The Statistician 32 (1983)

Assessment of a Beta Prior Distribution: PM Elicitation

KATHRYN M. CHALONER and GEORGE T. DUNCAN†

Department of Statistics and School of Urban and Public Affairs, Carnegie–Mellon University, Pittsburgh, PA 15213

Abstract: An interactive computer scheme is described for eliciting from an analyst a beta prior distribution on the parameter π of a binomial distribution. Information on the analyst's beta-binomial predictive distribution is obtained through a questioning and feedback algorithm based on modes.

1 Introduction

Practical implementation of subjective Bayesian methods requires the assessment of a prior distribution. This prior distribution is personal to the substantive expert, the analyst, who has primary concern for the results of the statistical analysis. The role of the statistician is to support the statistical analysis by providing a body of generally applicable technique and, in particular, in providing elicitation procedures for the personal prior distribution of the analyst.

Interactive statistical computing provides the opportunity for new and workable methods of prior distribution assessment. Kadane *et al.* (1980) demonstrated a method for eliciting a natural conjugate prior for the normal regression model. This article proposes interactive elicitation methodology for a natural conjugate beta prior distribution for a binomial parameter π.

We base our elicitation method on the persuasive arguments of Geisser (1980) and Kadane (1980) that the natural elements for statistical inference are characteristics of the predictive distribution of an analyst. That is, our concern is with the distribution of an observable quantity X, unconditional on any unobservable parameters π. This distribution differs from the sampling distribution of X, which is the distribution of X conditional on the value of the parameter π. Indeed, the predictive distribution function of X at any particular value x has a representation as a probability-weighted average over π of the sampling distribution functions. The probability weights are given by the prior distribution. An important fact is that specification of certain characteristics of the predictive distribution determines the prior distribution. This fact allows elicitation of a prior distribution to be carried out through elicitation of a predictive distribution, which is presumably a cognitively easier task since only potentially observable quantities are involved.

† Kathryn M. Chaloner is Assistant Professor, Department of Applied Statistics, University of Minnesota. George T. Duncan is Associate Professor of Statistics, School of Urban and Public Affairs and Department of Statistics, Carnegie–Mellon University. This research was supported in part by grants NSF PFR79-23350 and ONR N00014-77-C-0095. Presented at the 1982 International Conference on Practical Bayesian Statistics, The Institute of Statisticians, St John's College, Cambridge, England, 21–24 July.

Perhaps the first application of these ideas was by Bayes (1764). As interpreted by Stigler (1982), Bayes considered situations in which for any n Bernoulli trials, independent conditionally on the parameter π, the predictive distribution was uniform. That is, $P(X=x)=1/(n+1)$ for $x=0, 1, \ldots, n$. Bayes took this uniform predictive distribution for the observable X as indicative of lack of knowledge about the unobservable π.

Two comments with regard to Bayes's position, essentially made by Stigler (1982), set the stage for our elicitation scheme. First, the role of n is presumably arbitrary. No value of n is of special merit, so the uniformity of the distribution of X would hold consistently over n. Second, knowledge about the data generating process is presumably reflected in a non-uniform distribution for X, one in which a particular event $[X=x^*]$, say, has higher probability than some other event.

Central to our elicitation scheme is (1) over-determination of the parameters of the prior distribution and their reconciliation to give coherency over n and (2) identification of an observable event of highest probability and an assessment of the extent to which the predictive distribution for X departs from uniformity. To give a specific example, consider predictive distributions in the beta-binomial class with $\alpha=3$ and $\beta=2$ for $n=4$ and $n=6$. The two predictive distributions are displayed in Figure 1. Given a certain state of knowledge by the analyst of the nature of the data generating process as specified by $\alpha=3$ and $\beta=2$, coherency requires that if the predictive distribution is as at the top of Figure 1 for $n=4$,

