Bootstrapping Dependent Data

One of the key issues confronting bootstrap resampling approximations is how to deal with dependent data. Consider a sequence $\{X_t\}_{t=1}^{n}$ of dependent random variables. Clearly it would be a mistake to resample from the sequence scalar quantities, as the reshuffled resamples would break the temporal dependence. Our goal is most often to learn the variance of a general statistic $T_n(X_1, \ldots, X_n)$, we hereafter refer to the unknown variance as σ^2. The quantity σ^2 may not be calculable analytically because the dependence structure and the underlying distribution of the innovations are not assumed to be known.

In 1985, Hall examined the problem of bootstrap estimation for data that was spatial in character. His proposed methods could be applied to time-series data, although the specific details of his results cannot be directly applied.[1] For the fixed-block bootstrap, he proposes dividing the series into m nonoverlapping blocks of equal length, each block has length $\frac{n}{m}$. For the moving-block bootstrap, he proposes dividing the series into $n-m+1$ overlapping blocks of equal length $\frac{n}{m}$. To fix ideas, consider the sample $\{x_1, \ldots, x_4\}$ with block length 2. The fixed-block bootstrap is obtained by constructing the statistic of interest for each member of the set

$$\{(x_1, x_2), (x_3, x_4)\}.$$

The moving-block bootstrap is obtained by constructing the statistic of interest for each member of the set

$$\{(x_1, x_2), (x_2, x_3), (x_3, x_4)\}.$$

The intuition underpinning the fixed-block bootstrap is as follows. The moving-block bootstrap has many samples that share a large number of observations, in this way there is redundancy. The fixed-block bootstrap avoids such redundancy. Further, if m grows with n, then a statistic constructed from a given subsample will eventually behave as though it is independent of all but two (the adjacent two) of the statistics constructed from the other subsamples. In addition, m should grow with n to allow for long-lived dynamics to be captured. One natural choice for m would be $m = cn$, with $0 < c < 1$, as the subsamples would be of the same order of magnitude as the original data. Unfortunately, such an approach would

[1] To see why consider the case of two dimensional spatial data. Rather than a sequence of time-series variables, the underlying components are rectangles. One assumption is that the ratio of the lengths of two adjoining edges of the rectancle is constant, which has no natural anlaog in time-series data.

not provide enough subsamples, as we have only about $\frac{1}{c}$ subsamples regardless of n. We require that m increase more slowly, so that $\frac{m}{n} \rightarrow 0$.

In 1986, Carlstein independently developed the fixed-block bootstrap for stationary, α-mixing sequences. Formally, let $\{X_t, -\infty < t < \infty\}$ be a strictly stationary sequence defined on probability space (Ω, F, P). The function $T_n(x_1, \ldots, x_n)$, from $\mathbb{R}^1 \longmapsto \mathbb{R}^n$ is defined so that $T_n(X_1(\omega), \ldots, X_n(\omega))$ is F-measurable. Fixed blocks of data are defined as

$$X_m^t = (X_{t+1}, X_{t+2}, \ldots, X_{t+m}),$$

so the whole sample is denoted X_n^0. A general statistic defined for the fixed block is

$$T_m^t = T_m(X_m^t);$$

for example, the sample mean

$$\bar{X}_m^t = \frac{1}{m} \sum_{j=1}^{m} X_{t+j}.$$

The statistic is appropriately standardized, so that for the unknown variance

$$\lim_{n \rightarrow \infty} E\left[n^{\frac{1}{2}}\left(T_n^t - ET_n^0\right)\right]^2 = \sigma^2 \in (0, \infty),$$

which is clear for the case of the sample mean

$$\lim_{n \rightarrow \infty} n\left(\frac{1}{n^2} \sum_{t=1}^{n} E(X_t - \mu)^2\right) = \sigma^2.$$

The value of the statistic for each of the fixed blocks is denoted

$$T_m^{tm} : 0 \leq t \leq k_n - 1,$$

where $k_n = \left[\frac{n}{m}\right]$. For example, let $n = 100$ and $m = 8$, so $k_n = [12.5] = 12$,

$$T_8^0 = T_8(X_1, \ldots, X_8), \; T_8^1 = T_8(X_9, \ldots, X_{16}) \; \ldots, T_8^{11}(X_{89}, \ldots, X_{96}),$$

so the last four observations are not used. To construct the estimator of σ^2, first construct the average value of T across the subsamples

$$\bar{T}_m = \frac{1}{k_n} \sum_{t=0}^{k_n - 1} T_m^{tm}.$$

With the sample average in hand, the variance estimator is simply the standard variance estimator

$$\hat{\sigma}^2_{FBoot} = m \frac{1}{k_n} \sum_{t=0}^{k_n-1} \left(T_m^{tm} - \bar{T}_m\right)^2 .$$

For comparison, consider the variance of the sample mean

$$\hat{\sigma}^2 = nE\left(\bar{X}_n - \mu\right)^2 .$$

Observe that there is no randomness in the construction of $\hat{\sigma}^2_{FBoot}$, the statistic of interest is calculated for each subsample (that is, for each fixed block) and the variance is directly estimated. In this way, as Kunsch (1989) argues, the fixed-block bootstrap is really closer to the jackknife than the moving-block bootstrap. For the jackknife, one deletes each block of m consecutive observations once and calculates the sample variance of the statistics constructed from the $n-m+1$ samples of length $n-m$. Thus the jackknife differs from the fixed-block bootstrap in that overlapping subsamples are used (and that tapering is used to make a smooth transition between observations omitted and observations included). For the arithmetic mean, Kunsch argues that the fixed-block bootstrap and the jackknife are equivalent. For more complicated statistics they are not, and Kunsch argues that the jackknife outperforms the fixed-block bootstrap.

