Non-parametric Bootstrap example

Imagine that we have a set of $n$ random measurements of the height of 100 blades of grass from a lawn and we wish to estimate the true mean height of all blades of grass from that lawn. We put these 100 values into a spreadsheet and label the array "Data".

We need to go through three stages:

1. **Estimate the distribution from the data**

   For the non-parametric Bootstrap, we simply use the frequency distribution of the $n$ data values as our estimate.

   In relative frequency form, a continuous variable would simply be a list of values each with probability $1/n$. The data set plot however shows three peaks with probability $2/n$ because the measuring was done to 1/10mm and therefore three values occurred twice. A finer level of measurement would make each value individual.

2. **Simulate the data collection**

   We now perform a Monte Carlo experiment to replicate the process via which we acquired the data set. Each of the 100 values is replaced by a Discrete Uniform distribution.

3. **Calculate the sample statistic**

   We now run a large number of iterations, each one generating a new Bootstrap replicate and, for each iteration, we calculate the statistic of interest. In this case, we'll calculate the mean, standard deviation, and the difference between the 10th and 90th percentiles. Non-Parametric Bootstrap Model goes through the calculations. The resultant distributions are our uncertainty about the population statistics:

A very important advantage that the Bootstrap has over other statistical methods is the ability to analyze the correlation between uncertainty distributions. For example, the scatter plots below of the generated statistical estimates are correlated. It is therefore very convenient for us that the Bootstrap automatically creates the necessary correlations structure.
The links to the Non-Parametric Bootstrap software specific models are provided here:

Non_Parametric_Bootstrap_Model

To use the frequency distribution mentioned in step 1, we can use the distribution Discrete Uniform that is constructed with Crystal Ball’s Custom Distribution as the best guess estimate of the population distribution:
A cumulative plot of the empirical distribution is as follows:

In step 3, the resultant distributions are our uncertainty about the population statistics:

Graph of mean estimate:

Graph of standard deviation estimate:
To use the frequency distribution mentioned in step 1, we can use the distribution RiskDuniform(Data) as the best guess estimate of the population distribution:
In step 3, the resultant distributions are our uncertainty about the population statistics:

*Graph of mean estimate:*
Graph of standard deviation estimate:

Distribution for Mean (cm)/H5

![Histogram showing distribution of mean measurements with mean at 17.8331]

Graph of percentile range estimate:

Distribution for Standard deviation (cm)/H6

![Histogram showing distribution of standard deviation measurements with mean at 9.78564 and 5% confidence intervals at 8.5616 and 10.9807]
Distribution for 90th percentile - 10th percentile (cm)/H7

Mean=25.66664

5% 20.986 90% 30.13 5%