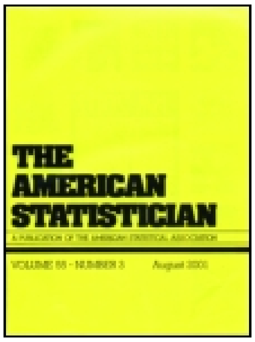

The American Statistician

ISSN: 0003-1305 (Print) 1537-2731 (Online) Journal homepage: www.tandfonline.com/journals/utas20

A Comparison of Bayes–Laplace, Jeffreys, and Other Priors

The Case of Zero Events

Frank Tuyl, Richard Gerlach & Kerrie Mengersen

To cite this article: Frank Tuyl, Richard Gerlach & Kerrie Mengersen (2008) A Comparison of Bayes–Laplace, Jeffreys, and Other Priors, The American Statistician, 62:1, 40-44, DOI: 10.1198/000313008X267839

To link to this article: https://doi.org/10.1198/000313008X267839

A Comparison of Bayes–Laplace, Jeffreys, and Other Priors: The Case of Zero Events

Frank TUYL, Richard GERLACH, and Kerrie MENGERSEN

Beta distributions with both parameters equal to 0, $\frac{1}{2}$, or 1 are the usual choices for "noninformative" priors for Bayesian estimation of the binomial parameter. However, as illustrated by two examples from the Bayesian literature, care needs to be taken with parameter values below 1, both for noninformative and informative priors, as such priors concentrate their mass close to 0 and/or 1 and can suppress the importance of the observed data. These examples concern the case of no successes (or failures) and illustrate the informativeness of the Jeffreys prior usually recommended as the "consensus prior." In particular, the second example suggests that when the binomial parameter is known to be very small, an informative prior from the beta$(1, b)$ family ($b > 1$) seems appropriate, while a beta(a, b) with $a < 1$ can be *too* informative. It is thus argued that sensitivity analysis of an informative prior should be based on a consensus posterior corresponding to the Bayes–Laplace prior rather than the Jeffreys prior.

KEY WORDS: Bayesian inference; Binomial distribution; Noninformative priors; Prior families.

1. INTRODUCTION

The Bayesian approach to estimation is to determine a posterior distribution for the unknown parameter(s), based on a prior distribution and the likelihood function. So-called conjugate priors lead to posteriors that have the same parametric form. The conjugate prior for the probability of success θ of the binomial distribution is the beta(a, b) distribution. After observing x successes in n trials, θ has the posterior distribution beta$(x + a, n - x + b)$.

Typically one of two questions is relevant in Bayesian prior formulation: how to define a prior distribution if we know "nothing," or how to match one to some available prior knowledge. These questions are not usually important when the sample size is large, as the data dominates the prior and the posterior tends to normality under certain conditions.

This work was largely undertaken while all authors were members of the Australian Research Council Centre of Excellence in Dynamic Systems and Control, University of Newcastle, Australia, where Frank Tuyl, Statistician at Hunter New England Population Health, Newcastle, Australia, is Conjoint Lecturer (E-mail: *frank.tuyl@newcastle.edu.au*). Richard Gerlach is Senior Lecturer in Econometrics and Business Statistics, Faculty of Economics and Business, University of Sydney, Australia. Kerrie Mengersen is Professor of Statistics, School of Mathematical Sciences, Queensland University of Technology, Australia. The authors thank editors and referees for many helpful suggestions.

This asymptotic result applies to the binomial parameter too, except when x is close to 0 or n. Quite clearly, the case $x = 0$ (or $x = n$) is the most extreme test. Several authors have reported on it (Louis 1981; Jovanovic and Levy 1997; Winkler, Smith, and Fryback 2002), and we consider it of significant practical interest. For example, Berry (1996, p. 227) analyzed a legal case based on $x = n = 3$ and prior expert knowledge. Here θ was the "true" proportion of glass breakages with nickel sulfide being present in the glass, which points to faulty manufacture. In our own consulting experience (industrial, financial, medical) we have also seen extreme outcomes relatively often.

