[Interval Estimation for a Binomial Proportion]: Comment
Author(s): Malay Ghosh
Source: *Statistical Science*, Vol. 16, No. 2 (May, 2001), pp. 124-125
Published by: Institute of Mathematical Statistics
Stable URL: https://www.jstor.org/stable/2676788
Accessed: 30-09-2018 02:41 UTC

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at https://about.jstor.org/terms

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve and extend access to *Statistical Science*

This content downloaded from 132.174.254.12 on Sun, 30 Sep 2018 02:41:36 UTC
All use subject to https://about.jstor.org/terms

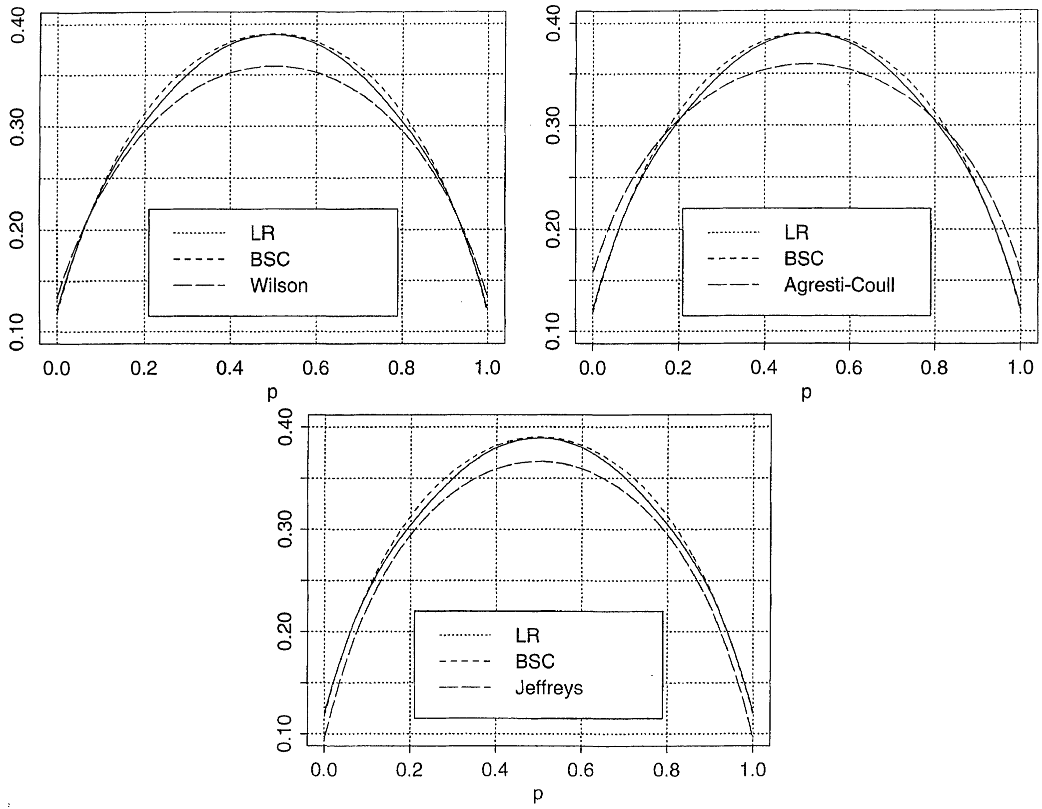

FIG. 2. *Expected lengths of BSC and LR intervals as a function of p compared, respectively, to Wilson, Agresti–Coull and Jeffreys intervals* ($n = 25$). *Compare to authors' Figure* 8.

Comment

Malay Ghosh

This is indeed a very valuable article which brings out very clearly some of the inherent difficulties associated with confidence intervals for parameters of interest in discrete distributions. Professors Brown, Cai and Dasgupta (henceforth BCD) are to be complimented for their comprehensive and thought-provoking discussion about the "chaotic" behavior of the Wald interval for the binomial proportion and an appraisal of some of the alternatives that have been proposed.

My remarks will primarily be confined to the discussion of Bayesian methods introduced in this paper. BCD have demonstrated very clearly that the modified Jeffreys equal-tailed interval works well in this problem and recommend it as a possible contender to the Wilson interval for $n \le 40$.

There is a deep-rooted optimality associated with Jeffreys prior as the unique *first-order probability matching prior* for a real-valued parameter of interest with no nuisance parameter. Roughly speaking, a probability matching prior for a real-valued parameter is one for which the coverage probability of a one-sided Baysian credible interval is asymptotically equal to its frequentist counterpart. Before giving a formal definition of such priors, we provide an intuitive explanation of why Jeffreys prior is a matching prior. To this end, we begin with the fact that if $X_1, \dots, X_n$ are iid $N(\theta, 1)$, then $\overline{X}_n = \sum_{i=1}^{n} X_i/n$ is the MLE of θ. With the uniform prior $\pi(\theta) \propto c$ (a constant), the posterior of θ

Malay Ghosh is Distinguished Professor, Department of Statistics, University of Florida, Gainesville, Florida 32611-8545 (e-mail: ghoshm@stat.ufl.edu).

