E0 232 Probability and Statistics **Oct 30, 2012**

Brief Note on Symmetric Confidence Intervals

Shivani Agarwal

Recall that a $(1-\alpha)$ confidence interval for a parameter θ based on data $X_1, \ldots, X_n$ is a (random) interval $[\widehat{\Theta}_n^-, \widehat{\Theta}_n^+]$, where $\widehat{\Theta}_n^- = g^-(X_1, \ldots, X_n)$ and $\widehat{\Theta}_n^+ = g^+(X_1, \ldots, X_n)$ are some functions of the data satisfying

$$\mathbf{P}_\theta\left(\widehat{\Theta}_n^- \le \theta \le \widehat{\Theta}_n^+\right) \;\ge\; 1-\alpha\,.$$

Often, it is natural to construct a symmetric confidence interval $[\widehat{\Theta}_n - \epsilon, \widehat{\Theta}_n + \epsilon]$ around some estimator $\widehat{\Theta}_n$. This reduces the problem to finding $\epsilon > 0$ such that

$$\mathbf{P}_\theta\left(\big|\widehat{\Theta}_n - \theta\big| \ge \epsilon\right) \le \alpha\,.$$

Example 1 (Confidence interval for mean of a bounded random variable)**.** Suppose $X_1, \ldots, X_n$ are iid random variables in some bounded range $[a, b]$ with unknown mean μ. Then a confidence interval for μ using the sample average estimator $\widehat{M}_n = \frac{1}{n}\sum_{i=1}^n X_i$ can be obtained using Hoeffding's inequality. In particular, taking $\epsilon = (b-a)\sqrt{\frac{1}{2n}\ln(\frac{2}{\alpha})}$ yields a $(1-\alpha)$ confidence interval $[\widehat{M}_n - \epsilon, \widehat{M}_n + \epsilon]$ for μ.

Example 2 (Confidence interval derived from an asymptotically normal estimator)**.** Suppose $X_1, \ldots, X_n$ are iid random variables whose distribution is governed by some unknown parameter θ, and $\widehat{\Theta}_n = g(X_1, \ldots, X_n)$ is an asymptotically normal estimator of θ, i.e. that

$$\frac{\widehat{\Theta}_n - \theta}{\sqrt{\mathbf{Var}_\theta(\widehat{\Theta}_n)}} \xrightarrow{D} \mathcal{N}(0,1)\,.$$

Unless $\mathbf{Var}_\theta(\widehat{\Theta}_n)$ is known, this does not directly yield a confidence interval for θ. However, if for some data-based estimator $\widehat{V}_n = h(X_1, \ldots, X_n)$ of $\mathbf{Var}_\theta(\widehat{\Theta}_n)$ we have

$$\frac{\widehat{\Theta}_n - \theta}{\sqrt{\widehat{V}_n}} \xrightarrow{D} \mathcal{N}(0,1)\,,$$

then for large n, this can be used to obtain a normal-based (approximate) confidence interval for θ. In particular, taking $\epsilon = \sqrt{\widehat{V}_n}\cdot z_{\alpha/2}$, where $z_{\alpha/2} = \Phi^{-1}(1-\frac{\alpha}{2})$, yields a $(1-\alpha)$ confidence interval $[\widehat{\Theta}_n - \epsilon, \widehat{\Theta}_n + \epsilon]$ for θ.

Example 3 (Confidence interval for mean of a normal random variable)**.** Suppose $X_1, \ldots, X_n$ are iid normal random variables with unknown mean μ and unknown variance σ^2, and let $\widehat{M}_n = \frac{1}{n}\sum_{i=1}^n X_i$ and $\widehat{S}_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \widehat{M}_n)^2$. Then

$$\frac{\widehat{M}_n - \mu}{\frac{\widehat{S}_n}{\sqrt{n}}} \sim t_{n-1}\,,$$

where t_{n-1} denotes the t distribution with $n-1$ degrees of freedom, and whose CDF Ψ_{n-1} is available in tables similarly to the standard normal CDF tables. In this case, one can obtain a t distribution based confidence interval for μ; in particular, taking $\epsilon = \frac{\widehat{S}_n}{\sqrt{n}} \cdot z'_{\alpha/2}$, where $z'_{\alpha/2} = \Psi_{n-1}^{-1}(1-\frac{\alpha}{2})$, yields a $(1-\alpha)$ confidence interval $[\widehat{M}_n - \epsilon, \widehat{M}_n + \epsilon]$ for μ. For large n, t_{n-1} is well-approximated by the standard normal $\mathcal{N}(0,1)$, and so one can again fall back on a normal-based approximation, but for small n, the t_{n-1}-based confidence interval should be used.