This article shows that in the context of $x = 0$ the noninformative beta$(\frac{1}{2}, \frac{1}{2})$ or Jeffreys prior can be too informative, while some choices of informative priors (beta(a, b), $a < 1$) may become excessively informative, in each case concentrating too much probability mass close to 0, thus suppressing the importance of the observed data. We revisit two examples in order to illustrate the informativeness of the priors chosen and to suggest alternative improved solutions.

Throughout this article we argue that the beta(1,1) or Bayes–Laplace (B–L) prior is a good "consensus" prior (Bernardo 2005) for all x. In their introduction to Bayesian reference analysis, which leads to the beta$(\frac{1}{2}, \frac{1}{2})$ prior, Bernardo and Ramon (1998, p. 128) stated, "Proper or improper, what must be required from nonsubjective priors is that, for any dataset, they lead to sensible, data-dominated, *posterior* distributions." Such consensus priors would thus be suitable as standards for scientific communication. It appears, however, that for the binomial parameter the reference/Jeffreys prior fails this requirement when $x = 0$.

Additionally, Bernardo and Ramon (1998) discussed the importance of such priors for the purpose of sensitivity analysis, that is, to check the sensitivity of results to the choice of an informative prior. A natural reference point for comparison of priors would be a consensus prior satisfying Bernardo and Ramon's above requirement. The beta$(1, b)$ prior appears to be the natural family of informative priors to consider when θ is known to be small, and again our results favor the B–L prior as the choice of consensus prior over the popular Jeffreys prior.

The article is set out as follows. Section 2 discusses noninformative priors and an example from Congdon (2001). Section 3 examines informative priors and revisits an example from Hanley and Lippman-Hand (1983); see also Winkler, Smith, and Fryback (2002). This includes an analysis of the performance of the Jeffreys and B–L priors. We further discuss these priors in Section 4.

  DOI: 10.1198/000313008X267839

2. NONINFORMATIVE PRIORS

We believe that Geisser's (1984) arguments in favor of the B–L prior should have put an end to the discussion of noninformative priors. Geisser compared predictive distributions based on the different Jeffreys priors resulting from binomial and negative binomial sampling with finite population results, based on a discrete uniform prior. Geisser showed that the posterior distributions for the hypergeometric and negative hypergeometric parameters coincide with the posterior predictive distribution based on the B–L prior, and that when the population size $N \to \infty$ they correspond to the B–L posterior. We would like to add that even without considering sampling rules it seems unclear at which stage, when $N \to \infty$ again, the generally accepted uniform prior for the hypergeometric parameter, that is, $p(\theta) = 1/(N+1)$ with $\theta = 0, 1/(N+1), 2/(N+1), \ldots, 1$, should give way to the U-shape of the various alternatives.

Invited comments to Geisser (1984), by proponents of alternative priors, indicated that more evidence was required. Geisser seemed to supply such in his reply, for example, pointing to the problems resulting from sampling-rule-dependence. However, more than two decades later the case for objective Bayesian analysis (Berger 2006) is strong, which includes the Jeffreys rather than the B–L prior for the binomial parameter. For the sake of completeness we discuss the Haldane and Zellner priors also, the resulting set of four called "plausible" by Berger (1985, p. 89).

2.1 Options

The Haldane prior beta$(0, 0)$ still seems popular partially due to the posterior mean being equal to x/n, that is, the maximum likelihood estimator. It also corresponds to a uniform prior for the log odds. Lindley (1965, Part 2, p. 145) recommended it and Hora and Kelley (1983) applied it to odds and risk ratio examples including samples with $x = 1$. However, when x equals 0 or n, the Haldane posterior is in fact improper which, as Bernardo (1979) stated, is less than adequate. To effectively use the Haldane prior, but still allow for the possibility of $x = 0$ or n, some analysts have adopted a beta prior with very small parameters; we examine the impact of this in the following example.