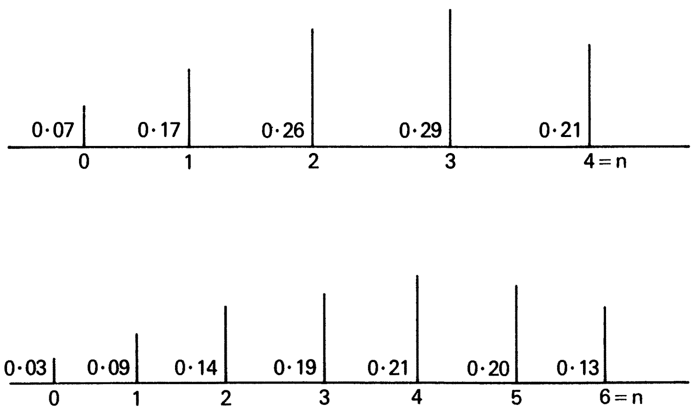

Fig. 1. Beta-binomial density functions for $\alpha=3$, $\beta=2$, $n=4$ and $n=6$.

then it must be as at the bottom of Figure 1 for $n=6$. In actual elicitation these predictive distributions will differ. Since a reasonable theory of decisionmaking argues that a person ought to be coherent, variation in the assessment of α and β over different values of n can be regarded as noise. It is desirable to reconcile the disparities to achieve coherency.

An analyst assessing his or her knowledge about a data generating process naturally tends to anchor (see Tversky and Kahneman, 1981) on the uniform distribution and begin to establish departures from uniformity. Arguably, a first step is to establish which outcomes of X are most likely, that is to specify modes of the predictive distribution of X as points of departure from uniformity. The first question posed according to our method elicits a mode of the predictive distribution. Hence we call the method PM elicitation, for predictive modal elicitation. A refinement is the second set of questions establishing the relative likelihood of the modes to adjacent values of X, thereby assessing the extent of departure from uniformity. Elicitation methods, such as PM elicitation which focus first on modes of X and then on the relative likelihood of the modes to adjacent values of X would seem to be in close accord with what is known empirically about how individuals assess uncertainty.

On the other hand, other methods for elicitation which have intuitive appeal would begin with elicitation of means or medians of the predictive distribution. For the analyst to assess the mean or median of the distribution of X requires, in the framework we have argued reasonably holds, additional cognitive processing.

In the case of the mean, the extent of this processing is so great that elicitation methods which demand it may well produce little more than random noise. To convince oneself of this, one can stare at the two predictive distributions displayed in Figure 1 and attempt to assess the mean without doing any arithmetic (off-limits also is conjuring up the formula for a beta-binomial mean in terms of n, α and β). Clearly in the case of skew distributions the mean is heavily influenced by small probabilities on extreme values of X. Thus the mean has little utility as a base for subjective elicitation in this context.

The median of X is not so bad as the mean of X as a base for subjective elicitation. It is certainly not unnatural for an analyst holding a certain predictive distribution for X to assess a value X_{median} for which even odds bets that X will be above or below X_{median} are equally attractive. But specification of X_{median} has no special advantages in analytically determining the prior parameters α and β, and eliciting it is not cognitively less demanding than eliciting a mode of X. Other quantiles than the median might also be used but, as with the mean, require further cognitive processing. Also extreme quantiles require knowledge of the tails of the predictive distribution which are difficult to quantify. Events of small probability are notoriously hard to specify (Savage, 1971).

2 The PM elicitation algorithm

The PM elicitation algorithm has the analyst choose, for each of various numbers n of trials, a modal number m of successes. For the basic case under consideration of a beta prior distribution which is single peaked the parameters α and β are each greater than one. In this case the number of trials n must be large enough for the count of successes and the count of failures at the specified mode to be at least one. If the analyst believes that, however large n is, either count is zero, then different methods are required.

Specification of a mode of X places some simple analytical constraints on the prior parameters α and β. If m is a mode of a beta-binomial distribution, then the ratio of the probability at m to the probability at the adjacent points $m-1$ and $m+1$ must be no greater than one. The form of the discrete density function for X is

$$f(x;\alpha,\beta)=\frac{\Gamma(n+1)\,\Gamma(\alpha+\beta)\,\Gamma(\alpha+x)\,\Gamma(\beta+n-x)}{\Gamma(n+\alpha+\beta)\,\Gamma(x+1)\,\Gamma(n-x+1)\,\Gamma(\alpha)\,\Gamma(\beta)}\,; \qquad x=0,1,\ldots,n$$

