the real problem stems from a desire to construct an interval to summarize the posterior distribution. The posterior distribution itself is invariant to transformations and is a much more informative summary of the statistical inference. It should be preferred over any particular Bayesian interval.

The second task, specifying a model for the sampling distribution (or likelihood), is truly subjective. In any given analysis some models are clearly inappropriate, but there always remain models among which the data are unable to distinguish. In some cases we make a parsimonious choice and in others the choice has little effect on the final analysis. In any case, specification of the sampling distribution is a subjective task common to all statistical analyses. The choice is critical, sometimes highly influential, and thus should be approached with care and checked when possible against the data, rather than holding to an arbitrary initial proposal.

I save the seemingly most potent criticism for last. Indeed in her discussion of Bayesian methods as a potential solution to the difficulties encountered by frequentist methods in the presence of nuisance parameters, Reid pointed to the necessary specification of a "prior [distribution] for a high-dimensional nuisance parameter" as justification for her conclusion that "the fact that the Bayesian approach is logically consistent strikes me as somewhat irrelevant" (Reid, 1995, see also McCullagh, 1995). Here, however, these concerns do not seem to apply. In particular, the prior distributions for nuisance parameters are neither subjective nor uninformative; they are based on calibration data and merely enable the inference to reflect uncertainty in the calibration variables. The parameter of interest is of low dimension, dimension one in the current model formulation, where $p(\mu) \propto 1$ is an obvious choice. Even with higher dimensional parameters, hierarchical models or hierarchical prior specifications serve to mitigate Reid's concern. The sensitivity of the final analysis to the choice of prior distribution as well as the frequency properties of the resulting intervals can be explored. Indeed, in this case, a prior distribution seems neither difficult to specify nor subjective, at least not when compared with the subjective nature of the principles underlying the alternatives.

ACKNOWLEDGMENT

The author thanks Tom Loredo for pointing out several references and the recent relevant workshops at CERN and Fermi Lab. Funding was partially provided by NSF Grant DMS-01-04129 and by NASA Contract NAS8-39073 (CXC).

Comment

Michael Woodroofe and Tonglin Zhang

We thank Professor Mandelkern for his informative review of statistical problems that have been plaguing physicists and his attempts to address them. We have some minor quibbles with the "desirable features," some brief comments on the Bayesian and unified methods with known b and σ^2, and more extensive comments on treating σ^2 as an estimated parameter instead of a known one.

Quibbles. In (i), statisticians have been searching for a general method that is neither arbitrary or subjective and makes intuitive sense for a long time now without any general consensus on what that method is. In (ii), there is certainly a need for a method that does not require prior information; but using prior information should not be precluded when it exists. Also, requiring equivariance under one-to-one transformations, as in (iii), rules out many intuitive optimality criteria.

Known b and σ^2. The unified method was developed explicitly to deal with problems of a restricted parameter space. It clearly provides an improvement over the Neyman intervals and has attracted a wide following among physicists. We agree with Mandelkern, however, that it can produce unbelievably short intervals. The Bayesian intervals are not especially short in the Poisson case, as is clear from Mandelkern's Figure 4. In the extreme case $N = 0$, the length of the Bayesian interval is $\log(1/\alpha)$, and this is the right answer in the absence of prior information. To elaborate,

Michael Woodroofe is Professor and Tonglin Zhang is a graduate student Department of Statistics, University of Michigan, 4082 Frieze Building, Ann Arbor, Michigan 48109 (e-mail: michael@umich.edu).

suppose that $N = B + S$, where $B \sim \text{Poisson}(b)$ and $S \sim \text{Poisson}(\theta)$ are independent, b is known, and θ is unknown. If $N = 0$, then $B = 0$ and $S = 0$, and common sense dictates that the confidence interval for θ should be the same as if we had observed (only) $S = 0$. When $S = 0$ is observed the Bayesian and Neyman upper credible/confidence bounds are both $\log(1/\alpha)$, and the unified bound is only slightly larger. So we do not believe that the Bayesian intervals are counterintuitive in the Poisson case.