Carlstein shows that if $m_n \to \infty$ and $\frac{m_n}{n} \to 0$, then $\hat{\sigma}^2_{FBoot} \xrightarrow{P} \sigma^2$. How should one choose m in practice? Increasing m reduces bias and captures more persistent dependence. Decreasing m reduces variance as more subsamples are available. The trade-off between bias and variance leads one to consider mean square error as the optimal criterion. Because construction of the MSE depends on knowledge of the underlying data generating process, no optimal results are available. For the special case in which

$$X_t = \beta X_{t-1} + U_t,$$

with $|\beta| < 1$ and $U_t \overset{iid}{\sim} N(0,1)$, the value of the block length that minimizes first-order MSE is $m_n^* = \left(\frac{2|\beta|}{1-\beta^2}\right)^{\frac{2}{3}} n^{\frac{1}{3}}$. Sensibly, the block length increases with the magnitude of β.

In a 1989 paper rich with results, Kunsch explored the moving-block bootstrap (as well as the jackknife, about which we have little to say here). Kunsch is clear that either the moving-block or fixed-block methods are only appropriate for statistics constructed from the empirical distribution function, as Hall makes clear from the outset in his book. To construct the potential blocks of data for the moving-block bootstrap, we again let $X_m^t = (X_{t+1}, X_{t+2}, \ldots, X_{t+m})$ and note that there are $n - m + 1$ possible overlapping blocks. (For the case in which $n = 100$ and $m = 8$, the fixed-block bootstrap used 12 nonoverlapping subsamples, while there are 93 potential (overlapping) blocks for the moving-block bootstrap. Unlike the fixed-block bootstrap, there is randomness for the moving-block bootstrap, as the potential overlap of the blocks does not make clear precisely which subsamples should be used. If we let S_t be a random variable distributed uniformly on the integers $\{0, 1, \ldots, n-m\}$, then the moving-block bootstrap begins by constructing a sample of length km (Kunsch assumes that $km = n$, so we do as well in what follows) as

$$\left\{X_m^{S_1 m}, X_m^{S_2 m}, \ldots, X_m^{S_k m}\right\}.$$

The statistic of interest is calculated for the entire bootstrap sample, rather than from the subsamples as in the fixed-block bootstrap, and is denoted

$$T_{n\square} = T\left(\left\{X_m^{S_1 m}, X_m^{S_2 m}, \ldots, X_m^{S_k m}\right\}\right).$$

The moving-block estimator of the variance is

$$\hat{\sigma}^2_{MBoot} = Var_*\left(T_{n\square}\right) = E_*\left[\left(T_{n\square} - E_* T_{n\square}\right)^2\right],$$

where E_* denotes expectation with respect to $S_1, \ldots, S_k$. Kunsch shows that if $m_n \to \infty$ and $\frac{m_n}{n} \to 0$, then $\hat{\sigma}^2_{MBoot} \xrightarrow{P} \sigma^2$.

There has been no mention of monte carlo resampling. That is because the bootstrap is defined literally as the variance of the statistic constructed from all possible subsamples. In most applications $\hat{\sigma}^2_{MBoot}$ must be evaluated my monte carlo simulation. To illustrate how such a quantity could be calculated without computer simulation, consider estimation of the (arithmetic) mean. We consider estimation of the mean from blocks of length m. Because we are estimating the mean, calculation of the statistic on the entire bootstrap sample is equivalent to calculating the statistic on each block, and averaging the block means

$$T_{n\square} = \frac{1}{k}\sum_{t=1}^{k} W_{n,t},$$

where $W_{n,t}$ is the average from block t. Because each block is equally likely to be sampled, the $W_{n,t}$'s are i.i.d. with

$$P\left(W_{n,t} = \frac{X_{j+1} + \cdots X_{j+m}}{m}\right) = \frac{1}{n-m+1} \text{ for each } j = 0, \ldots, n-m.$$

We have

$$
\begin{aligned}
E\left(T_{n^{\square}} | X_1, \ldots, X_n\right) &= EW_{n,1} = \frac{1}{n-m+1} \sum_{j=0}^{n-m} \frac{1}{m} \sum_{i=1}^{m} X_{j+i} \\
&= \frac{1}{n-m+1} \frac{1}{m} \sum_{t=1}^{n} X_t c_t,
\end{aligned}
$$

where $c_t = [\min(t-1, n-m) - \max(t-m, 0) + 1]$ is a counter that indexes the number of appearances of each X_t in the total sum. For example, X_1 appears in only the first block, so $c_1 = 1$. Similarly, X_2 appears in the first two blocks and $c_2 = 2$. We thus have an analytic expression for the expectation of the moving-block bootstrap estimator of the mean.

Of course, we are typically interested in the variance of the estimator. We have

$$
Var\left(T_{n^{\square}} | X_1, \ldots, X_n\right) = \frac{1}{k} Var\left(W_{n,1}\right) = \frac{1}{k} \frac{1}{n-m+1} \sum_{j=0}^{n-m} \frac{1}{m^2} \sum_{i=1}^{m} \left(X_{j+i} - EW_{n,1}\right)^2,
$$

which provides an analytic expression for the variance of the bootstrap estimator.

References

Carlstein, E., 1986, "The Use of Subseries Values for Estimating the Variance of a General Statistic from a Stationary Sequence" [illegible] 14, 1171-1179.

Hall, P., 1985, "Resampling a Coverage Pattern" [illegible] 20, 231-246.

Künsch, H., 1989, "The Jackknife and the Bootstrap for General Stationary Observations" [illegible] 17, 1217-1241.