Thus, if we denote the ratios of the probability at $m-1$ and $m+1$ to the probability at m by d_l and d_r, respectively,

$$d_l=\frac{f(m-1)}{f(m)}=\frac{(n-m)\,(m+\alpha)}{(m+1)\,(n-m+\beta-1)}\leqslant 1$$

and

$$d_r=\frac{f(m+1)}{f(m)}=\frac{m(n-m+\beta)}{(n-m+1)\,(m+\alpha-1)}\leqslant 1$$

Once the mode is specified, these two inequalities constrain α and β to lie within a particular cone in the (α, β)-plane. This cone is displayed with solid lines in Figure 2. If, in addition, the two ratios, d_l and d_r, of the probability at the mode to the probability at the adjacent values are specified, then α and β are determined as the intersection of the two

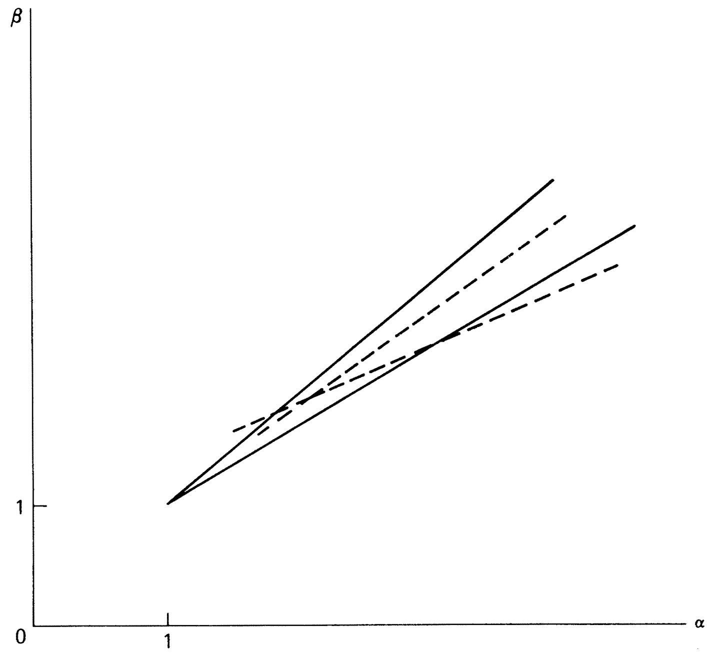

Fig. 2. Geometry of PM elicitation.

dotted lines in Figure 2. The analyst is then given feedback in the form of prediction intervals and allowed to modify the spread of the predictive distribution while leaving the mode fixed. This interactive process is the heart of PM elicitation.

We now describe the details of PM elicitation as it is implemented in an interactive Fortran program using the DEC Gigi colour graphics terminal. The first step is the specification by the analyst of a value of n, the number of trials to be considered. For many applications, $n=20$ appears to be suitable at this first step. Next, the analyst is asked for the value of m, the most likely number of successes in the n trials. This gives the mode of the analyst's predictive distribution. The terminal then displays as feedback to the analyst, a bar chart of binomial distribution probabilities having the same mode and number of trials. Specifically, probabilities for a binomial (n, m/n) distribution are displayed. This feedback gives the analyst an indication of what the analyst's predictive distribution would be if the analyst knew, without uncertainty, the probability π of success. The analyst then has an anchor or benchmark for the introduction of uncertainty. Specifically, the values of d_l and d_r for this binomial distribution, which are calculated and displayed, suggest a lower bound for d_l and d_r for the analyst's beta-binomial predictive distribution.

At the next stage of the elicitation process, the analyst provides the numerical values of d_l and d_r. Each numerical value determines a straight line relationship between α and β. To ensure that these two lines, shown as the dotted lines in Figure 2, intersect within the mode cone, the following inequality must be satisfied:

$$d_l d_r \leqslant \frac{m(n-m)}{(m+1)(n-m+1)}$$

If the analyst gives values violating this inequality, feedback indicates the violation and the analyst is asked to respecify d_l and d_r. Once values satisfying the inequality have been given, the program calculates the values α_1 and β_1 corresponding to the values of d_l and d_r. These values α_1 and β_1 are interpreted as initial "estimates" of the parameters α and β of the beta prior distribution. They also represent initial changes from the uniform distribution case of $(\alpha_0, \beta_0)=(1, 1)$.

