A Note on Bayes Factors for Log-Linear Contingency Table Models with Vague Prior Information

A. E. Raftery

Journal of the Royal Statistical Society. Series B (Methodological), Vol. 48, No. 2 (1986), 249-250.

Stable URL:
http://links.jstor.org/sici?sici=0035-9246%281986%2948%3A2%3C249%3AANOBFF%3E2.0.CO%3B2-2

J. R. Statist. Soc. B (1986)
48, *No.* 2, *pp.* 249–250

A Note on Bayes Factors for Log-linear Contingency Table Models with Vague Prior Information

By A. E. RAFTERY†

Trinity College, Dublin, Ireland
[Received July 1985. Revised January 1986]

SUMMARY

The approximate Bayes factor, B_{01}, for log-linear contingency table models proposed by Spiegelhalter and Smith (1982) is indeterminate if any of the cell frequencies is zero. It is noted that use of a standard Jeffreys prior overcomes this difficulty. It is pointed out that $-2 \log B_{01}$ is approximately equivalent to Schwarz's (1978) model selection criterion in large samples.

Keywords: BAYESIAN INFORMATION CRITERION; IMAGINARY OBSERVATIONS; JEFFREYS PRIOR

Spiegelhalter and Smith (1982)—hereafter SS—proposed an approximate method for calculating the Bayes factor for a log-linear model M_0 for a contingency table against the saturated model M_1, with vague prior information. We adopt their notation and denote by *SS*(n) equation (n) of SS.

Suppose $x_1, \ldots, x_k$ have a multinomial distribution with parameters $\phi_1, \ldots, \phi_k$ where $\phi_i \geqslant 0$, $\Sigma\phi_i = 1$. We write $y^T = (\log x_1, \ldots, \log x_k)$, $\theta_1^T = (\log \phi_1, \ldots, \log \phi_k)$, and $Y = \text{diag}\{x_1, \ldots, x_k\}$. Then if M_1 is the saturated model, and M_0 is the nested, log-linear, model defined by setting the contrasts $C\theta_1 = 0$, where C is an $s \times k$ matrix with rank s and rows summing to zero, the approximate Bayes factor B_{01} for M_0 against M_1 is given by SS (32).

This, however, is indeterminate if any of the cell frequencies in the table is zero. This is because of the use by SS of a prior density proportional to $(\Pi\phi_i)^{-1}$. If, however, we use, instead, the standard Jeffreys prior density proportional to $(\Pi\phi_i)^{-1/2}$, the problem no longer arises. Then, by the arguments of SS and Lindley (1964), the resulting Bayes factor is still given by SS (32), with x_i replaced by $x_i + \frac{1}{2}$ in the definitions of y and Y ($i = 1, \ldots, k$), and SS (33) replaced by $c_1^{-1} = (\frac{2}{3})^{s/2} |CC^T|^{1/2}$.

If this solution is adopted, the prior is proper, and so, in principle, the problem of assigning an arbitrary multiplicative constant, for which the SS approach was primarily devised, need not arise. One could, in theory, simply apply Bayes; theorem directly and so obtain the Bayes factor exactly. However, in practice, this is difficult to do, and I know of no general solution to the problem. Even for simpler, more specific, contingency table and related models, such as those of independence or equiprobability, finding the Bayes factor exactly is not too easy; see, for example, Crook and Good (1982), Altham (1971), Günel and Dickey (1974), Günel (1982), Broniatowski (1981), and references therein.

The second purpose of this note is to point out that, conditionally on M_0,

$$-2 \log B_{01} \overset{P}{\sim} \chi^2 - s \log n = BIC \tag{1}$$

where $n = \Sigma x_i$, $\overset{P}{\sim}$ denotes asymptotic equivalence in probability as $n \to \infty$, and χ^2 is the standard likelihood-ratio goodness-of-fit statistic. This is equivalent to the Schwarz (1978) model selection criterion for independent observations, extended to Markov chains by Katz

† *Address for correspondence*: Department of Statistics GN-22, University of Washington, Seattle, WA 98195, USA.

 0035–9246/86/48249 $2.00

(1981). The use of the AIC criterion for model selection in contingency tables, defined by replacing $\log n$ by 2 in (1), is discussed by Sakamoto (1984).

To show (1) we remark that, by SS (32),

$$-2 \log B_{01}/BIC = 1 + (Z^2 - \chi^2)/BIC + (\log \Gamma_n/BIC) \tag{2}$$

where $Z^2 = y^T C^T (CY^{-1}C^T)^{-1} Cy$ and $\Gamma_n = n^s c_1^2 | CY^{-1}C^T |$. (2) then follows from (1) by noting that $Z^2 - \chi^2 \to 0$ in probability as pointed out by SS (the proof is similar to that of Lemma 14.9-1 in Bishop, Fienberg and Holland (1975)), $\Gamma_n \to c_1^2 | C\Phi^{-1}C^T |$ in probability by the weak law of large numbers, where $\Phi = \text{diag}\{\phi_1, \ldots, \phi_k\}$, and $BIC \to -\infty$ in probability.

Our statement of (1) is conditional on M_0. Conditionally on M_1, $BIC \to \infty$ in probability by Jensen's inequality and the weak law of large numbers, so that BIC is a consistent criterion in the contingency table case.

ACKNOWLEDGEMENT

I would like to thank the referee for very helpful comments on an earlier version of this paper.

REFERENCES

Altham, P. M. E. (1971) Exact Bayesian analysis of an intraclass 2 × 2 table. *Biometrika*, **58**, 679–680.

Bishop, Y. M. M., Fienberg, S. E. and Holland, P. W. (1975) *Discrete Multivariate Analysis: Theory and Practice.* Cambridge, Mass: MIT Press.

Broniatowski, M. (1981) Méthode Bayesienne de test entre hypothèses disjointes—Application aux chaînes Markoviennes et à la loi binomiale négative. *Publ. Inst. Statist. Paris*, **26**, 13–26.

Crook, J. F. and Good, I. J. (1982) The powers and strengths of tests for multinomials and contingency tables. *J. Amer. Statist. Ass.*, **77**, 793–802.

Günel, E. and Dickey, J. M. (1974) Bayes factors for independence in contingency tables. *Biometrika*, **61**, 545–557.

Günel, E. (1982) Bayes factors in three-way contingency tables. *Commun. Statist—Theor. Meth.*, **11**, 911–931.

Katz, R. W. (1981) On some criteria for estimating the order of a Markov chain. *Technometrics*, **23**, 243–249.

Lindley, D. V. (1964) The Bayesian analysis of contingency tables. *Ann. Math. Statist.*, **35**, 1622–1643.

Sakamoto, Y. (1984) Efficient use of Akaike's information criterion on high dimensional contingency table analysis. *Metron*, **40**, 257–275.

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

Spiegelhalter, D. J. and Smith, A. F. M. (1982) Bayes factors for linear and log-linear models with vague prior information. *J. R. Statist Soc.* B, **44**, 377–387.