Nearly Best Wald Intervals

George Terrell

In Terrell 2017 it was argued that the best confidence interval for a parameter θ of size α was determined by the likelihood criterion $L(\theta) \geq C_{\alpha}$; where the constant is determined by the requirement that, before the experiment is carried out, the probability that the criterion will turn out to be met is 1 – α. Such an interval can be given in closed form only under special conditions, the most important of which is symmetry of the likelihood about its maximum. Terrell 2017 proposed the construction of nearly-best confidence intervals by special transformations of the parameter to near-symmetry in the case of exponential families. Then the central limit theorem allows us to use the symmetry of the normal location family, (whose best interval is $\bar{x} - z_{\alpha/2}\frac{\sigma}{2} \leq \mu \leq \bar{x} + z_{\alpha/2}\frac{\sigma}{2}$ where for Z standard normal $P(Z > z_{\alpha/2}) = 1 - \alpha/2$) to write down nearly-best intervals for those cases.

As important as the exponential families are, we have thereby found nearly-best intervals only for a limited number of cases. For a much wider range of cases, approximate confidence intervals are usually found by the Wald formula: letting $l(\theta)$ be the log-likelihood, the maximum likelihood estimate $\hat{\theta}$ is then the solution to $l'(\theta) = 0$. Then $Var(\hat{\theta})$ is asymptotically for large numbers of independent observations approximately $\frac{-1}{l''(\hat{\theta})}$ (Wald 1943). The Wald approximate confidence interval of size α is then $\hat{\theta} - z_{\alpha/2}\frac{1}{\sqrt{-l''(\hat{\theta})}} \leq \theta \leq \hat{\theta} + z_{\alpha/2}\frac{1}{\sqrt{-l''(\hat{\theta})}}$.

The Wald interval approximates the best interval well only for large sample sizes. Its imprecision is dominated by lack of symmetry in *l*. Terrell 2017 noted that there existed an asymmetric family, the Inverse Gaussian location family, for which best confidence intervals could be written down in closed form. This family and its mirror image about the origin can be thought of as generalizing the normal family to include arbitrary asymmetry (see e.g. Shuster 1968). We will therefore propose a nearly-best confidence interval for asymmetric likelihoods by using the more general Inverse Gaussian family to construct nearly–best Wald-type approximate intervals.

We established in Terrell 2017 that an Inverse Gaussian family T with location parameter μ and skewness parameter *b* had log-likelihood $l(\mu) = -\frac{b^2}{2t}\left(\frac{t}{\mu} - 1\right)^2$ (up to an irrelevant additive constant). It was then shown that its (exact) size α best confidence interval for μ was $t(1 + \frac{z_{\alpha/2}\sqrt{t}}{b})^{-1} \leq \mu \leq t\left(1 - \frac{z_{\alpha/2}\sqrt{t}}{b}\right)^{-1}$. Our approach for a parameter θ will be to match its maximum likelihood estimate $\hat{\theta}$ to *t*, its curvature to $l''(\hat{\theta})$, and its asymmetry to $l'''(\hat{\theta})$.

We find $\hat{\mu} = t$. Then $l''(\hat{\mu}) = -\frac{b^2}{t^3}$ and $l'''(\hat{\mu}) = 6\frac{b^2}{t^4}$. Matching $l(\theta)$ to $l(\mu)$ and solving we find $t = -6\frac{l''(\hat{\theta})}{l'''(\hat{\theta})}$ and $\frac{\sqrt{t}}{b} = \frac{l'''(\hat{\theta})}{6\left(-l''(\hat{\theta})\right)^{3/2}}$. Centering the interval at $\hat{\theta}$ and replacing terms in the best Inverse Gaussian interval, we get our proposed nearly-best Wald interval

$\hat{\theta} - z\alpha/_2 \frac{1}{\sqrt{-l''(\hat{\theta})}}\left(1 - z\alpha/_2 \frac{l'''(\hat{\theta})}{6\left(-l''(\hat{\theta})\right)^{3/2}}\right)^{-1} \leq \theta \leq \hat{\theta} + z\alpha/_2 \frac{1}{\sqrt{-l''(\hat{\theta})}}\Bigg(1 +$

$z\alpha/_2 \frac{l'''(\hat{\theta})}{6\left(-l''(\hat{\theta})\right)^{3/2}}\Bigg)^{-1}$ where the signs are reversed for reversed 3rd derivative.

We see that it is the classical Wald interval with a correction for asymmetry.

Let us revisit the case of estimating Binomial(p) from Terrell (2017), for which we had to condition Poisson variables because of the difficulty of carrying out the required global parameter transformation. The nearly-best Wald method is no longer covariant under transformations, so we will arrive at (slightly) different intervals when we transform p. Perhaps a good choice of parameter is the log-odds $\theta = \log\left(p/_{(1-p)}\right)$, the natural parameter when thinking of it as an exponential family. We find $l(\theta) = x\theta - n\log\left(1 + e^{\theta}\right)$. Then $l'(\theta) = x - n\left(1 - \frac{1}{1+e^{\theta}}\right)$ (so that of course $\hat{\theta} = \log\left(\hat{p}/_{(1-\hat{p})}\right)$ where $\hat{p} = x/_n$. Further $l''(\theta) = -n\frac{e^{\theta}}{\left(1+e^{\theta}\right)^2}$, so that $l''\left(\hat{\theta}\right) = -n\hat{p}(1-\hat{p})$. Then our asymmetry is $l'''(\theta) = -n\left(\frac{e^{\theta}(1-e^{\theta})}{\left(1+e^{\theta}\right)^3}\right)$, so that $l'''\left(\hat{\theta}\right) = -n\hat{p}(1-\hat{p})(1-2\hat{p})$. This gives an interval around θ of $\hat{\theta} \mp \frac{z\alpha/_2}{\sqrt{n\hat{p}(1-\hat{p})} \pm \frac{z\alpha/_2}{6}(1-2\hat{p})}$.

For example, let n = 40 and observe x = 4 successes (for a quite asymmetric likelihood for Binomial p) and nominal coverage .95. The Terrell(2017) method gives $.0317 \leq p \leq .218$ with coverage .946. Our new, more general method gives $.0325 \leq p \leq .216$ with coverage .943. Both method give likelihoods at the boundary equal to 3 significant figures. The difference from the standard Wald interval $.007 \leq p \leq .193$ is striking.

References:

Shuster, J. J. (1968) On the inverse Gaussian distribution function. Journal of the American Statistical Association. 63 pp. 1514-1516.

Terrell, G. R. (2017) Nearly best confidence intervals. Unpublished manuscript.