At this stage of the algorithm, we take the mode, $(\alpha-1)/(\alpha+\beta-2)$, of the beta prior distribution to have been elicited. The spread, however, may not yet have been adequately elicited. The program now keeps the mode of the beta distribution fixed and enters an interactive stage to elicit the proper spread. Note that fixing the mode of the beta distribution fixes the mode of the beta-binomial distribution, whereas fixing the mean of the beta distribution does not fix the mode of the beta-binomial distribution.

The program takes α_1 and β_1 and provides graphical feedback illustrating the smallest prediction interval of the beta-binomial distribution with at least 50 per cent probability content. Also calculated and displayed are the exact predictive probabilities in this interval and their sum, the probability content of the interval. The analyst is then asked if this predictive interval is (a) too short, (b) too long, or (c) the right length. If the analyst answers (c), the interval is consistent with the analyst's beliefs about the appropriate length, the elicitation for this value of n is complete. He is then presented with the values of α_1 and β_1.

If the 50 per cent prediction interval is too short, the analyst is given a "longer" prediction interval calculated by decreasing the values of α_1 and β_1 to α_2 and β_2. The 50 per cent prediction interval implied by the new values α_2 and β_2 may not literally be longer, but it will have less total probability on the points in the earlier prediction interval. On the other hand, if the 50 per cent prediction interval is too long, the analyst is given a "shorter" prediction interval calculated by increasing the values of α_1 and β_1 to α_2 and β_2. The analyst is then given the 50 per cent prediction interval implied by α_2 and β_2 and asked again whether it is of the right length. The process iterates until the analyst is satisfied with the length of the interval. The analyst is then presented with the values of α and β at the final step.

In this stage of adjusting the length of the prediction interval, the program chooses new values of α and β according to a process similar to binary search. Specifically, all choices of α_i and β_i are restricted to the ray from (1, 1) through (α_1, β_1). This restriction keeps the mode of the elicited beta prior distribution constant. If the analyst initially says the prediction interval is too long, the values α_1-1 and β_1-1 are doubled to obtain α_2-1 and β_2-1. This doubling of each successive pair is continued until the analyst says the prediction interval is too short. This gives an upper bound on α and β. Let the first time the analyst says the interval is too short be denoted by l. Then (α, β) is restricted to lie within the interval joining $(\alpha_{l-1}, \beta_{l-1})$ to (α_l, β_l). The value of $(\alpha_{l+1}, \beta_{l+1})$ is given by $\alpha_{l+1}=\frac{1}{2}(\alpha_l+\alpha_{l-1})$ and $\beta_{l+1}=\frac{1}{2}(\beta_l+\beta_{l-1})$. This process is continued as long as the analyst is dissatisfied with the length of the prediction interval. Specifically, at the $(l+j)$th iteration, if the analyst says the interval is too short, then $\alpha_{l+j+1}=\alpha_{l+j}-2^{-j-1}(\alpha_l-\alpha_{l-1})$ and $\beta_{l+j+1}=\beta_{l+j}-2^{-j-1}(\beta_l-\beta_{l-1})$. Similarly, if at this stage the analyst says the interval is too long then $\alpha_{l+j+1}=\alpha_{l+j}+2^{-j-1}(\alpha_l-\alpha_{l-1})$ and $\beta_{l+j+1}=\beta_{l+j}+2^{-j-1}(\beta_l-\beta_{l-1})$. Thus the iterations will give (α, β)-values which converge to finite values. The analyst has specified a particular beta-binomial distribution as the analyst's predictive distribution. If the analyst never specifies the interval is too short the analyst is indicating certainty about the value of π so that the binomial distribution adequately represents the analyst's predictive distribution.

