Bias Reduction Using the Bootstrap

Find $f_t$ (i.e., $t$) so that

$$
\mathbb{E}(f_t(P, P_n)|P) = 0
$$
or

$$
\mathbb{E}(T(P_n) - \theta(P) + t|P) = 0.
$$

Change the problem to the sample:

$$
\mathbb{E}(T(P_n^{(1)}) - T(P_n) + t_1|P_n) = 0,
$$
whose solution is

$$
t_1 = T(P_n) - \mathbb{E}(T(P_n^{(1)}|P_n),
$$
so the bias-reduced estimate is

$$
T_1 = 2T(P_n) - \mathbb{E}(T(P_n^{(1)}|P_n).$$
Bias Reduction

We may be able to compute $E(T(P_n)|P_n)$. If so, do so.

If not, estimate by Monte Carlo.
Resampling to Estimate Bias

Problem: given a random sample \((x_1, x_2, \ldots, x_n)\) from an unknown distribution with df \(P\), we want to estimate a parameter, \(\theta = \theta(P)\).

One of the most important properties of an estimator, 

\[ T = t(x_1, x_2, \ldots, x_n), \]

is its bias, \(E_P(T) - \theta\). A “good” estimator, of course, has a small bias, probably even zero, i.e., the estimator is unbiased.

The plug-in estimator, \(T = \theta(P_n)\), may or may not be unbiased, but its bias is often small (e.g., the plug-in estimator of \(\sigma^2\); the bias correction is \(n/(n-1)\)).

Of course, without fairly strong assumptions on the underlying distribution, it is unlikely that we would know the bias of an estimator. Resampling methods can be used to estimate the bias.

In particular, the *bootstrap estimate of the bias* is

\[ E_{P_n}(T) - \theta(P_n) \]
The Bootstrap Estimate of the Bias

The bootstrap estimate of the bias is the plug-in estimate of $E_P(T) - \theta$. The plug-in step occurs in two places, for estimating $E_P(T)$ and then for estimating $\theta$.

Consider the simple estimators (both plug-in estimators):

Sample mean:
$$E_{P_n}(T) - \theta(P_n) = 0.$$  
The sample mean is unbiased for the population mean (if the population mean exists).

Sample second central moment:
$$E_{P_n}(T) - \theta(P_n) = -\frac{\sum (x_i - \bar{x})^2}{n}.$$  
This is also what we would want.

This ideal bootstrap estimate must generally be approximated by Monte Carlo simulation.
The Monte Carlo Bootstrap Estimate of the Bias

The ideal estimate is an expected value of a function:

$$E_{P_n}(T) - \theta(P_n);$$

the Monte Carlo estimate of an expected value of a function is the Monte Carlo sample mean of that function.

For the Monte Carlo simulation, we generate a number of bootstrap samples, \((x_1^*, x_2^*, \ldots, x_n^*)\), drawn from the empirical distribution resulting from \((x_1, x_2, \ldots, x_n)\).

Letting \((x_1^{*j}, x_2^{*j}, \ldots, x_n^{*j})\) represent the \(j^{th}\) bootstrap sample, the Monte Carlo estimate of the bootstrap estimate of the bias is

$$\sum t(x_1^{*j}, x_2^{*j}, \ldots, x_n^{*j})/m - \theta(P_n)$$

For the Monte Carlo estimate we can define a resampling vector, \(P^*\), corresponding to each bootstrap sample as the vector of proportions of the elements of the original sample in the given bootstrap sample.
The Resampling Vector

If the bootstrap sample \((x_1^*, x_2^*, x_3^*, x_4^*)\) is really the sample \((x_2, x_2, x_4, x_3)\), the resampling vector \(P^*\) is

\[
(0, 1/2, 1/4, 1/4)
\]

The resampling vector has random components that sum to 1.

The bootstrap replication of the estimator \(T\) is a function of \(P^*\).

The Monte Carlo estimate of the bootstrap estimate of the bias can be improved if the estimator whose bias is being estimated is a plug-in estimator.
The Bootstrap Estimate of the Bias of a Plug-In Estimator

Consider the resampling vector, $P^0 = (1/n, 1/n, \ldots, 1/n)$.

Such a resampling vector corresponds to a permutation of the original sample. If the estimator is a plug-in estimator, then its value is invariant to permutations of the sample; and, in fact,

$$\theta(P^0) = \theta(P_n),$$

so the Monte Carlo estimate of the bootstrap estimate of the bias,

$$\sum t(x_1^*, x_2^*, \ldots, x_n^*)/m - \theta(P_n),$$

can be written as

$$\sum t(x_1^*, x_2^*, \ldots, x_n^*)/m - \theta(P^0).$$

Instead of using $\theta(P^0)$, we can increase the precision of the Monte Carlo estimate by using the individual $P^*$'s actually obtained:

$$\sum t(x_1^*, x_2^*, \ldots, x_n^*)/m - \theta(\sum P^*/m),$$

that is, by using the mean of the resampling vectors. Notice that for an unbiased plug-in estimator, e.g., the sample mean, this quantity is 0.
Variance Reduction in Monte Carlo

The use of $\theta(\bar{P}^*)$ is a type of variance reduction in a Monte Carlo procedure.

Remember that a Monte Carlo procedure is estimating an integral. Suppose the integral is

$$\int (f(x) + g(x)) \, dx = \int f(x) \, dx + \int g(x) \, dx,$$

and suppose we know $\int g(x) \, dx$. What is the best way to do the Monte Carlo, to do the integral of the sum, or just to do the integral of $f(x)$ and add on the known value of the integral of $g(x)$?

What makes one estimator better than another?

Is the variance of $f(X) + g(X)$ smaller than the variance of $f(X)$?

Remember that the variance of the Monte Carlo estimator, $\hat{I}$, of an integral, $\int h(x) \, dx$, is generally proportional to the variance of $h(X)$, where $X$ is a random variable. (There are, of course, different ways of using random variables in the Monte Carlo estimation of the integral.)
Variance Reduction in Monte Carlo

If the objective in Monte Carlo experimentation is to estimate some quantity, just as in any estimation procedure, we want to reduce the variance of our estimator (while preserving its other good qualities).

The basic idea is usually to reduce the problem analytically as far as possible, and then to Monte Carlo what is left.

Beyond that general reduction principle, in Monte Carlo experimentation, there are several possibilities for reducing the variance.
Variance Reduction in Monte Carlo

• judicious use of an auxiliary variable
  – control variates (any correlated variable, either positively or negatively correlated)
  – antithetic variates (in the basic uniform generator)
  – regression methods

• use of probability sampling
  – discrete: stratified sampling
  – continuous: importance sampling
Balanced Resampling

Another way of reducing the variance in Monte Carlo experimentation is to constrain the sampling so that some aspects of the samples reflect precisely some aspects of the population.

What about constraining $\bar{P}^*$ so as to equal $P^0$? This makes $\theta(\bar{P}^*) = \theta(P^0)$, and hopefully makes $\sum t(x_1^1, x_2^1, \ldots, x_n^1)/m$ closer to its expected value, while preserving its correlation with $\theta(\bar{P}^*)$.

This is called balanced resampling.

Hall (1990) has shown that the balanced resampling Monte Carlo estimator of the bootstrap estimator has a bias $O(m^{-1})$, but that the reduced variance generally more than makes up for it.