The Jeffreys prior beta$(\frac{1}{2}, \frac{1}{2})$, which at first glance appears a compromise between the Haldane and the B–L priors, follows from the invariance rule derived by Jeffreys (1946). It implies a uniform prior on $\sin^{-1}(\sqrt{\theta})$ and is preferred by objective Bayesians: for one-parameter problems, Jeffreys' rule is usually equivalent to both the data translated likelihood idea of Box and Tiao (1973) and the reference prior methodology originally introduced by Bernardo (1979). Its purpose (invariance under transformations) is perhaps not generally understood; for example, Henderson and Meyer (2001) stated that "if the actual θ is thought to be close to zero or one, the Jeffreys beta$(\frac{1}{2}, \frac{1}{2})$ prior is often recommended" and then pointed out that in this case the practitioner should know *which* value the parameter is near. However, noninformative priors are not *intended* to represent prior knowledge, so that symmetry is an obvious requirement, as the a priori labeling of "successes" and "failures" is arbitrary.

The Bayes–Laplace prior beta$(1, 1)$ is originally due to Bayes (1763). According to Berger, Liseo, and Wolpert (1999), Laplace suggested that parameterizations be chosen such that a uniform prior is reasonable, but in fact Bayes argued that a priori, in a state of complete ignorance, each number of successes should be equally probable, that is, $p(x) = 1/(n+1)$ for $x = 0, \ldots, n$. This is usually called the prior predictive distribution and calculated as $p(x) = \int p(x|\theta)p(\theta)d\theta$, so that $p(\theta) = 1$ is *implied* by this argument rather than assumed. The implied invariance under transformations $\phi = \phi(\theta)$, which here follows from fixing the prior predictive by simply setting $p(\phi) = |d\theta/d\phi|$, was hinted at by Edwards (1978). Note that the B–L posterior is calculated by simply normalizing the likelihood function, so that its mode is equal to x/n, perhaps just as attractive as the Haldane posterior property.

The fourth prior is based on a "maximal data information" approach due to Zellner (1977). The Zellner prior is not a beta distribution and is proportional to $\theta^{\theta}(1-\theta)^{1-\theta}$. It is proper with normalization constant 1.6186, which is also the pdf value at the extremes. Figure 1 shows that the Zellner prior sits between the B–L and Jeffreys priors. Clearly the flat B–L prior will be data-dominated for all x and n under Bayes' rule. Both the Haldane and Jeffreys priors shoot off to infinity at the endpoints, as Zellner remarked in his comment to Geisser (1984), and most of their mass is near 0 and 1. This appears to unduly affect the $x = 0$ examples below, whereas for x close to 0 or n it appears that the Zellner and B–L posteriors are similar. This is illustrated strongly for $x = 0$ by the example in Section 3.

Clearly, the posterior distributions corresponding to the above four priors have modes at 0 when $x = 0$, in which case we believe highest posterior density (HPD) intervals are preferable to central intervals, as they are one-sided here and thus include the area close to 0.

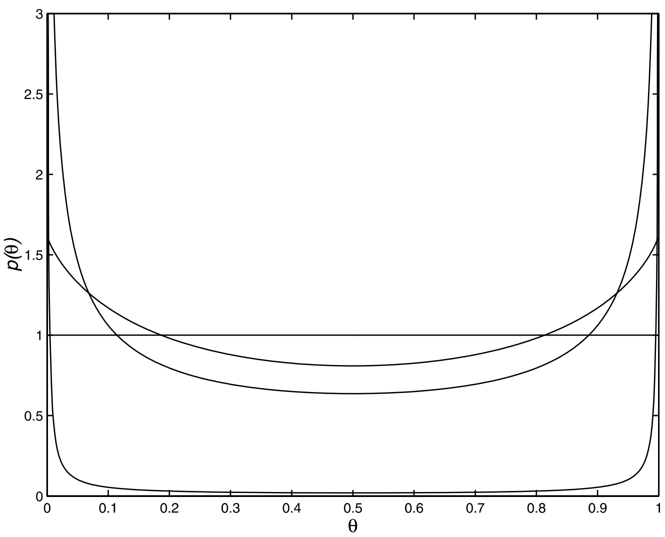

Figure 1. Four "noninformative" priors; from top to bottom (at $\theta = \frac{1}{2}$): Bayes–Laplace, Zellner, Jeffreys, and Haldane, the latter approximated by beta$(0.01, 0.01)$.

