# Monte Carlo Methods

## Module DES130: Computational Statistics

Dr. Erick A. Chacón Montalván (echacon@uni.edu.pe)

Escuela de Profesional de Ingeniería Estadística
Facultad de Ingeniería Económica, Estadística y Ciencias Sociales
Universidad Nacional de Ingeniería (UNI)
Lima – Perú

# Introduction

# Monte Carlo methods

**Definition**

Monte Carlo methods are a family of methods that use simulated random numbers to estimate some function or value associated to a probability distribution [@gentle2002elements].

**Note**

It can also be used to estimate deterministic functions by transforming a deterministic problem into an stochastic problem.

# Monte Carlo experiments

**Pseudo-algorithm**

1. Define the domain of the problem.
2. Generate *m* random values from a probability distribution.
3. Perform a deterministic computations using the generated random values.
4. Summarize the results.

**Note**

The quality of the inference depends of the number of random values *m*.

# Example to estimate $\pi$

Consider a circle around $(0, 0)$ with radio $r$ and an square región $[-r, r] \times [-r, r]$. Let $p$ be the probability of a point lying inside the cirle, then

$$p = \frac{\pi r^2}{(2r)^2} = \frac{\pi}{4} \qquad \Rightarrow \qquad \hat{\pi} = 4\hat{p}.$$

If we draw random points uniformly inside the square region,

$$\hat{p} = \frac{\text{\# points in the circle}}{\text{\# total points}} \quad \Rightarrow \quad \hat{\pi} = 4\left(\frac{\text{\# points in the circle}}{\text{\# total points}}\right).$$

# Example to estimate $\pi$

**Pseudo-algorithm**

1. Define the value of $r$.
2. Generate $m$ random values uniformly in $[-r, r] \times [-r, r]$.
3. Compute the number of points ($q$) inside the circle.
4. Compute $\hat{\pi} = 4\left(\frac{q}{m}\right)$.

# Motivation

- **Integration**: approximate integral in high dimensions.
- **Optimization**: optimize non-tractable functions.
- **Hypothesis testing**: perform statistical test for complex statistics.
- **Random number generation**: generate random numbers of high dimensional and complex distributions.
- **Simulation**: obtain properties of complex systems.

# Monte Carlo integration

# Definition

Let consider a quantity of interest $\mu$ which can be writen as the mean of $h(X)$ where $X$ is a random variable with density function $f(x)$,

$$\mu = \int h(x)f(x)dx.$$

Considering the random sample $X_1, X_2, \ldots, X_m$ of size $m$ with density $f(x)$, we define the Monte Carlo estimator

$$\hat{\mu}_{MC} = \frac{\sum_{i=1}^{m} h(X_i)}{m}.$$

The estimator improves as long as $m \to \infty$. A short overview can be fount at (Johansen 2010).

# Justification

The justification for the Monte Carlo estimator uses the strong law of large numbers.

**Strong law of large numbers**

$$\Pr\left(\lim_{m\to\infty} \frac{1}{m}\sum_{i=1}^{\infty} h(X_i) = \mathbb{E}\left[h(X)\right]\right) = 1.$$

In addition, the use of the central limit theorem implies that the Monte Carlo estimator is asimptotically normally distributed.

**Central limit theorem**

$$\lim_{m\to\infty} \sqrt{m}\left(\frac{1}{m}\sum_{i=1}^{\infty} h(X_i) - \mathbb{E}\left[h(X)\right]\right) \to N(0, \sigma^2)$$

# Example to estimate $\pi$

Let $p$ be the probability of a point lying inside the circle, then

$$
\begin{aligned}
\pi &= 4p \\
\pi &= 4\frac{\int_{x^2+y^2\leq r^2} 1 dxdy}{4r^2} \\
\pi &= 4\frac{\int_{-r}^{r}\int_{-r}^{r} I(x^2+y^2 \leq r^2)dxdy}{4r^2} \\
\pi &= 4\int_{-r}^{r}\int_{-r}^{r} I(x^2+y^2 \leq r^2)f(x,y)dxdy.
\end{aligned}
$$

