\phi_{i,j}$ is the jth element of ϕ_i and $S = \sum_{j=1}^{k} (\prod_{i=1}^{k} \phi_{i,j}^{1/k})$. The resulting pdf of these probabilities is the prior distribution capturing the beliefs of the expert, which we may call $p(\theta_{ROR})$, where *ROR* indicates the relative odds ratio method. We can see that this pdf is essentially nonparametric and hence will be nonconjugate for any standard likelihood function. Berger (1985) describes the assessment of priors by histograms, and as such can be seen to be a precursor to the method described here. Note that the method will be robust to minor amounts of inconsistency in the judgments. We have already seen that methods that constrain later responses to be consistent with earlier ones are only sensible if the earlier responses are accurate. In practice it may be difficult to determine which one of the elicited values is the source of the inconsistency. The above method can be used to sidestep this difficulty, as conceptually a type of averaging over judgments is used to obtain the

final probabilities. Note that the total directed divergence at p' is an indication of the extent of judgmental inconsistency in Ξ. Specifically, the total directed divergence at p' is

$$D^*(p', \phi) = \sum_{i=1}^{k} \sum_{j=1}^{k} p'_j \log \left(\frac{p'_j}{\phi_{i,j}} \right).$$

When all of the $\xi_{i,j}$ are cardinally consistent with one another (i.e., $\xi_{h,i} \times \xi_{i,j} = \xi_{h,j}$), the total directed divergence $D^*(p', \phi)$ will be zero; otherwise it will be positive. It is natural to consider dividing $D^*(p', \phi)$ by k to produce a scaled measure of inconsistency.

Several advantages accrue to the relative odds ratio prior methodology described above. First, the task is straightforward for the expert. Second, minor inconsistencies in the expert's judgments are tolerated and it is not necessary to "fix" inconsistency by assuming that previous judgments are correct and that subsequent judgments must conform to them. Third, the method is quite general and may be applied for example to real-valued parameters, strictly positive parameters, and parameters which exist on the unit interval. Hence, it is applicable to many types of modeling situations. Fourth, it is easy to use the method to produce automated graphical output that can be used to provide additional feedback to the expert. To be fair, a problem with this method is that tail areas may not be included. For example, there may be a very small but non-zero probability that the parameter takes on values above and below the range of the pdf. Two rejoinders may be offered, however. First, the expert is free to continue adding intervals to the pdf until he or she is feels it has been sufficiently well specified. Second, the use of a conjugate prior, while ensuring that tail areas are represented, does not ensure that tail areas are represented reasonably accurately. For a variety of reasons, the task of obtaining reasonably accurate estimates of very small tail-area probabilities is formidable and perhaps ill-advised (Savage, 1971), despite the fact that failure to represent tail areas accurately may have a substantial impact on the moments of the distribution (Berger, 1985). In using conjugate priors it is probably impossible to determine whether a particular distribution's decay toward the asymptote reasonably reflects the person's actual prior beliefs. Hence, the conjugate approach may be said to have problems representing tail areas also.

4.1. Multiple parameters

One of the features of the method is that is straightforward to extend it to the case of multiple parameters, including that when parameters are not independent of one another. When two parameters are dependent, the marginal distribution of one parameter, Θ_a, would be elicited as described previously. Then, the conditional distribution of the second parameter, Θ_b, would be elicited for each interval of the parameter space of Θ_a. Specifically, the expert would be asked to partition the parameter space of Θ_b into m intervals. The expert would be prompted to consider the case when Θ_a was fixed at its ith interval, i.e., to consider $\theta_{a,i}$. Then, he or she would be asked to provide the relative likelihood or odds of $\theta_{b,1}$ versus $\theta_{b,j}$ given $\theta_{a,i}$. This process would be repeated for all intervals at which point the joint distribution would be obtained. The method can of course be similarly extended to the case of more than 2 variables by the elicitation of the appropriate conditional distributions. As the dimensionality of the problem increases, the time required of the expert will increase. However, since the judgment tasks themselves are simple, the method remains relatively practicable when the number of parameters is not too large.