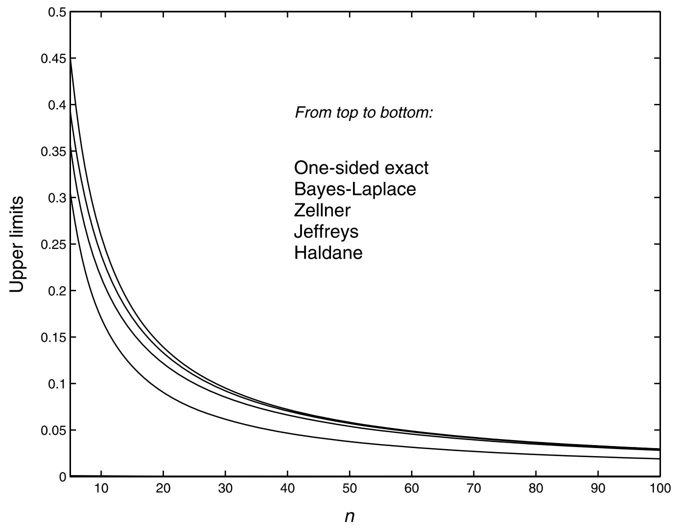

Figure 2. Various 95% upper limits for $x = 0$ and n between 5 and 100. The Haldane limits, derived from the "approximate" beta$(0.01, 0.01 + n)$ posterior, are indistinguishable from the x-axis.

2.2 Example

Congdon (2001, p. 31) gave an example based on the prior beta$(0.01, 0.01)$, as a "proper" approximation of the Haldane prior. The example is based on $n = 8$ and $x = 0$, for which Congdon gave the 95% credible interval for θ as [0, 0.007]. To call this interval "less than adequate" is an understatement. Table 1 shows that this "Haldane" interval seems to be a slightly inaccurate central interval, with a lower limit that is indeed hardly worth showing; however, the more sensible HPD interval is here even worse compared with the intervals resulting from the other three priors. Due to the extra weight in these priors near 0, the Jeffreys and Zellner intervals are also shorter than those based on the B–L prior.

For the above example, Congdon gave "the classical interval" as $[0, 0.31]$, referring to the calculation $\theta_u = 1 - \alpha^{1/n}$ given by Blyth (1986). Blyth considered one-sided exact limits only, when the two-sided exact interval Clopper and Pearson (1934) has upper limit $1 - (\alpha/2)^{1/n}$. While the two-sided version is known to be conservative, the one-sided is less so and is more comparable with HPD credible intervals when $x = 0$. In particular, the B–L upper limit is $1 - \alpha^{1/(n+1)}$.

Figure 2 shows how the above upper limits of these intervals change for increasing n. (The upper limits corresponding to $a = 0.01$ cannot be distinguished from the horizontal axis.) The increasing similarity of the one-sided exact, B–L and Zellner limits is clear: at $n = 100$ they are 0.0295, 0.0292, and 0.0282, respectively. These values reflect the so-called Rule of Three (Hanley and Lippman-Hand 1983; Jovanovic and Levy 1997): the one-sided exact limit is well approximated by $-\log(\alpha)/n$, or about $3/n$ for $\alpha = 0.05$, when n is not too small. Also, when $x = 0$ and n is large, the Zellner posterior is approximated by a beta$(1, n + 2)$ posterior. Thus the differences between the one-sided exact, B–L, and Zellner limits (in order of decreasing size) converge to 0.

