RESPONSE (TO COUSINS)

N. REID
Department of Statistics, University of Toronto, 100 St. George St., Toronto Canada M53 3G3
E-mail: reid@utstat.utoronto.ca

The paper by Cousins has provided an excellent overview of many of the problems arising with nuisance parameters. In my view it is easier and clearer to think about confidence limits from the approach using p-values, which is fairly common in the statistical literature but perhaps less so in the physics literature. Assume we have data x from a model $X \sim f(x;\theta)$ where the unknown parameters $\theta = (\psi, \nu)$ are partititoned into parameters of interest ψ and nuisance parameters ν. A p-value for testing $H_0 : \psi = \psi_0$ is given by $\Pr(T \geq t(x); \psi_0)$, where $T = t(X)$, and Pr is computed under model f for x.

Confidence intervals are easily obtained from p-values, by considering the function $p(\psi) = \Pr(T \geq t^{obs}; \psi)$: a $(1-\alpha)$ interval is obtained by finding $\{\psi : \alpha/2 \leq p(\psi) \leq 1-\alpha/2\}$, and a confidence bound can be obtained using $\{\psi : p(\psi) \geq \alpha\}$ or $\{\psi : p(\psi) \leq 1-\alpha\}$. The conversion from the function $p(\psi)$ to limits for ψ can be done exactly, approximately, or by simulation. For example, if $\bar{x} \sim N(\mu, 1)$, the p-value function for μ is $p(\mu) = 1 - \Phi\{\sqrt{n}(\bar{x} - \mu)\}$, and the confidence interval is $\bar{x} \pm \sqrt{n} z_{\alpha/2}$. If the p-value is obtained by using the approximation $-2 \ln L \sim \chi^2_1$, then the confidence interval is computed by interpolation, and is sometimes summarized as $\hat{\psi} - \sigma_-, \hat{\psi} + \sigma_+$.

There are two aspects to the definition of p-values: the choice of the summary statistic $t(X)$ to be used in assessing the consistency of the data with H_0, and the calculation of p which may be carried out exactly, by some approximation, or by simulation. I think it is helpful to separate these two aspects. For example we might be able to find a statistic $t(X)$ whose distribution is free of ν. While this distribution might be complicated, basing confidence intervals on it would guarantee coverage for all values of the nuisance parameter. The comparison in Cousins' Table 1 uses approximate calculation of the limits for the likelihood ratio, but exact calculation for the frequentist limits. However the likelihood ratio intervals could be inverted exactly and would then be identical to the frequentist shortest intervals.

There is considerable interest in the statistical literature on so-called 'matching' priors, which are priors for which the Bayesian posterior limit has good coverage. Defining the posterior limit $\theta^{(1-\alpha)}(x)$ by $\Pr\{\theta \geq \theta^{(1-\alpha)}(x) \mid x\} = \alpha$, the matching condition is $\Pr\{\theta^{(1-\alpha)}(X) \leq \theta \mid \theta\} = \alpha + \epsilon$, where $\epsilon = 1/\sqrt{n}$ or $1/n$, etc. Matching to $O(1/\sqrt{n})$ turns out to be too weak; all priors achieve this. Matching to $O(1/n)$ is uniquely achieved for models with just one parameter by Jeffreys' prior $\pi(\theta) \propto i^{1/2}(\theta)$. For the Poisson mean this gives $\pi(\mu) \propto \mu^{-1/2}$, and the corresponding interval for Cousins' Table 1 is (1.72, 5.27). This actually does match the frequentist interval, but the latter must be defined using the mid p-value $\Pr(T > t^{obs}) + (1/2) * \Pr(T = t^{obs})$.

For the ratio of Poisson means, we have a complete factorization of the likelihood $L(\lambda, \nu; x, y) = L_1(\tau) L_2(\lambda)$, where $\tau = \lambda(1+\nu)$, and L_2 is the binomial likelihood for x 'successes' out of $x+y$ events, and probability of 'success' $\lambda/(\lambda+1)$. Using L_2 is equivalent to choosing T to be X given $X+Y$, and has the advantage that distribution is free of τ. It also has the interpretation, perhaps more important, that it directly measures λ. Although it will not usually be the case, in this model the profile likelihood is identical to L_2. However the conditional distribution has fewer points of support than full distribution so it is less likely that we will be able to observe a p-value of exactly 0.05, and it is this discreteness that leads to overcoverage. This is a different phenomenon than the separation of the parameter of interest from the nuisance parameter.

Finally, I appreciate the uncertainty over the 'correct' form of adjustment to profile likelihood. The statistical literature is full of suggestions, but there is no magic bullet here either. One version that seems to improve the usual approximations involves a correction related to the observed information, and is described in my summary paper. Experience with the correction in many examples indicates that it is worth the effort to incorporate it, especially in small n situations. Liseo's[1] comparison of Bayesian analysis to methods based on modified likelihoods is somewhat misleading, as it does not use the more widely accepted likelihood approach based on higher order asymptotic theory.[2]

References

1. B. Liseo, *Biometrika* **80**, 295 (1993).
2. N. Reid, in *Bayesian Statistics V*, Ed. J.M. Bernardo *et al.*, Oxford University Press, 351, 1996.