**Bootstrapping**

Bootstrapping is a method in statistics to estimate a statistic of a population by resampling. There are two main types of bootstrap methods: parametric bootstrap and non-parametric bootstrap.

In parametric bootstrapping, first a parametric model (e.g., normal distribution, Poisson distribution, exponential distribution) is fitted to the sample data. Next, a new sample is generated from the sampled data using the assumed distribution with replacement. Then, the statistic of interest is calculated from the new sample. This process is continued for a large number of iterations. Finally, the statistic of interest is calculated from the distribution of the statistic of the bootstrap samples. Here is an example in Python.

# Parametric bootstrapping

import numpy as np

n\_population = 500 # number of data points

lmbda = 3 # lambda for exponential distribution

population = np.random.exponential(scale = lmbda, size = n\_population)

population\_mean = np.mean(population)

nboot = 1000 # the number of bootstrap samples

mboot = 200 # size of each bootstrap sample

# nboot bootstrap samples using exponential distribution of size mboot)

bootstrap\_samples = np.random.exponential(scale = lmbda, size = (mboot, nboot))

bootstrap\_means =np.mean(bootstrap\_samples, axis=0)

# true and estimated mean of population

true\_mean = lmbda

estimated\_mean = np.mean(bootstrap\_means)

print("True population mean age: ", true\_mean)

print("Estimated population mean age:", estimated\_mean)

The output is:

True population mean age: 3

Estimated population mean age: 2.9962374846284563

Non-parametric bootstrapping is another resampling technique with replacement. Non-parametric bootstrapping does not require any assumptions about the data distribution. It is a powerful tool that allows us to make inferences about the population parameters (e.g., mean, variance) from a finite number of samples. Even when we only have one sample, the bootstrap method provides a good approximation to the true population parameter. Below are the steps of non-parametric bootstrapping:

1. Consider that the sample is the population
2. Take repeated samples from the sample (“population") with replacement
3. Calculate the statistic of interest (e.g., mean)
4. Repeat the above steps for a large number of iterations
5. Create a data distribution of the statistic
6. Calculate the statistic of interest from the distribution

The statistic calculated from the distribution will be a close approximation of the true population statistic. Here is an example in Python.

import numpy as np

# population data

ages **=** [25, 30, 35, 40, 45, 50, 55, 60, 65, 70]

nboot = 1000 # number of bootstrap samples

mboot = 5 # bootstrap sample size

bootstrap\_means = np.zeros(nboot)

# Perform bootstrap sampling

for i in range(nboot):

bootstrap\_sample = np.random.choice(ages, size=mboot, replace=True)

bootstrap\_mean = np.mean(bootstrap\_sample)

bootstrap\_means[i] = bootstrap\_mean

true\_mean = np.mean(ages)

estimated\_mean = np.mean(bootstrap\_means)

print("True population mean age: ", true\_mean)

print("Estimated population mean age:", estimated\_mean)

The output is:

True population mean age: 47.5

Estimated population mean age: 47.619