



Next: [More about the theoretical](#) Up: [Lectures](#) Previous: [Balanced Bootstraps](#)

## Subsections

- [What is a Monte Carlo Method?](#)
  - [Antithetic Resampling](#)
  - [Importance Sampling](#)

# Monte Carlo

This is the method used for drawing a sample at random from the empirical distribution. I will start by giving a history and general remarks about Monte Carlo methods for those who have never studied them before.

## What is a Monte Carlo Method?

There is not necessarily a random component in the original problem that one wants to solve, usually a problem for which there is no analytical solution. An unknown parameter (deterministic) is expressed as a parameter of some random distribution, that is then simulated. The oldest well-known example is that of the estimation of  $\pi$  by Buffon, in his needle on the floorboards experiment, where supposing a needle of the same length as the width between cracks we have:

$$p(\text{needle crosses crack}) = \frac{2}{\pi} \text{ implying } \hat{\pi} = 2 \frac{\# \text{tries}}{\# \text{hits}}$$

In physics and statistics many of the problems Monte Carlo is used on is under the form of the estimate of an integral unknown in closed form:

$$\theta = \int_0^1 f(u) du \text{ which can be seen as the evaluation of } Ef(u), \text{ where } u \sim U(0,1)$$

1. The crude, or mean-value Monte Carlo method thus proposes to generate  $B$  numbers uniformly from  $(0,1)$  and take their average: to estimate  $\theta$ ,

$$\hat{\theta} = \frac{1}{B} \sum_{b=1}^B f(u_b)$$

2. The hit-or-miss Monte Carlo method generates random points in a bounded rectangle and counts the number of 'hits' or points that are in the region whose area we want to evaluate.

$$\hat{\theta} = \frac{\#hits}{\#total}$$

Which estimate is better?

This is similar to comparing statistical estimators in general.

There are certain desirable properties that we want estimators to have, consistency which ensures that as the sample size increases the estimates converges to the right answer is ensured here by the properties of Riemann sums. Other properties of interest are:

- Unbiasedness.
- Minimal Variance.

Both the above methods are unbiased, that is when repeated many times their average values are centred in the actual value  $\theta$ .

$$E(\hat{\theta}) = \theta$$

So the choice between them lies in finding the one which has the less variance. The heuristic I developed in class to see that the hit-and-miss has a higher variance is based on the idea that the variance comes from the added randomness of generating both coordinates at random, instead of just the abscissae in the crude Monte Carlo.

More precisely, the variance of crude Monte Carlo is

$$\sigma_M^2 = \frac{1}{n} \int_0^1 (f(u) - \theta)^2 du = \frac{E(f(u) - \theta)^2}{n} = \frac{1}{n} E(f(u)^2) - \frac{\theta^2}{n}$$

and that of hit and miss Monte Carlo, which is just a Binomial(n,) is:

$$\sigma_H^2 = \frac{\theta(1 - \theta)}{n}$$

The difference between these two variances is always positive:

$$\sigma_M^2 - \sigma_H^2 = \frac{1}{n} \int_0^1 f(u)(1 - f(u)) du > 0$$

Most improvements to Monte Carlo methods are variance-reduction techniques.

## Antithetic Resampling

Suppose we have two random variables that provide estimators for  $\theta$ ,  $X$  and  $Y$ , that they have the same variance but that they are negatively correlated, then  $\frac{1}{2}(X + Y)$  will provide a better estimate for  $\theta$  because its variance will be smaller.

This is the idea in antithetic resampling (see Hall, 1989). Suppose we are interested with a real-valued parameter, and that we have ordered our original sample  $x_1 < x_2 \cdots x_n$ , for each resample

$\mathcal{X}^* = \{x_{j_1}, x_{j_2}, \dots, x_{j_n}\}$  and statistic  $s(\mathcal{X}^*)$  we associate  $\mathcal{X}^{**}$  by taking a special permutation of the  $j_i$ 's that will make  $\text{cov}(\mathcal{X}^{**}, s(\mathcal{X}^*)) < 0$ , and as small as possible. If the statistic is a smooth function of mean for instance, then the 'reversal permutation' that maps 1 into  $n$ , 2 into  $n - 1$ , etc... is the best, the small sample values are transformed into the larger observations, and the average of these two estimates will give an estimate with smaller variance.

## Importance Sampling

This is often used in simulation, and is a method to work around the small area problem. If we want to evaluate tail probabilities, or very small areas, we may have very few hits of our random number generator in that area. However we can modify the random number generator, make that area more likely as long as we take that into account when we do the summation. Importance sampling is based on the equalities:

$$\int_0^1 f(u) du = \int_0^1 \frac{f}{g}(u) g(u) du = \int_0^1 \frac{f}{g}(u) dG(u)$$



**Next:** [More about the theoretical](#) **Up:** [Lectures](#) **Previous:** [Balanced Bootstraps](#)  
 Susan Holmes 2004-05-19