In the normal case, the Bayesian credible interval shrinks to $\{0\}$ as $x \to -\infty$; that is, letting $u(x) = x + d(x)$ denote the upper credible limit,

$$\lim_{x \to -\infty} u(x) = 0. \tag{1}$$

This may appear counterintuitive, for reasons given in the paper, but it is consistent with the solution to the Poisson case (which we maintain is the right solution). For if b is large, then $X = 2[\sqrt{N} - \sqrt{b}]$ is approximately normal with mean $\mu = 2[\sqrt{b+\theta} - \sqrt{b}]$ and unit variance; and if $N = 0$, then upper credible bounds for θ and μ are $\log(1/\alpha)$ and

$$2\left[\sqrt{b + \log\left(\frac{1}{\alpha}\right)} - \sqrt{b}\right] < \frac{1}{\sqrt{b}} \log\left(\frac{1}{\alpha}\right) = \frac{2}{|x|} \log\left(\frac{1}{\alpha}\right).$$

So (1) does not seem unreasonable when applied to $X = 2[\sqrt{N} - \sqrt{b}]$. To the extent that (1) appears counterintuitive, it does so because large values of $-x$ cast doubt on the model.

Estimated nuisance parameters. Reassessing the model can introduce biases, as Mandelkern says, but it is necessary sometimes and does not always introduce severe biases. In the present context, the values of σ^2 and b were assumed known. This is almost certainly an oversimplification (recalling that b was given by $b = 2.88 \pm 0.13$ in an example). We show below how treating σ^2 as an estimated, rather than known, parameter in the normal case leads to important differences in the nature of the confidence bounds for negative x. Thus, suppose that there are independent data $S^2 \sim \sigma^2 \chi_r^2 / r$ and $X \sim \text{Normal}(\mu, \sigma^2)$, where $0 \le \mu < \infty$ and $0 < \sigma^2 < \infty$ are both unknown, but $r \ge 1$ is known. Then σ^2 is unbiasedly estimated by S^2, and the likelihood function is

$$L(\mu, \sigma^2 \mid x, s^2) \propto \frac{1}{\sigma^{r+1}} \exp\left[-\frac{rs^2 + (x-\theta)^2}{2\sigma^2}\right].$$

This simple change in the model has a profound effect on the nature of the Bayesian credible intervals for large values of $-x/s$.

Estimated σ^2: The Bayesian view. In the enlarged model, credible intervals for μ may be obtained from the (improper) prior $d\mu\, d\sigma^2/\sigma^2$ over $0 \le \mu < \infty$, $0 < \sigma^2 < \infty$. After some routine calculation, the (marginal) posterior distribution of μ is

$$g(\mu \mid x, s) = \frac{1}{H_r(t)} h_r\left(\frac{\mu - x}{s}\right),$$

where $t = x/s$ and h_r and H_r are the density and distribution function of the t-distribution on r degrees of freedom. Equivalently, the posterior distribution of $(\mu - x)/s$, given $X = x$ and $S = s$, is a t-distribution with r degrees of freedom, conditioned to exceed $-t$. There is then a complete analogue with the results of Roe and Woodroofe (2001). Letting $t_0 = H_r^{-1}[1/(1+\alpha)]$, a level $1-\alpha$ Bayesian credible interval for μ has the form $[s\ell(t), su(t)] = [\max(0, x - sd), x + sd]$, where $d = H_r^{-1}[1 - \alpha H_r(t)]$ if $t \le t_0$ and $d = H_r^{-1}[\frac{1}{2} + \frac{1}{2}(1-\alpha)H_r(t)]$ if $t > t_0$. Further, a level $1-\alpha$ credible interval has (frequentist) coverage probability at least $(1-\alpha)/(1+\alpha)$, even without any ad hoc modification, and the latter bound is conservative. See Zhang and Woodroofe (2001) for the derivations.

