MONTE CARLO METHOD AND VARIANCE REDUCTION TECHNIQUES.

Domingo Morales González

d.morales@umh.es

Universidad Miguel Hernández de Elche

- Monte Carlo integration: estimate by simulation expectations of the form

$$\theta = E[\phi(X)] = E[\phi(X_1, \ldots, X_k)],$$

where

1. $\phi : (\boldsymbol{R}^k, B_k) \longrightarrow (\boldsymbol{R}, B)$ is measurable.
2. $X = (X_1, \ldots, X_k)$ is a random variable with dimension k.

- **Estimation procedure:** Given the simulated sample

$$X^{(1)} = (X_1^{(1)}, \ldots, X_k^{(1)}), \ldots, X^{(n)} = (X_1^{(n)}, \ldots, X_k^{(n)}),$$

estimator is

$$\widehat{\theta} = \frac{1}{n} \sum_{i=1}^{n} \phi(X^{(i)}).$$

- By applying the Central Limit Theorem, we get

$$\widehat{\theta} \sim N(\mu = \theta, \sigma^2 = \frac{\tau}{n}).$$

- A 95% confidence interval of θ is

$$\widehat{\theta} \pm 1.96\frac{\sqrt{\tau}}{\sqrt{n}}.$$

Remark:

1. The width of the interval is proportional to $\frac{1}{\sqrt{n}}$.

2. To reduce width to the its half, we have multiply the number of observations by 4.

3. Variance reduction techniques try to reduce the value of τ.

4. An adequate design of a simulation experiment can improve the precision of estimations for a given computational cost and viceversa.

Example: Estimate the integral

$$\int_0^1 \sqrt{1-x^2}\,dx.$$

Note: The target of the presented examples is to analyze the precision of estimators. For this reason we will consider integrals with known numerical value. In this example, numerical value is $\frac{\pi}{4}$.

- Integral is the area in the first quadrant of the unity circle.
- We generate uniform random points in the unity square $(0,1) \times (0,1)$.

- n is the number of generated points.
- R is the number of points in the region A under the curve $y = \sqrt{1 - x^2}$ (number of hits).
- R/n is the estimation of the probability of hitting the region A.
- $\widetilde{\pi} = 4R/n$ is the estimate of π.
- $V\left[\widetilde{\pi}\right] = 2.697/n$

- Let $a < b \in \boldsymbol{R}$.
- Let $\phi : (a, b) \longrightarrow \boldsymbol{R}^+$ be a measurable function such that $0 \leq \phi(x) \leq c$ for all $x \in (a, b)$.

We want to estimate

$$\theta = \int_a^b \phi(x)\, dx;$$

this is to say, the area under the curve $\phi(x)$.

- Generate pairs (U, V), where $U \sim \mathcal{U}(a, b)$ y $V \sim \mathcal{U}(0, c)$.
- R is the number of points (U, V) under the curve $\phi(x)$.
- n is the number of generated points.

The hit or miss Monte Carlo estimators is

$$\tilde{\theta} = c(b-a)\widehat{P}(V \leq \phi(U)) = c(b-a)\frac{R}{n}.$$

- $E[\tilde{\theta}] = \theta$
- $V[\tilde{\theta}] = \frac{\theta}{n}[c(b-a) - \theta]$

Let $\phi : (\boldsymbol{R}^k, \mathcal{B}_k) \longrightarrow (\boldsymbol{R}, \mathcal{B})$ be measurable. The pure Monte Carlo method estimates

$$\theta = E[\phi(X)] = \int_{\boldsymbol{R}^k} \phi(x) f(x)\, dx$$

with

$$\widehat{\theta} = \frac{1}{n} \sum_{i=1}^{n} \phi(X^{(i)}),$$

where $X^{(1)}, \ldots, X^{(n)}$ is a random sample simulated from $X = (X_1, \ldots, X_n) \sim f(\cdot)$.

Corollary.

- $E[\widehat{\theta}] = \theta$
- $V[\widehat{\theta}] = \frac{1}{n} \int_{\boldsymbol{R}^n} (\phi(x) - \theta)^2 f(x)\, dx$

Example ($I = \int_0^1 \sqrt{1-x^2}\,dx$, continuation).

- $\int_0^1 \sqrt{1-x^2} \cdot f_{\mathcal{U}(0,1)}(x) \cdot dx = E[\sqrt{1-U^2}]$
- $U \sim \mathcal{U}(0,1)$.
- Let $U_1, \ldots, U_n$ be i.i.d. $\mathcal{U}(0,1)$.
- $\widehat{I} = \frac{1}{n}\sum_{i=1}^{n}\sqrt{1-U_i^2}$
- $V[\widehat{I}] = 0.0498/n$.
- $\widehat{\pi} = 4\widehat{I}$
- $V[\widehat{\pi}] = 0.7968/n$.

Remark. The same amount of generated random numbers $\mathcal{U}(0,1)$ should be considered when comparing simulation methods.

$$V[\text{Hit or Miss Monte Carlo}] = \frac{2.697}{n}$$

$$V[\text{Pure Monte Carlo}] = \frac{0.7968}{2n} = \frac{0.398}{n}.$$

For hit or miss Monte Carlo the estimator variance is approximately 7 times greater than for pure Monte Carlo.

