Biometrika (1990), 77, 4, pp. 901–4
Printed in Great Britain

Prior distributions for the bivariate binomial

By NICK POLSON AND LARRY WASSERMAN
Department of Statistics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, U.S.A.

SUMMARY

We derive prior distributions for the bivariate binomial model using Bernardo's (1979) method. These priors are compared to the Jeffreys prior and to a prior proposed by Crowder & Sweeting (1989). The priors possess desirable symmetry properties since we allow them to depend on the parameter of interest.

Some key words: Jeffreys prior; Multivariate binomial; Nuisance parameter; Reference prior.

1. INTRODUCTION

Crowder & Sweeting (1989) considered the following model. Each of m spores has a probability p of germinating. Of the r spores that germinate, each has a probability q of bending in a particular direction. Let s be the number that bend in the specified direction. The probability model for these data is

$$f(r, s \mid p, q, m) = \binom{m}{r} p^r (1-p)^{m-r} \binom{r}{s} q^s (1-q)^{r-s}$$

for $r = 1, \ldots, m$ and $s = 1, \ldots, r$. This is referred to as the bivariate binomial.

The Fisher information matrix for this model is

$$I(p, q) = m \operatorname{diag}[\{p(1-p)\}^{-1}, p\{q(1-q)\}^{-1}].$$

Thus, the Jeffreys prior is

$$\pi_J(p, q) \propto (1-p)^{-\frac{1}{2}} q^{-\frac{1}{2}} (1-q)^{-\frac{1}{2}}.$$

This differs from that of Crowder & Sweeting (1989, eqn (3·2)). Note that the Jeffreys prior is proper.

Crowder & Sweeting are dissatisfied with the Jeffreys prior for this model because it is asymmetric in p and $1-p$. They also claim that the Bernardo prior (Bernardo, 1979) exhibits the same undesirable behaviour. They argue in favour of the improper prior

$$\pi_{CS}(p, q) \propto p^{-1}(1-p)^{-1} q^{-1} (1-q)^{-1}.$$

Their argument is based on invariance considerations, together with the assumptions that p and q should be independent and that θ and ϕ should be independent, where $\theta = pq$ and $\phi = p(1-q)/(1-pq)$.

Note that $p = \theta + \phi - \theta\phi$ and $q = \theta/(\theta + \phi - \theta\phi)$. Thus, the Fisher information matrix in the θ, ϕ parameterization is

$$I(\theta, \phi) = m \operatorname{diag}[\{\theta(1-\theta)\}^{-1}, (1-\theta)\{\phi(1-\phi)\}^{-1}].$$

In the present paper, we derive priors for this model using methods originally proposed by Bernardo (1979). For more recent details on these methods, see an unpublished report by J. O. Berger and J. M. Bernardo. The priors are essentially based on a stepwise Jeffreys approach. We show that these priors are proper, and thus lead to coherent inferences. Following Bernardo (1979), the construction of the prior will depend on the parameter of interest. Thus, we derive a family of priors, one for each parameter of interest. In § 3 we point out that each of these priors possesses a natural symmetry and independence property for the parameter of interest. The idea that a reasonable prior should depend on the choice of parameter of interest is implicit in Efron's (1973) discussion. In this problem there are four natural choices for the parameter of interest, namely, p, q, θ and ϕ.

In § 2, we present the four priors. In § 3, we discuss our results and the results of Crowder & Sweeting. The details of the derivations are in the Appendix.

2. THE PRIORS

We begin by reviewing Bernardo's (1979) method for finding prior distributions when little prior information is available.

Suppose the model is $\{f(x \mid \omega), \omega \in \Omega\}$, where ω is the unknown parameter and ω is decomposable as $\omega = (\psi, \lambda)$ where ψ is the parameter of interest and λ is a nuisance parameter. Let $J(\lambda \mid \psi)$ denote the normalized Jeffreys prior for λ with ψ fixed (Jeffreys, 1961, § 3.10). The marginal model for x conditional on ψ only, is

$$f(x \mid \psi) = \int f(x \mid \psi, \lambda) J(\lambda \mid \psi)\, d\lambda.$$

The next step is to find the Jeffreys prior $J(\psi)$ using the marginal model $f(x \mid \psi)$. The prior advocated by Bernardo is defined to be $\pi_\psi(\psi, \lambda) \propto J(\psi) J(\lambda \mid \psi)$. Bernardo (1979) calls this a reference prior. But this term was used in a different context by Box & Tiao (1973). If the entire parameter ω is considered to be the parameter of interest, then Bernardo's method reduces to the Jeffreys prior.