Graphs of the $\ell(t)$ and $u(t)$ are included in Figure 1 for selected α and r. Comparing the latter figure with Mandelkern's Figure 2 shows that including an unknown σ^2 in the model changes the nature of the upper credible limit qualitatively for large values of $-x$. In the enlarged model the upper confidence limit is *decreasing* in x for fixed s when $-x$ is large and even approaches ∞ as $x \to -\infty$. The explanation for this behavior is that if $-x$ is large, then the posterior distribution of σ^2 can be quite diffuse, even if

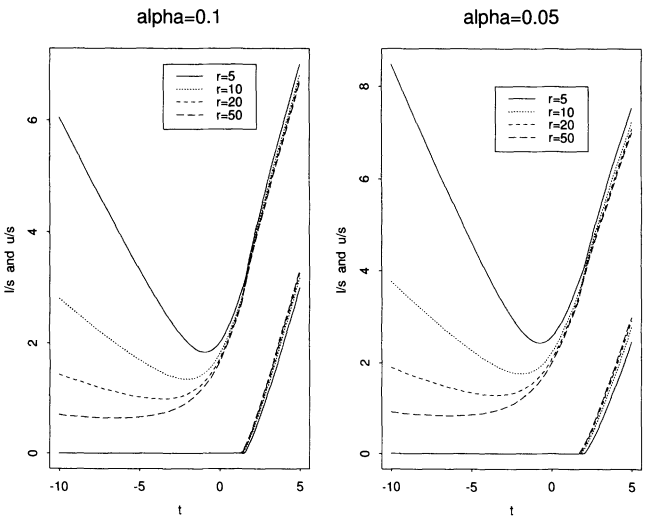

FIG. 1. *Bayesian confidence limits when $s = 1$.*

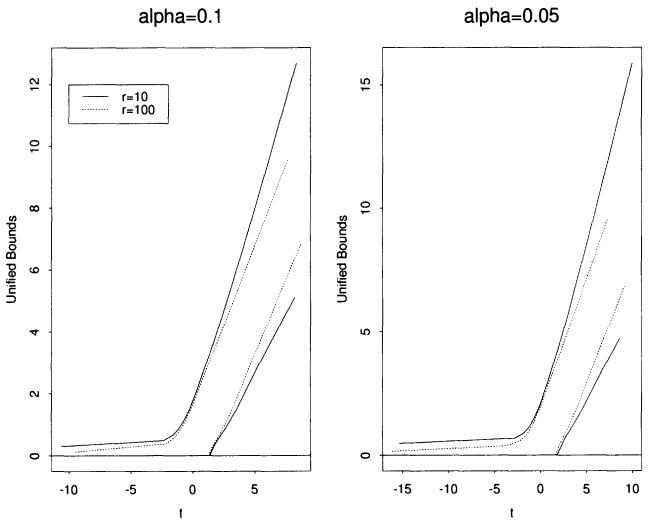

FIG. 2. *Unified confidence limits for* $\delta = \mu/\sigma$.

$s = 1$. Comparing Figure 1 with Mandelkern's Figure 3 shows that the Bayesian approach with estimated σ^2 discounts large values of $-x$ to an even greater extent that the maximum likelihood approach (with known σ^2).

The unified approach. Including an estimated σ^2 in the model causes some problems for the unified approach, which is naturally suited to one parameter families or to constructing confidence regions for the vector of all parameters in a multiparameter model. There are three possible ways to proceed. It is straightforward to construct confidence regions for (μ, σ^2), but then projections on the μ axis will lead to very long intervals. It is natural to reduce by invariance. The unified method can be applied to the distributions of $T = X/S$ to construct confidence intervals for μ/σ. The family of noncentral t-distributions is hard to use in this way, however. A simpler approach is to use the likelihood ratio statistic for composite hypotheses, $H_\delta : \mu/\sigma = \delta$. That is, letting