4.2. Non-uniform intervals

Another extension involves relaxing the notion that the probability density is uniform within each interval. In particular, we may be willing to assume that the rate of change is approximately linear across the interval, which itself is a relatively small portion of Θ. If so, we may obtain the expert's belief by means of the following method. We may split the interval θ_i at its midpoint to obtain θ_i^-, the subinterval on the left, and θ_i^+, the subinterval on the right. The expert may then be asked to provide the relative odds regarding θ_i^- and θ_i^+. Then, the probability densities of both θ_i^- and θ_i^+ can be found from this information. We may then plot a line which has rise equal to the difference in the probability densities of θ_i^+ and θ_i^- and run equal to the length of θ_i. In doing so, the rectangle formerly associated with θ_i will be replaced by a trapezium. Thus, the expert's prior will consist of k trapezia as opposed to k rectangles. Finally, by subdividing an interval into more than two sections, more complex forms could be elicited. For example, if an expert were to feel that the middle portion of an interval was more likely than the endpoints, the interval could be divided into four equal subintervals. Subsequent elicitations would produce a pentagonal shape for the interval. This approach would be particularly useful where it was felt that a broader interval could use adjustment. A distribution suitable for representing these kinds of beliefs appears in the Appendix. Although the use of this approach is straightforward, a referee has pointed out that increasing k may be a sensible alternative, particularly if linearity across the interval seems implausible.

5. Examples

5.1. Inference about a proportion

A number of examples could be proposed due to the generality of the methodology. We present two examples here that, while remaining simple, may be more suggestive of the variety of contexts that may be handled by the method. Consider a situation in which a challenging computer systems certification course is taught. It is of interest to determine the proportion, π, of students who obtain at least satisfactory performance on a preliminary standardized exam that is administered. Expert opinion exists with regard to this parameter and it is of interest to formally incorporate it into inference on π. In particular, the expert has observed that in a given class the proportion of students with satisfactory performance on the exam is typically rather substantial. However, it is regularly observed that for some classes the proportion of students with satisfactory performance is very low. That is, there is a secondary mode at low values of the parameter.

After consultation with the expert, the pdf intervals were specified as follows: $0 \leq \pi < .2$, $.2 \leq \pi < .35$, $.35 \leq \pi < .45$, $.45 \leq \pi < .55$, $.55 \leq \pi < .65$, $.65 \leq \pi < .75$, $.75 \leq \pi < .85$, and $.85 \leq \pi \leq 1$. The expert provided judgments of the relative odds ratios for the intervals, and a judgment matrix was formed from these (see Table 5.1). The boldface entries in Table 5.1 are the judgments supplied by the expert, while the remaining entries are derived from her judgments. Note that the judgment matrix contains the relative odds ratios for a pair of intervals. For example, consider the first interval and the second interval; the expert judged it was three times more likely for π to fall in the first interval than the second.

The preliminary probabilities for each interval were obtained from p' which was computed from the judgment matrix. Here, the probabilities for the intervals were .0738, .0375, .0620, .1179, .1715, .2823, .1717, and .0834 respectively. Trapezia were elicited for the third, fourth, fifth, seventh and eighth intervals. The respective odds of θ^+ to θ^- for these intervals were 11/10, 5/4, 21/20, 20/21 and 5/8. Interval 1 was subdivided equally. For each subinterval a trapezium was elicited such that

Table 1. Judgment matrix for π

Interval for θ_i	*Interval for θ_j*							
	0.0-.20	.20-.35	.35-.45	.45-.55	.55-.65	.65-.75	.75-.85	.85-1.0
0.0-.20	1/1	**3/1**	**2/1**	**1/2**	**1/4**	**1/4**	**1/3**	**1/1**
.20-.35	1/3	1/1	**1/2**	**1/4**	**1/5**	**1/5**	**1/3**	**1/2**
.35-.45	1/2	2/1	1/1	**1/2**	**1/4**	**1/4**	**1/2**	**1/1**
.45-.55	2/1	4/1	2/1	1/1	**1/2**	**1/3**	**1/2**	**2/1**
.55-.65	4/1	5/1	4/1	2/1	1/1	**1/3**	**1/1**	**1/1**
.65-.75	4/1	5/1	4/1	3/1	3/1	1/1	**2/1**	**2/1**
.75-.85	3/1	3/1	2/1	2/1	1/1	1/2	1/1	**3/1**
.85-1.0	1/1	2/1	1/1	1/2	1/1	1/2	1/3	1/1

the second trapezium was a reflection of the first, thus producing a pentagonal form. The odds ratio for the left subinterval was 3/2 (and thus that for the right was 2/3). Like Interval 1, Interval 6 was also subdivided and similar elicitations were performed to generate a pentagonal form. The odds ratio for the left and right subintervals were 4/3 and 3/4. No further elicitations were performed for the second interval. The value of $D^*(p',\phi)/k$ was .05, indicating the inconsistency in judgments was low. Figure 5.1 displays the expert's prior distribution.

