Right or Wrong, Our Confidence Intervals

A wonderful thing about tenure is that once I had it, I never had to control my irresistible urge to waste my time on the most useless of all things. The other day, a close friend said to me, *but, I was almost right*! I have not the slightest notion why this pedantic remark of a friend made me wonder if our everyday confidence intervals (sets) are almost right even when they are wrong, and squarely right when they are right. At the clear risk of saying things that were all done a long time ago, I want to report a few simple, but perhaps interesting, facts on how right are our confidence sets when they are right, and how wrong are they when they are wrong, and how does the dimension of the problem affect the answers, precisely.

Simplicity has its virtues. So, how about starting with a simple example that we can easily relate to. Take the t interval, say C_n, $\bar{X} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$ for the mean μ of a one dimensional CDF F with a finite variance. Its margin of error is of course $\delta_n = t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$. When our t interval misses the true μ, the amount by which it misses, say d_n, is the distance of μ from the appropriate endpoint of C_n. Expressed in units of the margin of error, the amount by which we miss is $w_n = \frac{d_n}{\delta_n}$; d_n and δ_n both go down at the rate $\sqrt{n}$, and it seemed as though w_n is a better index practically, than simply d_n. I wanted to understand how large w_n is when the t interval fails, for example, what is $E_F(w_n \,|\mu \notin C_n)$.

Of course, I did simulate it first. I simulated for seven choices of $F, N(0,1)$, standard double exponential, $t_3, U[-1,1], \mathrm{Beta}(1/2,1/2), \mathrm{Poisson}(4)$, and χ^2_4, using in each case a simulation size of $8,000$ and $\alpha = .05, n = 50$, a gentle sample size. My simulation averages of w_n (conditioned on failure) in the seven cases were $.23, .18, .19, .22, .21, .20$, and $.24$. I understood the simulations to mean that the $95\%\,t$ interval misses μ by about 20% of the margin of error when it misses. But why are the simulation averages all so tantalizingly close to 20% although the distributions simulated are very different? We must then expect that there is a theorem here. It turns out that whenever F has a finite variance, $E_F(w_n \,|\mu \notin C_n) \to \frac{2\phi(z_{\alpha/2})}{\alpha z_{\alpha/2}} - 1 = .1927$ for $\alpha = .05$, and this explains why my simulation averages all hovered around .2. We can say more; we have, for $w > 0, P_F(w_n > w \,|\mu \notin C_n) \to \frac{2[1-\Phi(z_{\alpha/2}+w)]}{\alpha}$. I will apply this to predicting a US Presidential election in closing. Higher order expressions for $P_F(w_n > w \,|\mu \notin C_n)$ are derivable (in nonlattice cases) by using results in Hall (1987, AOP).

The other side of the coin is how right is the interval when it is right, for example, $E_F(\frac{|\bar{X}-\mu|}{\delta_n} \,|\mu \in C_n)$. And here, it turned out that this converges to $\frac{2[\phi(0)-\phi(z_{\alpha/2})]}{(1-\alpha)z_{\alpha/2}} = .3657$ for $\alpha = .05$; that is, when we succeed, *whatever be our* F, the true μ is about 63% deep inside the interval from its boundary. I will let others decide if these two numbers .1927, .3657 are good or bad.

For the extension to higher dimensions, a little more notation is unavoidable. I let F be a CDF in p-space with a covariance matrix Σ, which I treat as known, and as my confidence set I take the usual (Gaussian) ellipsoid centered at the sample mean and oriented by Σ. The known Σ assumption does not affect first order asymptotics in this problem, if p is held fixed. One can write a formula; $E(w_n \,|\mu \notin C_n) = \frac{\sqrt{2}\Gamma(\frac{p+1}{2})}{\alpha\Gamma(\frac{p}{2})\sqrt{\chi^2_{\alpha,p}}} P(\chi^2_{p+1} > \chi^2_{\alpha,p})$.

Now, the analogous limit result on $E_F(w_n \,|\mu \notin C_n)$ needs a bit more work, as one needs to use higher order Stirling approximations to the Gamma function, and Edgeworth expansions for a χ^2 statistic, and Cornish-Fisher expansions for a χ^2 percentile, and then collect terms. My personal curiosity was about large p, and it turned out that $E(w_n \,|\mu \notin C_n) = \frac{\frac{\phi(z_\alpha)}{\alpha} - z_\alpha}{\sqrt{2p}} + O(p^{-1})$; so, in units of the margin of error, the amount by which the ellipsoid misses when it does goes down with the number of dimensions at the rate $\frac{1}{\sqrt{p}}$. Higher the dimension, when we miss, the true μ is more *just around the corner.*

I close the circle by returning to one dimension. Take the case of predicting a very close US Presidential election. Stratification and nonresponse aside, we are dealing with a binomial p. If we poll $n \geq 6765$ voters, and use a 90% Wald interval, then the pollster may state that the poll's margin of error is at most1%, and in case, by misfortune, the poll is wrong, the true p is within at most another half a percentage point with a 90% probability. Very many public polls use only about 1000 voters. If we poll only 1000 voters, we can claim that our margin of error is at most 2.6%, and in case our poll is wrong, the true p is within at most another 1.5% with a 90% probability.

This story remains the same for essentially all *LAN* problems. The corresponding Bayesian problems are similar. And now, I must find myself some other completely useless thought to keep me entertained!