This content downloaded from 132.174.254.12 on Sun, 30 Sep 2018 02:41:36 UTC
All use subject to https://about.jstor.org/terms

is $N(\overline{X}_n, 1/n)$. Accordingly, writing z_α for the upper $100\alpha\%$ point of the $N(0, 1)$ distribution,

$$P(\theta \leq \overline{X}_n + z_\alpha n^{-1/2} | \overline{X}_n)$$
$$= 1 - \alpha = P(\theta \leq \overline{X}_n + z_\alpha n^{-1/2} | \theta)$$

and this is an example of perfect matching. Now if $\hat{\theta}_n$ is the MLE of θ, under suitable regularity conditions, $\hat{\theta}_n|\theta$ is asymptotically (as $n \to \infty$) $N(\theta, I^{-1}(\theta))$, where $I(\theta)$ is the Fisher Information number. With the transformation $g(\theta) = \int^\theta I^{1/2}(t)$, by the delta method, $g(\hat{\theta}_n)$ is asymptotically $N(g(\theta), 1)$. Now, intuitively one expects the uniform prior $\pi(\theta) \propto c$ as the asymptotic matching prior for $g(\theta)$. Transforming back to the original parameter, Jeffreys prior is a probability matching prior for θ. Of course, this requires an invariance of probability matching priors, a fact which is rigorously established in Datta and Ghosh (1996). Thus a uniform prior for $\arcsin(\theta^{1/2})$, where θ is the binomial proportion, leads to Jeffreys Beta (1/2, 1/2) prior for θ. When θ is the Poisson parameter, the uniform prior for $\theta^{1/2}$ leads to Jeffreys' prior $\theta^{-1/2}$ for θ.

In a more formal set-up, let $X_1, \ldots, X_n$ be iid conditional on some real-valued θ. Let $\theta_\pi^{1-\alpha}(X_1, \ldots, X_n)$ denote a posterior $(1-\alpha)$th quantile for θ under the prior π. Then π is said to be a first-order probability matching prior if

$$\begin{aligned} P(\theta \leq \theta_\pi^{1-\alpha}(X_1, \ldots, X_n) | \theta) \\ = 1 - \alpha + o(n^{-1/2}). \end{aligned} \tag{1}$$

This definition is due to Welch and Peers (1963) who showed by solving a differential equation that Jeffreys prior is the unique first-order probability matching prior in this case. Strictly speaking, Welch and Peers proved this result only for continuous distributions. Ghosh (1994) pointed out a suitable modification of criterion (1) which would lead to the same conclusion for discrete distributions. Also, for small and moderate samples, due to discreteness, one needs some modifications of Jeffreys interval as done so successfully by BCD.

This idea of probability matching can be extended even in the presence of nuisance parameters. Suppose that $\theta = (\theta_1, \ldots, \theta_p)^T$, where θ_1 is the parameter of interest, while $(\theta_2, \ldots, \theta_p)^T$ is the nuisance parameter. Writing $I(\theta) = ((I_{jk}))$ as the Fisher information matrix, if θ_1 is orthogonal to $(\theta_2, \ldots, \theta_p)^T$ in the sense of Cox and Reid (1987), that is, $I_{1k} = 0$ for all $k = 2, \ldots, p$, extending the previous intuitive argument, $\pi(\theta) \propto I_{11}^{1/2}(\theta)$ is a probability matching prior. Indeed, this prior belongs to the general class of first-order probability matching priors

$$\pi(\theta) \propto I_{11}^{1/2}(\theta) h(\theta_2, \ldots, \theta_p)$$

as derived in Tibshirani (1989). Here $h(\cdot)$ is an arbitrary function differentiable in its arguments.

In general, matching priors have a long success story in providing frequentist confidence intervals, especially in complex problems, for example, the Behrens–Fisher or the common mean estimation problems where frequentist methods run into difficulty. Though asymptotic, the matching property seems to hold for small and moderate sample sizes as well for many important statistical problems. One such example is Garvan and Ghosh (1997) where such priors were found for general dispersion models as given in Jorgensen (1997). It may be worthwhile developing these priors in the presence of nuisance parameters for other discrete cases as well, for example when the parameter of interest is the difference of two binomial proportions, or the log-odds ratio in a 2×2 contingency table.

Having argued so strongly in favor of matching priors, I wonder, though, whether there is any special need for such priors in this particular problem of binomial proportions. It appears that any Beta (a, a) prior will do well in this case. As noted in this paper, by shrinking the MLE X/n toward the prior mean 1/2, one achieves a better centering for the construction of confidence intervals. The two diametrically opposite priors Beta (2, 2) (symmetric concave with maximum at 1/2 which provides the Agresti–Coull interval) and Jeffreys prior Beta (1/2, 1/2) (symmetric convex with minimum at 1/2) seem to be equally good for recentering. Indeed, I wonder whether any Beta (α, β) prior which shrinks the MLE toward the prior mean $\alpha/(\alpha + \beta)$ becomes appropriate for recentering.

The problem of construction of confidence intervals for binomial proportions occurs in first courses in statistics as well as in day-to-day consulting. While I am strongly in favor of replacing Wald intervals by the new ones for the latter, I am not quite sure how easy it will be to motivate these new intervals for the former. The notion of shrinking can be explained adequately only to a few strong students in introductory statistics courses. One possible solution for the classroom may be to bring in the notion of continuity correction and somewhat heuristcally ask students to work with $(X + \frac{1}{2}, n - X + \frac{1}{2})$ instead of $(X, n - X)$. In this way, one centers around $(X + \frac{1}{2})/(n + 1)$ a la Jeffreys prior.

This content downloaded from 132.174.254.12 on Sun, 30 Sep 2018 02:41:36 UTC
All use subject to https://about.jstor.org/terms