Comparing Methods for Constructing Confidence Intervals

A 95% Coverage Analysis of Four Approaches

Blessing Ofori-Atta

Introduction

Nonparametric bootstrap sampling provides a robust alternative to traditional parametric methods for statistical inference, particularly when the underlying distribution of the data is unknown or difficult to specify. Unlike parametric approaches, which rely on assumptions about the distributional form (e.g., normality), nonparametric bootstrap methods use the observed data itself to approximate the sampling distribution of a statistic. Despite these advantages, parametric methods, such as the Wald-based confidence interval, remain widely used in practice. When distributional assumptions are reasonably accurate and sample sizes are sufficiently large, parametric approaches can provide reliable results with greater computational efficiency. The Wald-based confidence interval, derived from inverting a large-sample Wald test, serves as a prime example of this. Efron (1979)[1] re-introduced the bootstrap principle, including the percentile and bootstrap-t methods. Hall (1987)[2] later proposed the symmetric bootstrap-t method, which modifies the bootstrap-t approach to produce symmetric confidence intervals. The percentile method chooses appropriate quantiles of an ordered bootstrap sample of coefficient estimates to form the interval, while the bootstrap-t method chooses appropriate quantiles of an ordered bootstrap sample of t-statistics to use as critical values in constructing the interval (Elias (2015)[3]). For more complex statistics without a simple standard error formula, a bootstrap estimate of the standard error can be computed for each bootstrap sample, requiring two levels of nested bootstrap sampling. The inner loop is used to estimate the standard error of the statistic of interest, while the outer loop is used to estimate the distribution of the statistic. This study compares variations of the symmetric bootstrap-t method—one using a plug-in estimate of the standard error and the other using a bootstrap estimate—with the percentile and Wald methods for constructing 95% confidence intervals for the population mean. The performance of these methods is evaluated through a simulation study across various sample sizes, using coverage rate as the metric. Coverage rate is the proportion of times the true parameter falls within the confidence interval, based on a specified number of replications, $R \gg 1$.

Description of the Solution Plan

To investigate the coverage rate of nominal 95% confidence intervals by simulating *mcrep* data sets of size n from a normal distribution with a known mean μ and standard deviation σ. If the population is not normally distributed, these methods still work.

Symmetric Bootstrap-t (Plug-in Estimate of Standard Error)

1. Generate a bootstrap sample $X_1^\star, \ldots, X_n^*$.
2. Compute $\hat{\theta}^\star(b)$, $\hat{se}^\star(b)$, and
$$T^\star(b) = \frac{\hat{\theta}^\star(b) - \hat{\theta}}{\hat{se}^\star(b)}$$
based on the bootstrap sample generated in step 1.
3. Repeat steps 1 and 2 B times to obtain $T^\star(b), b = 1, \ldots, B$.

Results

```
library(MASS)
```