Of the 18 students in the course, 12 obtained at least satisfactory performance on the exam. Inference on π was conducted via MCMC methods. The values from an initial burn-in of 1000 iterations were discarded, and posterior inference was based on a subsequent 20,000 iterations. Given the straightforward nature of the model, convergence was almost immediate and the Markov chain showed no evidence of poor mixing. Figure 5.1 also displays the exact posterior density of π as well as a kernel density estimate based on the MCMC output. The posterior mean and standard deviation of π were .667 and .084 respectively, with a 95% credible interval for the parameter ranging from .485 to .819. We may contrast these results with those of a 'naive' binomial analysis utilizing the classical point estimate ($\hat{\pi} = .667$) and 95% confidence interval obtained via $\hat{\pi} \pm 1.96\sqrt{(\hat{\pi}(1-\hat{\pi}))/n}$. Here, the 95% confidence interval ranges from .449 to .884. By contrast, the credible interval is narrower due to the information supplied by the expert's prior, particularly with respect to higher values of p.

5.2. Inference about a regression parameter

Lee (1997, p. 169) presented data on rainfall in York in the months of November and December during the years 1971 through 1980. Interest centered on predicting the amount of rainfall in December (y) from the amount of rainfall in November (x) by means of simple linear regression. The model can be expressed in the form

$$y_i \sim N(\mu_i, \tau)$$

$$\mu_i = \alpha + \beta(x_i - \bar{x}),$$

where $N(\cdot)$ denotes the Normal distribution. In this formulation, μ_i is the linear predictor at the ith observation. Note that y depends on both μ and τ, where τ is the reciprocal of the variance of y and is referred to as the precision of y. In turn, the linear predictor μ is a function of the coefficients α and β as well as the independent variable, x. Lee mentioned in passing having a prior belief that the amounts of rainfall in these two months would be positively correlated, i.e.,
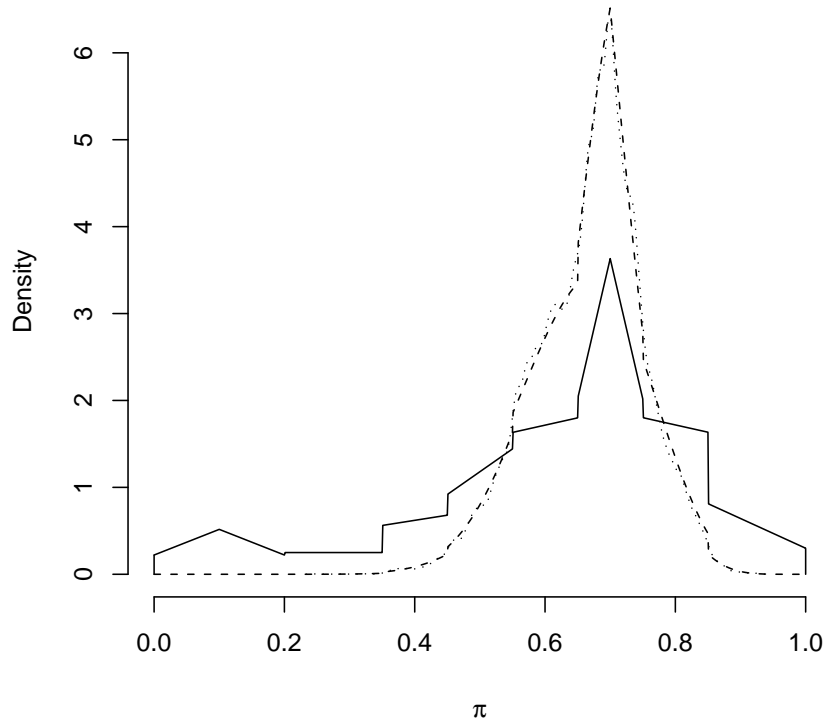
Fig. 1. Prior and posterior for π: the solid line is the exact prior, the dashed line is the exact posterior, and the dotted line is the kernel density estimate of posterior.