There are two aspects to this method of constructing priors. The first is the use of Jeffreys prior at each stage of the process. Bernardo's (1979) justification for this is that, under suitable regularity conditions, Jeffreys prior maximizes the missing asymptotic Shannon information for the parameter. This might be thought of as a justification of Jeffreys prior as being noninformative. We prefer to think of it simply as a way of modelling the notion that an experiment is carried out because one expects to learn about the parameters.

The second aspect of Bernardo's approach is the stepwise nature of the prior construction. He presents no formal justification for proceeding in this way though he notes that his method produces reasonable priors in a wide class of problems where Jeffreys's prior is not successful.

Note that the prior depends on which parameter is chosen as the parameter of interest. However, the prior is invariant under 1-1 transformations of either the parameter of interest or the nuisance parameter. Thus the invariance possessed by the Jeffreys prior is partially obeyed by the stepwise prior. The dependence of the prior on the choice of parameter of interest seems reasonable to us. The fact that a particular parameter is considered to be the parameter of interest is relevant information and should affect the choice of prior. Indeed, the most important innovation of Bernardo (1979) is the stepwise approach to the construction of such priors.

In the Appendix, we derive the four priors π_p, π_q, π_θ, π_ϕ, depending on whether the parameter of interest is p, q, θ or ϕ. The priors are

$$\pi_p(p, q) = \pi_q(p, q) \propto p^{-\frac{1}{2}}(1-p)^{-\frac{1}{2}} q^{-\frac{1}{2}}(1-q)^{-\frac{1}{2}},$$

$$\pi_\theta(\theta, \phi) = \pi_\phi(\theta, \phi) \propto \theta^{-\frac{1}{2}}(1-\theta)^{-\frac{1}{2}} \phi^{-\frac{1}{2}}(1-\phi)^{-\frac{1}{2}}.$$

The priors π_p and π_q are symmetric in $(p, 1-p)$ and $(q, 1-q)$ and they make p and q independent. Also, the priors π_θ and π_ϕ are symmetric in $(\theta, 1-\theta)$ and $(\phi, 1-\phi)$ and they make θ and ϕ independent. We discuss this issue in § 3. It is of interest to note that π_p is the product of two Jeffreys priors for the usual binomial model while the Crowder-Sweeting prior is the product of two improper Haldane (1948) priors.

3. Discussion

Crowder & Sweeting argued that independence of p and q together with independence of θ and ϕ imply that the prior must be of the form $p^{\alpha+\gamma-1}(1-p)^{\beta-1}q^{\alpha-1}(1-q)^{\gamma-1}$ for some α, β, γ. Furthermore, they argued that the prior should be symmetric in $(p, 1-p)$, in $(q, 1-q)$, in $(\phi, 1-\phi)$ and in $(\theta, 1-\theta)$. In the above equation, this implies that $\alpha=\beta=\gamma=0$ which leads to π_{CS}. They point out that the Jeffreys prior violates this symmetry condition and say 'The same applies to the reference prior of Bernardo (1979)'. We have shown that this latter claim must be interpreted cautiously. Indeed, π_p, π_q, π_θ and π_ϕ do satisfy the coveted symmetry and independence requirements for the parameter of interest. For example, if p were the parameter of interest, we are recommending the prior π_p. This prior is symmetric in $(p, 1-p)$ and $(q, 1-q)$ and makes p and q independent. None of these priors satisfies all the symmetry and independence requirements simultaneously, but each is symmetric for the parameter of interest. If we transform π_p into the θ, ϕ parameterization, the absolute value of the Jacobian is $(1-\theta)/(\theta+\phi-\theta\phi)$ and we get

$$\pi_p(\theta, \phi) \propto \phi^{-\frac{1}{2}}(1-\phi)^{-\frac{1}{2}}\theta^{-\frac{1}{2}}(\theta+\phi-\theta\phi)^{-\frac{1}{2}}$$

which is not symmetric in θ and $1-\theta$. We emphasize, however, that this is not the appropriate prior when θ is the parameter of interest. In this instance, we must use $\pi_\theta(\theta, \phi)$ which has exactly the same functional form as $\pi_p(p, q)$.

We do not advocate the unfettered use of these priors. In their report J. O. Berger and J. M. Bernardo say: '... the danger in using a noninformative prior is that it might contain hidden features that have a dramatic (and unrecognized) effect on the answers'. Indeed, we believe it is essential to investigate the behaviour of the various priors derived in § 2. A sensitivity analysis of the posteriors that result from these priors is essential for a complete analysis.

