Chapter 7: Monte Carlo Methods in Inference

Monte Carlo methods may refer to any method in statistical inference or numerical analysis were simulation is used.

We have so far learned about Monte Carlo methods for estimation.

We will now look at Monte Carlo methods to estimate coverage probability for confidence intervals, Type I error of a test procedure, and power of a test.

In statistical inference there is uncertainty in an estimate. We will use repeated sampling (Monte Carlo methods) from a given probability model to investigate this uncertainty.

1 Monte Carlo Estimate of Coverage

1.1 Confidence Intervals

Recall from your intro stats class that a 95 confidence interval for μ (when σ is known and $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$) is of the form

Interpretation:

Comments:

1.

2.

Mathematical interpretation:

Definition 1.1 For $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, σ known, the $(1-\alpha)100$ *confidence interval for* μ is

$$\left(\overline{x} - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \overline{x} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right),$$

where

$$z_{1-\frac{\alpha}{2}} = 1 - \frac{\alpha}{2} \text{ quantile of } N(0,1).$$

In general,

So, if we have formulas for L and U, we can use Monte Carlo integration to estimate α.

An estimate of $1-\alpha$ tells us about the behavior of our estimator $[L, U]$ in practice.

1.2 Vocabulary

We say $P(L < \theta < U) = P(\text{CI contains } \theta) = 1 - \alpha$.

$1 - \alpha =$

$1 - \hat{\alpha} =$

1.3 Algorithm

Let $X \sim F_X$ and θ is the parameter of interest.

Example 1.1

Consider a confidence interval for θ, $C = [L, U]$.

Then, a Monte Carlo Estimator of Coverage could be obtained with the following algorithm.

1.4 Motivation

Why do we want empirical and nominal coverage to match?

Example 1.2 Estimates of $[L, U]$ are biased.

Example 1.3 Estimates of $[L, U]$ have variance that is smaller than it should be.

Example 1.4 Estimates of $[L, U]$ have variance that is larger than it should be.

Your Turn

We want to examine empirical coverage for confidence intervals of the mean.

1. Coverage for CI for μ when σ is known, $\left(\overline{x} - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \overline{x} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$.

    a. Simulate $X_1, \dots, X_n \overset{iid}{\sim} N(0, 1)$. Compute the empirical coverage for a 95 confidence interval for $n = 5$ using $m = 1000$ MC samples.

    b. Plot 100 confidence intervals using `geom_segment()` and add a line indicating the true value for $\mu = 0$. Color your intervals by if they contain μ or not.

    c. Repeat the Monte Carlo estimate of coverage 100 times. Plot the distribution of the results. This is the Monte Carlo estimate of the distribution of the coverage.

2. Repeat part 1 but without σ known. Now you will plug in an estimage for σ (using `sd()`) when you estimate the CI using the same formula that assumes σ known. What happens to the empirical coverage? What can we do to improve the coverage? Now increase n. What happens to coverage?

3. Repeat 2a. when the data are distributed $\text{Unif}[-1, 1]$ and variance unknown. What happens to the coverage? What can we do to improve coverage in this case and why?