Where $f(x, y) = 1/4r^2$ for $x \in [-1, 1]$ and $y \in= [-1, 1]$. If we draw random $m$ points $(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)$, then

$$\hat{\pi}_{MC} = 4\frac{\sum_{i=1}^{m} I(x_i^2 + y_i^2 \leq r^2)}{m}.$$

# Variance of the Monte Carlo estimator

The central limit theorem implies that $V[\hat{\mu}_{MC}] = \sigma^2/m$, such as the variance of the Monte Carlo estimator is usually considered as

$$\hat{V}[\hat{\mu}_{MC}] = \frac{\sum_{i=1}^{m}(g(x_i) - \hat{\mu}_{MC})^2}{m(m-1)}.$$

## Sampling importance resampling algorithm

Let consider a density function of interest $f(x)$ and $g(x)$ a density easy to generate random values. Then weights and standardized weights are defined as

$$w^*(x_i) = f(x_i)/g(x_i), \qquad w(x_i) = w^*(x_i)/\sum_{i=1}^{m} w^*(x_i).$$

**Algorithm**

1. Sample candidates $y_1, y_2, \ldots, y_n$ from $g(\cdot)$.
2. Compute the standardized importance weights $w(y_1), \ldots, w(y_m)$.
3. Resample $x_1, \ldots, x_n$ from $y_1, y_2, \ldots, y_n$ with probabilities $w(y_1), \ldots, w(y_m)$.

See Givens and Hoeting (2012).

# Sampling importance resampling algorithm

- Rejection sampling is perfect as long a $n \to \infty$.
- Importance sampling used pre-defined number of values to generate a sample of size $n$.
- For convergence, it should hold that $n/m \to 0$.
- SIR is is sensitive to the choice of $g(x)$. It should have heavier tails then $f(x)$ to ensure that $f(x)/g(x)$ is never too large.

# Monte Carlo integration with importance sampling

Consider the Monte Carlo integration

$$\mu = \int h(x)f(x)dx.$$

We can introduce a *importance sampling function* $g(x)$ such as

$$\mu = \int h(x)\frac{f(x)}{g(x)}g(x)dx = \int h(x)w^*(x)g(x)dx$$

where $w^*(x) = f(x)/g(x)$ are known as weights or *importance ratios*. Hence the importance sampling estimator is defined as

$$\hat{\mu}^*_{IS} = \frac{\sum_{i=1}^m h(X_i)w^*(X_i)}{n}.$$

$\hat{\mu}^*_{IS}$ is an unbiased estimator or $\mu$.

# Monte Carlo integration with importance sampling

Similar to the previous estimator, we introduce $g(x)$ as

$$\mu = \frac{\int h(x)f(x)dx}{\int f(x)dx} = \frac{\int h(x)[f(x)/g(x)]g(x)dx}{\int [f(x)/g(x)]g(x)dx}$$

$$\mu = \frac{\int h(x)w^*(x)g(x)dx}{\int w^*(x)g(x)dx}.$$

Then the importance sampling estimator is defined as

$$\hat{\mu}_{IS} = \frac{\sum_{i=1}^m h(X_i)w^*(X_i)}{\sum_{i=1}^m w^*(X_i)}.$$

Defining $w(X_i) = w^*(X_i)/\sum_{i=1}^m w^*(X_i)$ as the standardized weights, we obtain that

$$\hat{\mu}_{IS} = \sum_{i=1}^m h(X_i)w(X_i).$$

$\hat{\mu}_{IS}^*$ is a biased estimator or $\mu$ and have lower variance when $w*(X)$ and $g(X)w^*(X)$ are strongly correlated.

# Control Variates

Another way to improve the MC estimator of $\mu = \mathbb{E}\left[h(X)\right]$ is by using the estimator of a known integral $\theta = \mathbb{E}\left[c(Y)\right]$. Asumming that $\text{Cov}\left[X_i, X_j\right] = \text{Cov}\left[Y_i, Y_j\right] = \text{Cov}\left[X_i, Y_j\right] = 0$,

$$\hat{\mu}_{CV} = \hat{\mu}_{MC} + \lambda(\hat{\theta}_{MC} - \theta).$$

The variance is

$$\mathbb{V}\left[\hat{\mu}_{CV}\right] = \mathbb{V}\left[\hat{\mu}_{MC}\right] + \lambda^2\mathbb{V}\left[\hat{\theta}_{MC}\right] + 2\lambda\text{Cov}\left[\hat{\mu}_{MC}, \hat{\theta}_{MC}\right],$$

which is minimized by

$$\lambda = \frac{-\text{Cov}\left[\hat{\mu}_{MC}, \hat{\theta}_{MC}\right]}{\mathbb{V}\left[\hat{\theta}_{MC}\right]}.$$

The values of $\lambda$ can be estimated.