$$R_\delta(t) = \frac{\sup_{\mu=\delta\sigma} L(\mu, \sigma^2 \mid x, s^2)}{\sup_{\mu,\sigma\geq 0} L(\mu, \sigma^2 \mid x, s^2)},$$

where $L(\mu, \sigma^2 \mid x, s^2)$ is the likelihood function, the unified confidence intervals for δ are $\{\delta : R_\delta(t) \geq c_\delta\}$, where c_δ are determined by $P_\delta[R_\delta(T) \geq c_\delta] = 1 - \alpha$. Here $R_\delta(t)$ depends only on t by scale invariance, and $T = X/S$ has the noncentral t-distribution with r degrees of freedom and noncentrality parameter δ. After some calculation,

$$R_\delta(t) = \left[\frac{r + t_-^2}{r+1}\right]^{(r+1)/2} \psi_\delta(t)^{r+1} \cdot \exp\left(-\frac{1}{2}\delta^2 + \frac{1}{2}\delta t \psi_\delta(t)\right),$$

where t_-^2 is the square of the negative part of t and

$$\psi_\delta(t) = \frac{\sqrt{\delta^2 t^2 + 4(r+1)(t^2+r)} + \delta t}{2(t^2+r)}.$$

If $\delta = 0$, then $\psi_0(t) = \sqrt{(r+1)/(r+t^2)}$, $R_0(t) = 1$ for $-\infty < t \leq 0$, and $R_0(t) = [r/(r+t^2)]^{(r+1)/2}$ is decreasing in $0 \leq t < \infty$. For $\delta > 0$, let $\tau_\delta =$

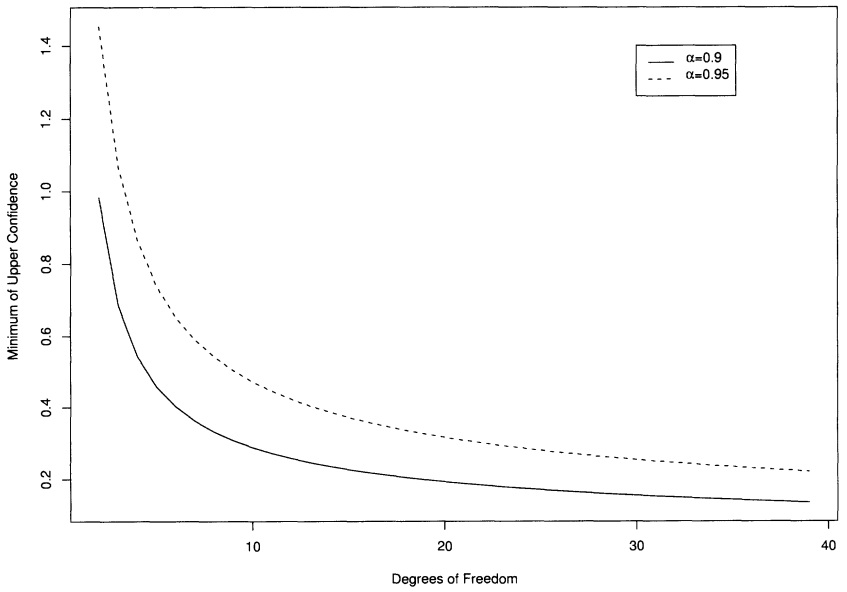

FIG. 3. *The minimum unified upper confidence limit for* δ.

$\delta\sqrt{r/(r+1)}$. Then, differentiation shows that $R_\delta(t)$ is increasing in $-\infty < t \le \tau_\delta$ and decreasing in $\tau_\delta \le t < \infty$. Further, $\lim_{t\to\infty} R_\delta(t) = 0$, and

$$\lim_{t\to-\infty} R_\delta(t) = \left[\frac{\sqrt{\delta^2+4(r+1)}-\delta}{2\sqrt{r+1}}\right]^{r+1} \cdot \exp\left[-\frac{\delta^2+\delta\sqrt{\delta^2+4(r+1)}}{4}\right]$$

