Homework 5: Bootstrap Sampling Correlation and Bagging

Di Gravio, Chiara

Tue Mar 31 11:50:14 2020

Suppose x_i, i = 1, . . . , N are iid (μ, σ^2). Let $\bar{x}_1^*$ and $\bar{x}_2^*$ be two bootstrap realizations of the sample mean. Show that the sampling correlation $cor(\bar{x}_1^*, \bar{x}_2^*) = \frac{n}{2n-1} \approx$ 50%. Along the way derive $var(\bar{x}_1^*)$ and the variance of the bagged mean $\bar{x}_{bag}$. Here $\bar{x}$ is a linear statistic; bagging produces no reduction in variance for linear statistics.

First, we compute the variance of $\bar{x}_1^*$ (the variance of $\bar{x}_2^*$ will be the same). Let, $\bar{x}_j^* = \frac{1}{N}\sum_{i=1}^N x_{ji}^*$ where x_{ji}^* are drawn with replacement from $\{x_1, ..., x_N\}$. We know that within each bootstrap sample; thus, conditioning on the xs in the sample:

$$E[\bar{x}_1^*|x_1, ..., x_N] = \frac{1}{N}\sum_{i=1}^N E[x_{1i}^*] = \bar{x} \implies E[\bar{x}_1^*] = E[E[\bar{x}_1^*|x_1, ..., x_N]] = \mu$$

and:

$$Var(\bar{x}_1^*|x_1, ..., x_N) = \frac{Var(x_{1i}^*)}{N} = \frac{1}{N}\frac{\sum_i (x_{1i} - \bar{x}_1)^2}{N} = \frac{(N-1)S^2}{N^2}$$

where $S^2 = \frac{1}{N-1}\sum_{i=1}(x_{1i} - \bar{x}_1)^2$.

Using the law of iterative expectation:

$$\begin{aligned} Var(\bar{x}_1^*) &= E[Var(\bar{x}_1^*|x_1, ..., x_N)] + Var(E[\bar{x}_1^*|x_1, ..., x_N]) = E\left[\frac{(N-1)S^2}{N^2}\right] + Var\left(\bar{x}\right) = \frac{N-1}{N^2}E[S^2] + \frac{\sigma^2}{N} \\ &= \frac{(N-1)\sigma^2 + N\sigma^2}{N^2} = \frac{(2N-1)\sigma^2}{N^2} \end{aligned}$$

The same result holds for $Var(\bar{x}_2^*)$. Second, we compute the $Cov(\bar{x}_1^*, \bar{x}_2^*)$.

$$Cov(\bar{x}_1^*, \bar{x}_2^*) = \frac{1}{N^2}Cov\left(\sum_{i=1}^N x_{1i}^*, \sum_{i=1}^N x_{2i}^*\right) = \frac{1}{N^2}\sum_{i=1}^N Cov(x_{1i}^*, x_{2i}^*) = \frac{\sigma^2}{N}$$

since $Cov(x_{1i}^*, x_{2i}^*)$ is zero if $x_{1i}^* \neq x_{2i}^*$ and σ^2 if $x_{1i}^* = x_{2i}^*$. Putting the variance and covariance together we find:

$$cor(\bar{x}_1^*, \bar{x}_2^*) = \frac{Cov(\bar{x}_1^*, \bar{x}_2^*)}{\sqrt{Var(\bar{x}_1^*)Var(\bar{x}_2^*)}} = \frac{\sigma^2/N}{\frac{(2N-1)\sigma^2}{N^2}} = \frac{N}{2N-1}$$

The variance of the bagged mean $\bar{x}_{bag}$ is given by:

$$
\begin{aligned}
Var(\bar{x}_{bag}) &= Var\left(\frac{1}{2}(\bar{x}_1^* + \bar{x}_2^*)\right) = \frac{1}{4}Var(\bar{x}_1^* + \bar{x}_2^*) = \frac{1}{4}\left[Var(\bar{x}_1^*) + Var(\bar{x}_2^*) + 2Cov(\bar{x}_1^*, \bar{x}_2^*)\right] \\
&= \frac{1}{4}\left[2\frac{(2N-1)\sigma^2}{N^2} + 2\frac{\sigma^2}{N}\right] = \frac{1}{4}\left[\frac{4N\sigma^2 - 2\sigma^2 + 2N\sigma^2}{N^2}\right] = \frac{(3N-1)\sigma^2}{2N^2}
\end{aligned}
$$