that $\beta > 0$. However, no prior beliefs regarding α or τ were mentioned. Here, we re-consider Lee's analysis using vague priors for α and τ, and an expository relative odds ratio prior for β. For illustrative purposes, we suppose that Lee's prior on β ranged from -1 to $+1$ and was skewed toward moderately positive values of β. Values outside the range -1 to $+1$ imply that small changes in x will be accompanied by large changes in y, which seems *a priori* unsupportable given the forecasting challenges faced by meteorologists. Subsequently, the range of β was partitioned into the intervals $-1 \leq \beta < -.5$, $-.5 \leq \beta < -.1$, $-.1 \leq \beta < .1$, $.1 \leq \beta < .3$, $.3 \leq \beta < .5$, $.5 \leq \beta < .7$, and $.7 \leq \beta < 1$. The probabilities for the intervals were subsequently obtained from the relative odds ratio prior methodology, and their values were .068, .101, .138, .221, .285, .125, and .064. A histogram of 20,000 simulated values from the prior appears in Figure 5.2.

For our analysis, posterior inference was based on 20,000 iterations after an initial burn-in of 1000 iterations. Again, convergence was almost immediate. Table 5.2 compares the results from the current analysis with the results obtained by Lee under a non-informative prior for β. We see that
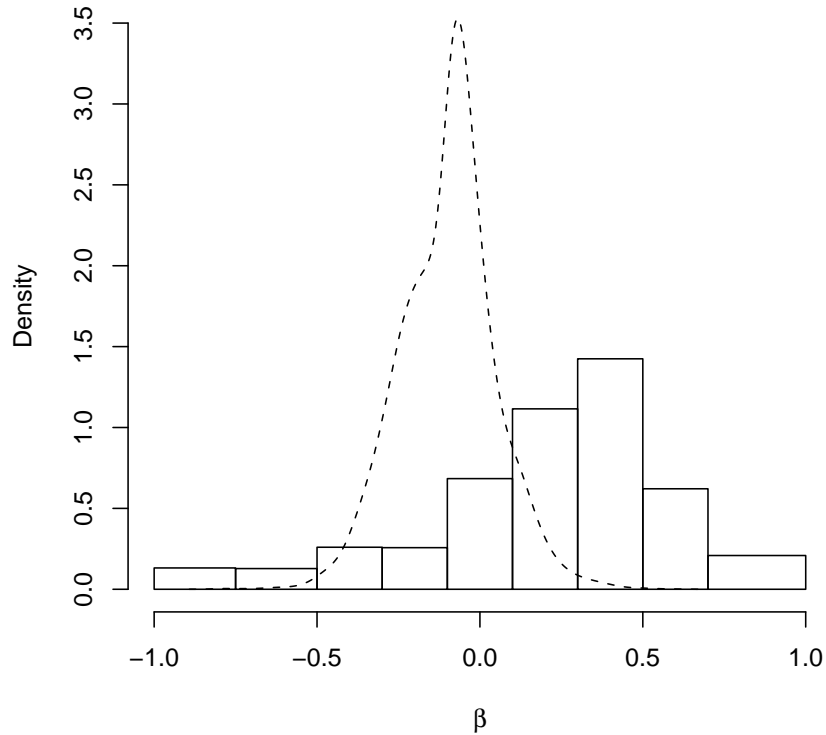
Fig. 2. Prior and posterior for β

under the informative prior the point estimate for β is shifted somewhat in the positive direction. Figure 5.2 shows the plot of the posterior density for β. The effect of the informative prior is also visible in Figure 5.2 where it can be seen that the right tail of the posterior is heavier than the left.