In contrast, the Jeffreys limit is 0.0190 when $n = 100$. In fact, limiting values can be found by adapting a result by Leemis and Trivedi (1996). The 95% upper credible limit θ_u resulting from a beta(a, b) prior and $x = 0$ follows from $2(n+a+b-1)\theta_u \to \chi^2_{2a,\alpha}$, so that the Jeffreys prior leads to a rule of 1.92. While the above beta$(0.01, 0.01)$ prior leads to an extreme rule of 0.0034, these results suggest that *any* beta(a, a) priors with $a < 1$ are undesirable.

Of course, objective Bayesians may argue that for the above example the Jeffreys interval is *not* too short. However, given that one of Bernardo and Ramon's (1998) desiderata for consensus priors concerns sampling properties, we consider that the poor (minimum) frequentist coverage of the Jeffreys interval is relevant, specifically due to $x = 0$ (n). This persists for large n: it can be shown that for θ just above the 95% upper limit for $x = 0$, coverage approaches 84.0%, based on the Poisson approximation and HPD intervals for all x. Instead, the B–L HPD interval has better frequentist properties with minimum coverage approaching 92.7% and mean coverage equal to nominal for all n.

3. INFORMATIVE PRIORS

We revisit an example for which a consensus prior is useful for sensitivity analysis, after first discussing possible informative priors in the context of small θ. This example shows that (1) informative beta(a, b) priors with $a \ll 1$ are dangerous, and (2) the Jeffreys posterior is unsuitable to compare with posteriors based on informative priors, for the purpose of sensitivity analysis.

3.1 Options

The beta(a, b) family seems mostly sufficiently flexible to represent prior information about θ. However, Chen and Novick (1984) added three and four parameter extensions, while Diaconis and Ylvisaker (1985) and Robert (1994, p. 106) employed mixtures of betas. Jovanovic and Levy (1997) proposed the

Table 1. Binomial $(x = 0, n = 8)$ example: 95% credible intervals based on four "noninformative" priors.

Prior	HPD Lower	HPD Upper	Central Lower	Central Upper
Haldane (approx.) beta$(0.01, 0.01)$	0	0.0004	4.7×10^{-168}	0.0062
Jeffreys beta$(0.5, 0.5)$	0	0.208	0.00006	0.262
Zellner	0	0.254	0.0022	0.307
Bayes-Laplace beta$(1, 1)$	0	0.283	0.0028	0.336

Table 2. Contrast agent example ($x = 0$ and $n = 167$): aspects of six priors and corresponding posteriors (between brackets).

Beta prior:	(.042, 27.96)	(1, 666)	(1, 398)	(1,1)	Zellner	$(\frac{1}{2}, \frac{1}{2})$
Mean	**0.15%** (0.022%)	**0.15%** (0.12%)	0.25% (0.18%)	50% (0.59%)	50% (0.58%)	50% (0.30%)
0.95 Quantile	**0.75%** (0.11%)	0.45% (0.36%)	**0.75%** (0.53%)	95.0% (1.77%)	96.7% (1.73%)	99.4% (1.14%)
Pr(θ < 0.75%)	**0.950** (0.995)	0.993 (0.998)	**0.950** (0.986)	0.0075 (0.718)	0.012 (0.728)	0.055 (0.887)
Pr(θ < 0.15%)	0.893 (0.961)	0.632 (0.714)	0.450 (0.572)	0.0015 (0.223)	0.0024 (0.229)	0.025 (0.521)
Minimum n	111	1330	1598	1995	1988	1280

beta$(1, b)$ family of priors for use when θ is known to be close to 0, which is equivalent to having observed $b - 1$ prior trials without a success. Tebbs, Bilder, and Moser (2001) argued that this prior is reasonable in group-testing (pooled-testing), which also deals with small θ.