At this point the analyst may end the elicitation process and use the beta prior distribution of π which has been determined. Or, as seems preferable, the analyst may repeat the PM elicitation scheme using a different value of n, the number of trials. If the analyst is coherent, in the full sense with no elicitation errors, the elicited values of (α, β) for the different values of n will all be the same. In practice, these values will differ and the analyst will have several "estimates", say, $(\alpha^1, \beta^1), (\alpha^2, \beta^2), \ldots, (\alpha^k, \beta^k)$ one for each of the k values of n used. There are three evident approaches to reconciling these values. One approach would be to ask the analyst to reconcile them in any way the analyst may choose, which might result in the selection of one the analyst finds most appealing, for whatever reason. A second approach would be to mechanically just choose some center, say the mean or median, of α^i and β^i, $i=1, \ldots, k$. Either of these two approaches seems perfectly adequate when the values of (α^i, β^i) are not too disparate. A third approach, is one in which the statistician (or the analyst) uses the elicited values as data and estimates in a Bayesian fashion the analyst's postulated underlying values of α and β. This formal Bayes approach would require the assumption that the elicitation method prompted responses with errors associated with them, the specification of a prior distribution for α and β and, perhaps less easily, the joint sampling distribution of the elicited values. As one aspect of the difficulty of the latter task, we note that the elicited values are presumably not independent, even conditionally on α and β.

In any case the overspecification of estimates of α and β seems to be an essential ingredient of a good elicitation scheme. In particular, it allows the statistician and the analyst to examine aspects of the analyst's elicited responses which have led to incoherency. It also permits a check on whether the underlying beta-binomial model is adequate to represent the analyst's beliefs.

3 Discussion

There has been some discussion in the statistical literature of how to elicit a beta prior distribution; see, for example, Bunn (1978, 1979). Proposed methods have concentrated on asking directly for the mean and variance of the beta prior distribution or have asked for the mean or variance of the posterior distribution after imaginary future results. We suggest that PM elicitation is an improvement for several reasons.

First, the method utilizes the current facilities of a computer, providing the interactive capability of structured questions, instant feedback, and graphical feedback. The instant graphical feedback appears in practice to be an especially helpful aspect of PM elicitation. Second, PM elicitation deals with the predictive distribution, the distribution of observable quantities, which we regard as more basic than the underlying prior distribution. Third, in the particular context of the beta-binomial distribution PM elicitation is appealing in that it deals with modes rather than with means or medians. Modes appear to be easier to process cognitively.

The beta-binomial distribution is a special case of the Dirichlet-multinomial distribution. We have expanded and adapted PM elicitation to this case (Chaloner and Duncan, 1982). A unimodality property of Dirichlet-multinomial distributions is proved and utilized in this paper. The coherency of analysts over different values of n has also been investigated empirically and we will report findings in a later paper.

In conclusion, we note that the major strength of Bayesian inference is that it optimally combines the expertise of the analyst with the information of the data. In effectively using this strength of introducing the expertise of the analyst, practical Bayesian statistics requires good methods for the elicitation of expert opinion. We view PM elicitation as a start on this task for inference regarding count data.

References

Bayes, Thomas (1764) An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, **53**, 370–418. Reprinted in Barnard (1958). *Biometrika*, **45**, 293–315.
Bunn, D. W. (1978). The estimation of a Dirichlet prior density. *Omega*, **6** (4), 371–3.

Bunn, D. W. (1979) Estimation of subjective probability distributions in forecasting and decision making. *Technological Forecasting and Social Change*, **14** (1), 205–16.
Chaloner, K. M. and Duncan, G. T. (1982). *Interactive Elicitation of a Dirichlet-multinomial Distribution in the Spirit of Bayes.* Technical Report 208, Carnegie–Mellon University.
Geisser, S. (1980). *The Estimation of Distribution Functions and the Prediction of Future Observations*, pp. 193–208. Academia Sinica, Taiwan.
Kadane, J. B. (1980). Predictive and structural methods for eliciting prior distributions. In *Bayesian Analysis in Econometrics and Statistics.* North-Holland, Amsterdam.
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**, 845–54.
Savage, L. J. (1971). Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association*, **66**, 565–614.
Stigler, Stephen M. Thomas Bayes's Bayesian inference. Manuscript. Department of Statistics, University of Chicago.
Tversky, Amos and Kahneman, Daniel (1981). The framing of decisions and the psychology of choice. *Science*, **211** (30 January), 453–8.