6. Conclusion

Bayesian statistics is predicated on the notion that the posterior distribution is proportional to the likelihood and the prior. In the past decade, MCMC techniques have had a major impact on the practice of Bayesian statistics. For the most part, these techniques have been used to conduct inference in situations where the likelihood function exhibits some complexity. However, these same techniques can be used to conduct inference using non-standard prior distributions, and there are opportunities for further exploration in this area. For example, as pointed out by a referee, in the method described here the geometric mean of the relevant values of ϕ is not the only mechanism by which a representation of the expert's belief may be obtained. Other estimators which would appear to be at least intuitively reasonable can be readily generated (e.g., the arithmetic mean of the

Table 2. Regression parameter estimates

	Informative Prior		Non-informative Prior	
	Point	*95% Credible*	*Point*	*95% Credible*
Variable	*Estimate*	*Interval*	*Estimate*	*Interval*
α	40.83	(30.6, 51.14)	40.81	(30.70, 50.90)
β	-0.098	(-0.386, 0.192)	-0.161	(-0.435, 0.113)
σ^2	266.6	(89.64, 733.7)	192	(77.45, 705.6)

relevant values of ϕ). Here, minimizing the Kullback-Leibler divergence from the true underlying distribution serves as a criterion by which one particular representation may be selected as having desirable properties. We also see that a review of the literature strongly suggests that methods for eliciting a more flexible class of priors can be an important part of a statistician's toolkit. The four principles outlined here should help to guide further research into the development of new methods for eliciting informative priors. The methodology described here is consistent with these principles and offers several advantages to both expert decision makers and statisticians.

Appendix

Here we describe the pentagonal distribution mentioned above. The pentagonal pdf is of the form

$$p(x|a,b,c) = \begin{cases} \dfrac{1}{a+b}\left(a + \dfrac{2bx}{c}\right) & \text{if } x < c \\[2ex] \dfrac{1}{a+b}\left(a + \dfrac{2b(1-x)}{(1-c)}\right) & \text{if } x \geq c \end{cases}$$

with $0 \leq x \leq 1$ where a is associated with the height of the parallel vertical sides of the pentagon, b is associated with the slope of the distribution's "roof" and c is the mode. Clearly $0 \leq c \leq 1$ in order for c to be the mode. The trapezoidal distribution arises as a special case when c is either 0 or 1. The cdf is useful for simulation purposes and can be shown to be

$$F(x|a,b,c) = \begin{cases} \dfrac{x}{a+b}\left(a + \dfrac{bx}{c}\right) & \text{if } x < c, \\[2ex] \dfrac{ax(1-c) - b(c + x(x-2))}{(a+b)(1-c)} & \text{if } x \geq c. \end{cases}$$

In the current context, we need to be able to assign values to the parameters of the distributions. For a trapezoidal distribution, it is easy to show that the appropriate values for a and b are

$$a = \frac{3\,O(\theta_i^-) - O(\theta_i^+)}{2\left(O(\theta_i^-) + O(\theta_i^+)\right)} \qquad \text{and} \qquad b = \frac{O(\theta_i^+) - O(\theta_i^-)}{O(\theta_i^-) + O(\theta_i^+)}.$$

Using as an example the third interval's trapezium from Section 5.1, we find $a = .4524$ and $b = .0476$. For a pentagonal distribution the values for a and b may be determined similarly using the odds ratio for the left subinterval. In the pentagonal distributions considered here, c is always taken to be $1/2$, although this could be changed if desired. Further discussion of a related category of trapezoidal distributions and their generalizations appears in van Dorp and Kotz (2003).

Acknowledgements

References

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. New York: Springer.

Berger, J. O. (2000). Bayesian analysis: A look at today and thoughts of tomorrow. *Journal of the American Statistical Association*, **95**, 1269-1276.

Carlin, B. P., Kass, R. E., Lerch, F. J. and Huguenard, B. R. (1992). Predicting working memory failure: A subjective Bayesian approach to model selection. *Journal of the American Statistical Association*, **87**, 319-327.

Chaloner, K. (1996). Elicitation of prior distributions. In *Bayesian Biostatistics* (D. A. Berry and D. K. Stangl, eds.), pp. 141-156. New York: Marcel Dekker.

Chaloner, K. M. and Duncan, G. T. (1983). Assessment of a beta prior distribution: PM elicitation. *The Statistician*, **32**, 174-180.

Dalal, S. R. and Hall, W. J. (1983). Approximating priors by mixtures of natural conjugate priors. *Journal of the Royal Statistical Society, Series B*, **45**, 278-286.