Winkler, Smith, and Fryback (2002) argued that the beta$(1, b)$ family is unnecessarily limited and reverted to the general beta(a, b) distribution. Incidentally, Jovanovic and Levy (1997) pointed out that a should not exceed 1, as in that case the posterior mode is not at 0 when $x = 0$, but they gave no reason as to why a should not be less than 1 either. Of course it is true that the additional parameter gives more flexibility, but in the following example we show that this may be excessive.

3.2 Example

Winkler, Smith, and Fryback (2002) discussed an example from Hanley and Lippman-Hand (1983), concerning a new contrast agent, used by radiologists, which has caused no reaction in 167 patients. The historical information on the old agent is expressed as "about 15 of every 10,000." Hanley and Lippman-Hand performed the obvious calculation $p(x = 0|\theta = 0.15\%, n = 167) = 0.778$ and concluded that finding no reaction is not at all surprising. In other words, with the old agent too, quite long runs with no reaction can be expected. In their pre-analysis Winkler et al. gave the Rule of Three upper limit of $3/167 = 1.8\%$, well above 0.15%. This rule implies that for the one-sided upper exact limit to be below 0.15%, a minimum of approximately 2,000 patients with no reaction is required; the one-sided exact result is actually 1996.

Winkler et al. (2002) adopted the prior beta$(0.042, 27.96)$ with prior mean equal to the historical value of 0.15% and 95th percentile equal to, rather arbitrarily, five times this value. In the context of the beta$(1, b)$ prior these constraints correspond to $b = 666$ and $b = 398$, respectively. Before studying further, we can see a major problem with the Winkler et al. prior, specifically because $a \approx 0$. For example, this prior has 80% of its mass to the left of $\theta = 0.01\%$ (i.e., *one fifteenth* of the historical mean), as opposed to 4% for the beta$(1, 398)$ prior.

Table 2 summarizes the above three priors and the corresponding posteriors. The three noninformative priors also displayed in this table are discussed later. We consider the decreases in the Winkler et al. prior mean and 95th percentile far too large. In contrast, the results from the two beta$(1, b)$ priors appear to be reasonable; given that we already know how many patients with no reaction are required for statistical significance (from a "noninformative" frequentist point of view), the fairly minor changes in the posterior estimates corresponding to the beta$(1, b)$ priors make sense. Overall, results based on the beta$(1, 666)$ and beta$(1, 398)$ priors are also quite similar.

Given the suggested amount of historical data, the question of whether the new contrast agent is better than the old one, might be addressed by the probability $\Pr(\theta < 0.15\%|x, n)$. For the quite informative beta$(1, 666)$ prior, for example, this probability has increased from 0.63 to 0.71, but is still considerably smaller than Winkler et al.'s *prior* probability of 0.89. Thus the beta$(0.042, 27.96)$ prior appears to be excessively informative.

Table 2 also shows the minimum sample sizes required to satisfy $\Pr(\theta < 0.15\%|x = 0, n) \geq 0.95$. The very small number suggested by the Winkler et al. (2002) prior is arguably quite irresponsible. Next, the Jeffreys sample size is smaller than that of the beta$(1, b)$ priors, which seems undesirable also. In contrast, the Zellner and B–L sample sizes are very close, as are their other entries in the table. In fact, the Zellner sample size is similar to that under a beta(a, a) prior with $a = 0.995$; for the other rows a needs to be slightly smaller to produce the equivalent posterior result, but never less than 0.96. In other words, for this example any beta(a, a) prior with $a < 0.96$ is arguably more informative than the Zellner prior, a strong indication that any $a < 1$ could be considered informative.

Figure 3 shows $\Pr(\theta < 0.15\%|x = 0, n)$ as a function of n. As expected, based on the Winkler et al. (2002) prior this probability approaches 1 for relatively small n. While the curve corresponding to the Zellner prior (not shown) is virtually identical to that based on the B–L prior, the informative nature of the Jeffreys prior for this application can also be seen here. When $n \geq 273$ the probability of interest based on the Jeffreys prior exceeds that based on the beta$(1, 398)$ prior and when $n \geq 1039$ it exceeds that based on the beta$(1, 666)$ prior.

