14.2 Example: the spatial test data

Our next example, the *spatial test data*, demonstrates the need for improvements on the percentile and bootstrap-t methods. Twenty-six neurologically impaired children have each taken two tests of spatial perception, called "A" and "B." The data are listed in Table 14.1 and displayed in Figure 14.1. Suppose that we wish to find a 90% central confidence interval for $\theta = \text{var}(A)$, the variance of a random A score.

The plug-in estimate of θ based on the $n = 26$ data pairs $x_i = (A_i, B_i)$ in Table 14.1 is

$$\hat{\theta} = \sum_{i=1}^{n}(A_i - \bar{A})^2/n = 171.5\,, \qquad (\bar{A} = \sum_{1}^{n} A_i/n). \quad (14.1)$$

Notice that this is slightly smaller than the usual unbiased estimate of θ,

$$\bar{\theta} = \sum_{i=1}^{n}(A_i - \bar{A})^2/(n-1) = 178.4. \quad (14.2)$$

The plug-in estimate $\hat{\theta}$ is biased downward. The BC$_a$ method automatically corrects for bias in the plug-in estimate, which is one of its advantages over the percentile method. [1]

A histogram of 2000 bootstrap replications $\hat{\theta}^*$ appears in the left panel of Figure 14.2. The replications are obtained as in Figure 6.1: if $\mathbf{x} = (x_1, x_2, \cdots, x_{26})$ represents the original data set of Table 14.1, where $x_i = (A_i, B_i)$ for $i = 1, 2, \cdots, 26$, then a bootstrap data set $\mathbf{x}^* = (x_1^*, x_2^*, \cdots, x_{26}^*)$ is a random sample of size 26 drawn with replacement from $\{x_1, x_2, \cdots, x_{26}\}$; the bootstrap replication $\hat{\theta}^*$ is the variance of the A components of $\mathbf{x}^*$, with $x_i^* = (A_i^*, B_i^*)$.

$$\hat{\theta}^* = \sum_{i=1}^{n}(A_i^* - \bar{A}^*)^2/n \qquad (\bar{A}^* = \sum_{i=1}^{n} A_i^*/n). \quad (14.3)$$

$B = 2000$ bootstrap samples $\mathbf{x}^*$ gave the 2000 bootstrap replications $\hat{\theta}^*$ in Figure 14.2. [2] These are *nonparametric* bootstrap

[1] The discussion in this chapter, and the algorithms bcanon and abcnon in the Appendix, assume that the statistic is of the plug-in form $\hat{\theta} = t(\hat{F})$.

[2] We don't need the second components of the x_i^* for this particular calculation, see Problem 14.2.