Diaconis, P. and Ylvisaker, D. (1985). Quantifying prior opinion. In *Bayesian Statistics 2* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.), pp. 133-156. Amsterdam: North-Holland.

Dickey, J. M. and Chen, C.-H. (1985). Direct subjective-probability modelling using ellipsoidal distributions. In *Bayesian Statistics 2* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.), pp. 157-182. Amsterdam: North-Holland.

Garthwaite, P. H. and Dickey, J. M. (1988). Quantifying expert opinion in linear regression problems. *Journal of the Royal Statistical Society, Series B*, **50**, 462-474.

Gelfand, A. E., Mallick, B. K. and Dey, D. K. (1995). Modeling expert opinion arising as a partial probabilistic specification. *Journal of the American Statistical Association*, **90**, 598-604.

Hogarth, R. M. (1975). Cognitive processes and the assessment of subjective probability distributions. *Journal of the American Statistical Association*, **70**, 271-289.

Jaynes, E. T. (1963). Brandeis lectures. Reprinted in *E. T. Jaynes: Papers on Probability, Statistics, and Statistical Physics* (in R. D. Rosenkrantz, ed., 1989), pp. 39-76. Dordrecht: Kluwer.

Kadane, J. B., Dickey, J. M., Winkler, R. L., Smith, W. S. and Peters, S. C. (1980). Interactive elicitation of opinion for a normal linear model. *Journal of the American Statistical Association*, **75**, 845-854.

Kass, R. E. and Wasserman, L. (1996). The selection of prior distributions by formal rules. *Journal of the American Statistical Association*, **91**, 1343-1370.

Kullback, S. (1959). *Information Theory and Statistics*. Mineola, NY: Dover.

Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, **22**, 79-86.

Lee, P. M. (1997). *Bayesian Statistics: An Introduction* (2nd ed.). Arnold: London.

McCulloch, R. E. (1988). Information and the likelihood function in exponential families. *The American Statistician*, **42**, 73-75.

O'Hagan, A. (1998). Eliciting expert beliefs in substantial practical applications. *The Statistician*, **47**, 21-35.

O'Hagan, A., Glennie, E. B. and Beardsall, R. E. (1992). Subjective modeling and Bayes linear estimation in the U.K. water industry. *Applied Statistics*, **41**, 563-577.

Phillips, L. D. and Edwards, W. (1966). Conservatism in a simple probability inference task. *Journal of Experimental Psychology*, **72**, 346-354.

Raiffa, H. A. and Schlaifer, R. S. (1961). *Applied Statistical Decision Theory*. Boston: Graduate School of Business Administration, Harvard University.

Savage, L. J. (1971). Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association*, **66**, 783-801.

Shmueli, G., Minka, T. P., Kadane, J. B., Borle, S. and Boatwright, P. (2001). Using computational and mathematical methods to explore a new distribution: The ν-Poisson. Technical Report 740, Department of Statistics, Carnegie Mellon University. Submitted to JASA.

Simon, H. A. (1957). *Models of Man: Social and Rational: Mathematical Essays on Rational Human Behavior in a Social Setting*. New York: Wiley.

Smith, L. H. (1967). Ranking procedures and subjective probability distributions. *Management Science*, **14**, B-236–B-249.

Soofi, E.S. (1994). Capturing the intangible concept of information. *Journal of the American Statistical Association,* **89**, 1243-1254.

Staël von Holstein, C-A. S. (1971). The effect of learning on the assessment of subjective probability distributions. *Organizational Behavior and Human Performance*, **6**, 304-315.

van Dorp, J. R. and Kotz, S. (2003). Generalized trapezoidal distributions. *Metrika*, **58**, 85-97.

Wallsten, T. S. and Budescu, D. V. (1983). Encoding subjective probabilities: A psychological and psychometric review. *Management Science*, **29**, 151-173.

Winkler, R. L. (1967). The assessment of prior distributions in Bayesian analysis. *Journal of the American Statistical Association*, **67**, 776-800.

Winkler, R. L., Smith, J. E. and Fryback, D. G. (2002). The role of informative priors in zero-numerator problems: Being conservative versus being candid. *The American Statistician*, **56**, 1-4.

Winkler, R. L., Smith, W. S. and Kulkarni, R. B. (1978). Adaptive forecasting models based on predictive distributions. *Management Science*, **24**, 977-986.