Clearly, it is possible for the Jeffreys posterior to have more weight near 0 than posteriors based on the beta$(1, b)$ prior, even when b is quite large. This goes against the idea of a consensus prior. In this respect the B–L posterior is satisfactory: the impli-

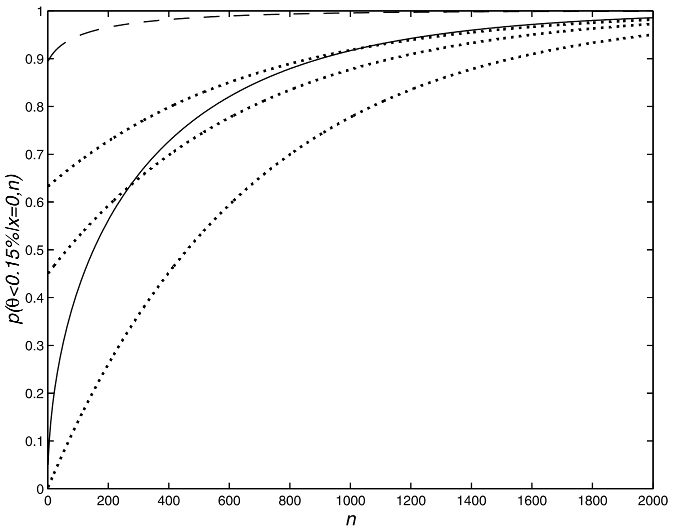

Figure 3. Contrast agent example: posterior probability of an improvement given the number of patients with no reaction. Dashed curve: based on beta(0.042, 27.96) prior; solid curve: based on beta($\frac{1}{2}$, $\frac{1}{2}$) prior; dotted curves, bottom to top: based on beta(1,1), beta(1,398), and beta(1,666) priors.

cations of any informative beta(a, b) prior are clear, in terms of "prior" successes and failures, as long as $a, b \geq 1$.

4. DISCUSSION AND CONCLUSIONS

This article has shown that in the binomial(x, n) case with $x = 0$ (or, equivalently, $x = n$) and any n, the use of a prior with small beta parameters ($a, b < 1$) should be avoided, both for noninformative and informative priors. This includes the popular Jeffreys prior. It is clear that the improper Haldane prior does not cater for $x = 0$ or $x = n$, and it is inappropriate to use a proper prior that approximates it (such as beta(0.01,0.01), which has nearly all its weight near 0 and 1). For informative priors in the context of $x = 0$, it was shown that the choice of the beta(1, b) family by Jovanovic and Levy (1997) is not as arbitrary as it seems: for the type of data on hand, a should not be less than 1 as this is too informative, and a greater than 1 would cause the posterior mode to be away from 0 even when $x = 0$. When a is fixed at 1 the choice of b simply causes the prior to represent the equivalent of $b - 1$ failures, when compared with the Bayes–Laplace (B–L) prior. This property is not shared by the Zellner prior which otherwise is very close to, while slightly more informative than, the B–L prior.

In summary, the evidence from these two examples with $x = 0$, such as (1) agreement with the Rule of Three and (2) intuitive comparisons with sensible informative conjugate priors, together with (3) impressive frequentist properties of the B–L HPD interval, which is clearly (4) data-dominated for all x, and (5) Geisser's (1984) arguments, leads us to recommend the B–L prior as a consensus prior, with the caveat that all four "plausible" priors (Berger 1985) will give highly similar results for $0 \ll x \ll n$.

[Received February 2007. Revised September 2007.]

REFERENCES

Bayes, T. R. (1763), "An Essay Towards Solving a Problem in the Doctrine of Chances," *Philosophical Transactions of the Royal Society of London*, 53, 370–418.

Berger, J. O., Liseo, B., and Wolpert, R. L. (1999), "Integrated Likelihood Methods for Eliminating Nuisance Parameters," *Statistical Science*, 14, 1–22.

