BOOTSTRAP CONFIDENCE INTERVAL

Measures of accuracy of statistical estimates provide a sense of how good an estimate of the truth that the estimate actually is. The statistical estimate of interest in this paper is the optimal split point on age at cochlear implantation. Since this estimate is obtained via binary partitioning, which is a non-parametric analytical method, the measure of accuracy cannot be simply calculated by a closed form formula. Therefore, the bootstrap algorithm, applying uniform weights on the uniquely sampled observations, is used to provide a confidence interval; as outlined in Efron and Tibshirani [1] we employ the bootstrap percentile interval.

The bootstrap sampling units $(x_1^*, x_2^*, \ldots, x_n^*)$, are a random sample of size n drawn with replacement from the population of n sampling units, $(x_1, x_2, \ldots, x_n)$. Thus a bootstrap data set can be expressed as, $\boldsymbol{x}^* = (\boldsymbol{x}_1^*, \boldsymbol{x}_2^*, \ldots, \boldsymbol{x}_n^*)$. Now, define θ to be the statistic of interest, so that the bootstrap replicate $\hat{\theta}^* = s(\boldsymbol{x}^*)$ is computed.

Ideally, when the number of bootstrap replications is infinite, the $1 - 2\alpha$ percentile interval is defined by the α and $1 - \alpha$ percentiles of the cummulative distribution function of $\hat{\theta}^*$,

$$[\hat{\theta}_{\%,lo}, \hat{\theta}_{\%,up}] \quad = \quad [\hat{\theta}^{*(\alpha)}, \hat{\theta}^{*(1-\alpha)}].$$

The percentile interval has probability exactly $1 - 2\alpha$ of containing the true value of θ: $Prob_\theta\{\theta < \hat{\theta}_{\%,lo}\} = \alpha$, and $Prob_\theta\{\theta > \hat{\theta}_{\%,up}\} = \alpha$. In fact, this is called an equal-tailed confidence interval.

Since obtaining an infinite number of bootstrap replications is not practical, we generate a finite number, B, of independent bootstrap data sets $\boldsymbol{x}^{*1}, \boldsymbol{x}^{*2}, \ldots, \boldsymbol{x}^{*B}$ so that replications $\hat{\theta}^*(b) = s(\boldsymbol{x}^{*b})$ are computed, where $b = 1, 2, \ldots, B$. Let $\hat{\theta}_B^{*(\alpha)}$ be the $100 \cdot \alpha th$ empirical percentile of the $\hat{\theta}^*(b)$ values, that is the $B \cdot \alpha th$ value in the ordered list of the B replications of $\hat{\theta}^*$. Likewise, let $\hat{\theta}_B^{*(1-\alpha)}$ be the $100 \cdot (1-\alpha)th$ empirical percentile. Thus, the bootstrap percentile interval can be expressed as,

$$[\hat{\theta}_{\%,lo}, \hat{\theta}_{\%,up}] \quad = \quad [\hat{\theta}_B^{*(\alpha)}, \hat{\theta}_B^{*(1-\alpha)}].$$

References

[1] Efron B, Tibshirani RJ. *An Introduction to the Bootstrap* Chapman and Hall: New York, 1993.