by direct calculation. So $\{t: R_\delta(t) \ge c_\delta\} = \{t: a_\delta \le t \le b_\delta\}$, where $-\infty \le a_\delta < b_\delta < \infty$ are determined by $P_\delta[a_\delta \le T \le b_\delta] = 1-\alpha$, and $R_\delta(a_\delta) \ge R_\delta(b_\delta)$ with equality if $a_\delta > -\infty$. It is straightforward to compute a_δ and b_δ numerically and to solve the equation $a_u = t$ and $b_\ell = t$ for $u(t)$ and $\ell(t)$, which then serve as the upper and lower boundaries of a level $1-\alpha$ confidence interval for δ. Graphs of $\ell(t)$ and $u(t)$ are included in Figure 2 for selected α and r.

In this case, the upper boundary is increasing in t and has a positive limit, δ_0 say, as $t \to -\infty$. Letting $H_{r,\delta}$ denote the noncentral t-distribution with r degrees of freedom and noncentrality parameter δ, δ_0 is the solution to the equation $R_\delta(-\infty) = R_\delta[H_{r,\delta}^{-1}(1-\alpha)]$ and is graphed as a function of r in Figure 3 for selected values of α. So shorter intervals are again obtained for large values of $-t$, but they do not shrink to 0 as $t \to -\infty$. In any case a small value of δ can arise from a small μ, a large σ, or both.

Rejoinder

Mark Mandelkern

I think I can speak for physicists in appreciating the interest of the statistical community in the problem of confidence intervals for bounded parameters. The variety of comments by five distinguished mathematical statisticians suggests that our community has not overlooked a satisfactory procedure that has previously been published. The comments have two main themes: (1) that a Bayesian solution may be most suitable, perhaps with frequentist modification to rationalize the coverage properties as suggested by Professor Casella in his comment and previously by Mandelkern and Schultz (2000a, b) and Roe and Woodroofe (2001); (2) that enlargement or respecification of the model, even a posteriori, may be appropriate. A number of distinct suggestions have been made in the latter regard.

While the procedure used to compute a confidence interval is usually discussed in the original experimental work, it is rarely carried forward to subsequent experimental papers, reviews and theoretical analyses. For this reason it is important for intervals to be evaluated in a consistent and uniform way. Consistency is certainly more important than are the absolute values obtained. It would be particularly valuable if statisticians working in this area would propose a procedure for computing confidence intervals, which could then be adopted as a standard by the experimental physics community.

Finally, it may be most appropriate to, at least in ambiguous cases, give up the notion of characterizing experimental uncertainty with a confidence interval and instead, as suggested by Professor Gleser in his comment, to present the likelihood function for this purpose. It is interesting that Enrico Fermi, who introduced the likelihood method to physicists (Orear, 1958, 1982), suggested that likelihood functions for different experiments be multiplied for overall estimation of parameters.

ADDITIONAL REFERENCES

BROWN, L. (1967). The conditional level of Student's t-test. *Ann. Math. Statist.* **38** 1068–1071.

BUEHLER, R. J. (1959). Some validity criteria for statistical inferences. *Ann. Math. Statist.* **30** 845–863.

CASELLA, G. (1987). Conditionally acceptable recentered set estimators. *Ann. Statist.* **15** 1363–1371.

CASELLA, G. and BERGER, R. L. (2002). *Statistical Inference*, 2nd ed. Duxbury, Pacific Grove, CA.

FISHER, R. A. (1956). *Statistical Methods and Scientific Inference*. Oliver and Boyd, Edinburgh.

FRASER, D. A. S. and REID, N. (1995). Ancillaries and third order significance. *Util. Math.* **47** 33–53.

FRASER, D. A. S. and REID, N. (2001). Ancillary information for statistical inference. *Empirical Bayes and Likelihood Inference. Lecture Notes in Statist.* **148** 185–207. Springer, New York.

GARWOOD, F. (1936). Fiducial limits for the Poisson distribution. *Biometrika* **28** 437–442.