

STA 3180 Statistical Modelling: Bootstrapping

Lecture Notes on Bootstrapping for STA 3180 Statistical Modelling

Bootstrapping:

Bootstrapping is a statistical technique used to estimate the sampling distribution of an estimator by resampling with replacement from the original sample. It is a non-parametric approach to statistical inference that can be used to estimate standard errors, confidence intervals, and other measures of statistical accuracy associated with an estimator.

Key Concepts:

- **Resampling:** The process of taking random samples from a population with replacement.
- **Sampling Distribution:** The probability distribution of a statistic computed from a sample.
- **Estimator:** A statistic used to estimate a population parameter.
- **Non-parametric:** A type of statistical analysis that does not make assumptions about the underlying distribution of the data.

Definitions:

- **Bootstrapping:** A statistical technique used to estimate the sampling distribution of an estimator by resampling with replacement from the original sample.
- **Bootstrap Sample:** A sample created by randomly selecting observations from the original sample with replacement.
- **Bootstrap Replicate:** A statistic computed from a bootstrap sample.

Rules:

- Bootstrapping should only be used when the sample size is large enough to ensure that the bootstrap samples are representative of the population.
- Bootstrapping should not be used if the original sample is not representative of the population.
- Bootstrapping should not be used if the original sample is not randomly selected.

Examples:

Suppose we are interested in estimating the mean height of adult males in the United States. We take a random sample of 100 adult males and measure their heights. To estimate the mean height of all adult males in the United States, we can use bootstrapping. We can create a bootstrap sample by randomly selecting observations from the original sample with replacement. We can then compute the mean height of the bootstrap sample and repeat this process many times to create multiple bootstrap replicates. We can then use the distribution of the bootstrap replicates to estimate the sampling distribution of the mean height and compute a confidence interval for the mean height of all adult males in the United States.