In conclusion, the priors based on the methods of Bernardo (1979) satisfy a restricted form of the desired independence and symmetry properties. Specifically, the prior corresponding to a particular parameter of interest has the appropriate symmetry and independence properties for that parameter. Clearly, the results here generalize for the multivariate binomial model. In a future paper, we hope to present a formal justification for Bernardo's stepwise approach to the construction of priors.

Acknowledgements

The authors are grateful to Rob Kass, Isabella Verdinelli, the editor and two referees for helpful comments. The second author would like to acknowledge the Natural Sciences and Engineering Research Council of Canada for financial support.

Appendix

Derivation of priors

Here we derive the stepwise priors. We give details only for π_p and π_q. The derivations for π_θ and π_ϕ are similar.

Case 1: *Parameter of interest is p.* First, note that $J(q \mid p) \propto q^{-\frac{1}{2}}(1-q)^{-\frac{1}{2}}$. The marginal model is

$$f(r, s \mid p, m) = \int_0^1 f(r, s \mid p, q, m) J(q \mid p)\, dq = \left\{\binom{m}{r}\binom{r}{s} p^r(1-p)^{m-r} B(s+\tfrac{1}{2}, r-s+\tfrac{1}{2})\right\} \Big/ B(\tfrac{1}{2}, \tfrac{1}{2}),$$

where $B(a, b)$ is the beta function. Let l be the log likelihood for the marginal model. Now,

$$E\left(-\frac{\partial^2 l}{\partial p^2}\right) = \frac{E(r \mid p)}{p^2} + \frac{m - E(r \mid p)}{(1-p)^2}.$$

But

$$E(r \mid p) = \int_0^1 E(r \mid p, q) J(q \mid p)\, dq = mp \int_0^1 J(q \mid p)\, dq = mp.$$

Therefore,

$$J(p) \propto p^{-\frac{1}{2}}(1-p)^{-\frac{1}{2}}, \quad \pi_p(p, q) \propto p^{-\frac{1}{2}}(1-p)^{-\frac{1}{2}} q^{-\frac{1}{2}}(1-q)^{-\frac{1}{2}}.$$

Case 2: *Parameter of interest is q*. In this case, $J(p\,|\,q) \propto p^{-\frac{1}{2}}(1-p)^{-\frac{1}{2}}$ so that

$$f(r, s\,|\,q, m) = \left\{ \binom{m}{r}\binom{r}{s} q^r(1-q)^{r-s} B(r+\tfrac{1}{2}, m-r+\tfrac{1}{2}) \right\} \Big/ B(\tfrac{1}{2}, \tfrac{1}{2}),$$

where $0 \leqslant s \leqslant r \leqslant m$. Then,

$$E\left(-\frac{\partial^2 l}{\partial q^2}\right) = \frac{E(s\,|\,q)}{q^2} + \frac{E(r\,|\,q) - E(s\,|\,q)}{(1-q)^2}.$$

But,

$$E(s\,|\,q) = \int_0^1 E(s\,|\,p, q) J(p\,|\,q)\, dp = qmc,$$

where c is a constant not depending on q. Similarly, $E(r\,|\,q) = mc$. Substituting these in the expression for the expected value of the second derivative we see that, $J(q) \propto q^{-\frac{1}{2}}(1-q)^{-\frac{1}{2}}$ so that

$$\pi_q(p, q) = \pi_p(p, q) \propto p^{-\frac{1}{2}}(1-p)^{-\frac{1}{2}} q^{-\frac{1}{2}}(1-q)^{-\frac{1}{2}}.$$

REFERENCES

BERNARDO, J. M. (1979). Reference posterior distributions for Bayesian inference (with discussion). *J. R. Statist. Soc.* B **41**, 113-48.

BOX, G. E. P. & TIAO, G. C. (1973). *Bayesian Inference in Statistical Analysis*. Reading, Mass: Addison-Wesley.

CROWDER, M. & SWEETING, T. (1989). Bayesian inference for a bivariate binomial. *Biometrika* **76**, 599-604.

EFRON, B. (1973). Discussion of paper by A. P. Dawid, M. Stone and J. Zidek. *J. R. Statist. Soc.* B **35**, 35.

HALDANE, J. B. S. (1948). The precision of observed values of small frequencies. *Biometrika* **35**, 297-303.

JEFFREYS, H. (1961). *Theory of Probability*, 3rd ed. Oxford: Clarendon Press.

[*Received November* 1989. *Revised May* 1990]