# Control Variates

The following estimators are used for $\hat{\lambda}$.

$$\mathbb{V}\left[\hat{\theta}_{MC}\right] \approx \frac{1}{m}\frac{\sum_{i=1}^{m}(c(Y_i)-\bar{c})^2}{m-1}.$$

$$\text{Cov}\left[\hat{\mu}_{MC}, \hat{\theta}_{MC}\right] \approx \frac{1}{m}\frac{\sum_{i=1}^{m}(h(X_i)-\bar{h})(c(Y_i)-\bar{c})}{m-1}.$$

Note that $\hat{\mu}_{CV}$ have lower variance than $\hat{\mu}_{MC}$.

$$\mathbb{V}\left[\hat{\mu}_{CV}\right] \approx \mathbb{V}\left[\hat{\mu}_{MC}\right] - \frac{\text{Cov}\left[\hat{\mu}_{MC}, \hat{\theta}_{MC}\right]^2}{\mathbb{V}\left[\hat{\theta}_{MC}\right]}.$$

# Monte Carlo for optimization

# Optimization problem

**Minimization**

$$\text{Find } x_0 \in A \ / \ S(x_o) \leq S(x) \text{ for all } x \in A.$$

where $S(x)$ is un unknown function $\mathbb{E}\left[\tilde{S}(x,\xi)\right]$ where $\xi$ is a random variable and $\tilde{S}$ is a known function.

- We need to optimize a function $S(x)$ that is unknown.
- We can approximate $\nabla S(x)$ and use standard optimization algorithms.
- Another approach would be to approximate $S(x)$ directly.
- This problem is common in models with random effects.

# Stochastic approximation

**Algorithm**

1. Initialize $x_1$
2. Estimate the $\triangledown S(x_t)$
3. Determine a step size $\beta_t$
4. Set $x_{t+1} = x_t - \beta_t \hat{\triangledown} S(x_t)$

This algorithm requires that the series $\{\beta_t\}$ should hold the following conditions:

$$\sum_{t=1}^{\infty} \beta_t = \infty, \sum_{t=1}^{\infty} \beta_t^2 < \infty.$$

# Another approach

**Minimization:**

$$\text{Find } x_0 \in A \ / \ \hat{S}(x_o) \leq \hat{S}(x) \text{ for all } x \in A.$$

where $S(x)$ is un unknown function $\mathbb{E}\left[\tilde{S}(x,\xi)\right]$ where $\xi$ is a random variable and $\tilde{S}$ is a known function.

$$\hat{S}(x) = \frac{\sum_{i=1}^{m} \tilde{S}(x,\xi)}{m}.$$

It becomes a deterministic approach.

# Monte Carlo maximum likelihood

Consider the following model with random effects:

$$Y_i \mid Z_i \sim f_\theta(y_i)$$
$$Z_i \sim \pi(z_i).$$

Then the likelihood is defined as

$$L(\theta) = \int L(\theta \mid \boldsymbol{z})\pi(\boldsymbol{z})d\boldsymbol{z}.$$

The estimation involves

$$\max L(\theta) = \max \int L(\theta \mid \boldsymbol{z})\pi(\boldsymbol{z})d\boldsymbol{z}.$$

Where the likelihood function can be estimated as

$$\hat{L}_{MC}(\theta) = \sum_{i=1}^{m} L(\theta \mid \boldsymbol{z}_i)/m.$$

# Tests and sampling

# Upcoming topics

**Hypothesis testing**

Next class.

**Sampling (MCMC)**

Week 6 and 8.

# References

# References


Givens, Geof H., and Jennifer A. Hoeting. 2012. *Computational Statistics*. John Wiley & Sons.

Johansen, A. M. 2010. “Monte Carlo Methods.” In *International Encyclopedia of Education (Third Edition)*, edited by Penelope Peterson, Eva Baker, and Barry McGaw, 296–303. Oxford: Elsevier. https://doi.org/10.1016/B978-0-08-044894-7.01543-8.