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

——— (2006), "The Case for Objective Bayesian Analysis," *Bayesian Statistics*, 1, 385–402.

Bernardo, J. M. (1979), "Reference Posterior Distributions for Bayesian Inference" (with discussion), *Journal of the Royal Statistical Society*, Series B, 41, 113–147.

——— (2005), "Reference Analysis," in *Bayesian Thinking: Modeling and Computation*, eds. D. K. Dey and C. R. Rao, Amsterdam: Elsevier, pp. 17–90.

Bernardo, J. M., and Ramon, J. M. (1998), "An Introduction to Bayesian Reference Analysis: Inference on the Ratio of Multinomial Parameters," *The Statistician*, 47, 101–135.

Berry, D. A. (1996), *Statistics—A Bayesian Perspective*, Belmont, CA: Duxbury.

Blyth, C. R. (1986), "Approximate Binomial Confidence Limits," *Journal of the American Statistical Association*, 81, 843–855.

Box, G. E. P., and Tiao, G. C. (1973), *Bayesian Inference in Statistical Analysis*, New York: Wiley Classics.

Chen, J. J., and Novick, M. R. (1984), "Bayesian Analysis for Binomial Models with Generalized Beta Prior Distributions," *Journal of Educational Statistics*, 9, 2, 163–175.

Clopper, C. J., and Pearson, E. S. (1934), "The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial," *Biometrika*, 26, 404–416.

Congdon, P. (2001), *Bayesian Statistical Modeling*, Chichester: Wiley.

Diaconis, P., and Ylvisaker, D. (1985), "Quantifying Prior Opinion," in *Bayesian Statistics* (vol. 2), eds. J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith, Amsterdam: North-Holland, pp. 163–175.

Edwards, A. W. F. (1978), "Commentary on the Arguments of Thomas Bayes," *The Scandinavian Journal of Statistics*, 5, 116–118.

Geisser, S. (1984), "On Prior Distributions for Binary Trials," *The American Statistician*, 38, 244–251.

Hanley, J. A., and Lippman-Hand, A. (1983), "If Nothing Goes Wrong, is Everything All Right?," *Journal of the American Medical Association*, 249, 1743–1745.

Henderson, M., and Meyer, M. C. (2001), "Exploring the Confidence Interval for a Binomial Parameter in a First Course in Statistical Computing," *The American Statistician*, 55, 337–344.

Hora, S. C., and Kelley, G. D. (1983), "Bayesian Inference on the Odds and Risk Ratios," *Communications in Statistics—Theory and Methods*, 12, 725–738.

Jeffreys, H. (1946), "An Invariant Form for the Prior Probability in Estimation Problems," *Proceedings of the Royal Society of London*, Series A, 186, 453–461.

Jovanovic, B. D., and Levy, P. S. (1997), "A Look at the Rule of Three," *The American Statistician*, 51, 137–139.

Leemis, L. M., and Trivedi, K. S. (1996), "A Comparison of Approximate Interval Estimators for the Bernoulli Parameter," *The American Statistician*, 50, 63–68.

Lindley, D. V. (1965), *Introduction to Probability and Statistics from a Bayesian Viewpoint*, New York: Cambridge University Press.

Louis, T. A. (1981), "Confidence Intervals for a Binomial Parameter After Observing No Successes," *The American Statistician*, 35, 154–154.

Robert, C. P. (1994), *The Bayesian Choice* (2nd ed.), New York: Springer–Verlag.

Tebbs, J. M., Bilder, R. B., and Moser, B. K. (2001), "An Empirical Bayes Group-Testing Approach to Estimating Small Proportions," *Communications in Statistics—Theory and Methods*, 32, 983–995.

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.

Zellner, A. (1977), "Maximal Data Information Prior Distributions," in *New Developments in the Applications of Bayesian Methods*, eds. A. Aykac and C. Brumat, Amsterdam: North-Holland, pp. 211–232.