Proposition. Let $\phi : (a,b) \longrightarrow \boldsymbol{R}^{+}$,

$0 \leq \phi(\cdot) \leq c$, $\theta = \int_a^b \phi(x)\,dx$. Then

$$V[\widehat{\theta}] \leq V[\tilde{\theta}]; \qquad \text{with} \quad \text{“}=\text{”} \Longleftrightarrow \phi(\cdot) \equiv c.$$

Conclusion. Hit or miss Monte Carlo **never** should be used, because its variance is always greater than the one of pure Monte Carlo.

Problem: estimate

$$\theta = E[\phi(X)] = E[\phi(X_1, \ldots, X_k)];$$

Basic solution:

- $Y_1 = \phi(X_1^{(1)}, \ldots, X_k^{(1)}), \ldots, Y_n = \phi(X_1^{(n)}, \ldots, X_k^{(n)})$
- $\widehat{\theta} = \bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$
- $E[\bar{Y}] = \theta$
- $V[\bar{Y}] = \frac{V[Y_i]}{n}$

Target: Reduce $V[\bar{Y}]$.

Idea: Let Y_1 e Y_2 be identically distributed random variables with $E[Y_i] = \theta$. Then

$$\begin{aligned} E\left[\frac{Y_1+Y_2}{2}\right] &= \theta, \\ V\left[\frac{Y_1+Y_2}{2}\right] &= \frac{1}{2}V[Y_1] + \frac{1}{2}\mathrm{Cov}(Y_1, Y_2). \end{aligned}$$

Conclusion: It is preferred that Y_1 and Y_2 be negatively correlated than they be independent.

Definition.

- Let $X_1, \ldots, X_n$ be independent random variables such that $X_i \sim F_i$.
- $X_i = F_i^{-1}(U_i)$, where $U_1, \ldots, U_k$ i.i.d. $\mathcal{U}(0,1)$.

Y_1 and Y_2 are **antithetic** if and only if

$$\begin{aligned} Y_1 &= \phi\left(F_1^{-1}(U_1), \ldots, F_k^{-1}(U_k)\right) \\ Y_2 &= \phi\left(F_1^{-1}(1-U_1), \ldots, F_k^{-1}(1-U_k)\right), \end{aligned}$$

where $\phi : (\boldsymbol{R}^k, \mathcal{B}_k) \longrightarrow (\boldsymbol{R}, \mathcal{B})$:

- measurable
- be monotonously increasing (or decreasing) in all its arguments.

Corollary. If Y_1 and Y_2 are antithetic, then they are identically distributed and dependent.

Proposition. If Y_1 and Y_2 are antithetic, then $\mathrm{Cov}(Y_1, Y_2) \leq 0$.

Definition. The **antithetic estimator** of θ is

$$\widehat{\theta}_A = \frac{1}{2n} \sum_{i=1}^{n} \left(Y_i^{(1)} + Y_i^{(2)} \right),$$

$$\begin{aligned} Y_i^{(1)} &= \phi(F_1^{-1}(U_1^{(i)}), \ldots, F_k^{-1}(U_k^{(i)})) \\ Y_i^{(2)} &= \phi(F_1^{-1}(1 - U_1^{(i)}), \ldots, F_k^{-1}(1 - U_k^{(i)})) \end{aligned}$$

$U_1^{(1)}, \ldots, U_k^{(1)}, U_1^{(2)}, \ldots, U_k^{(2)}, \ldots, U_1^{(n)}, \ldots, U_k^{(n)}$ son v.a.i.i.d. $\mathcal{U}(0,1)$.

Corollary.

$$E[\widehat{\theta}_A] = \theta, \qquad V[\widehat{\theta}_A] = \frac{1}{2n} V[Y_i^{(1)}] + \frac{1}{2n} \mathrm{Cov}(Y_i^{(1)}, Y_i^{(2)}).$$

Corollary. $V[\widehat{\theta}_A] \leq \frac{1}{2} V[\widehat{\theta}]$.

Remark. The estimator $\widehat{\theta}_A$ is preferred to $\widehat{\theta}$ because it has lower variance. Variance is reduced at least to a half.

Miguel Hernández

Example. $\theta = \int_0^1 \sqrt{1-x^2}\,dx$, continuation).

The antithetic estimator of θ is

$$\widehat{\theta}_A = \frac{1}{2n}\sum_{i=1}^{n}\left(\sqrt{1-U_i^2} + \sqrt{1-(1-U_i)^2}\right).$$

$$E[\widehat{\theta}_A] = \theta = \pi/4, \qquad V[\widehat{\theta}_A] = \frac{0.01355}{2n},$$

Observación. Based on n generated uniform random numbers, we have

$$V[\widehat{\theta}] = 0.0498/n, \qquad V[\widehat{\theta}_A] = 0.006775/n.$$

Conclusion: Variance of $\widehat{\theta}_A$ is approximately 7 times lower than variance of $\widehat{\theta}$.

UNIVERSITAS Miguel Hernández

- We want to estimate $\theta = E[\phi(X_1, \ldots, X_k)]$.
- We have a measurable function

$$\psi : (\boldsymbol{R}^k, \mathcal{B}_k) \longrightarrow (\boldsymbol{R}, \mathcal{B})$$

such that $E[\psi(X)] = E[\psi(X_1, \ldots, X_k)] = \mu$ is known.

We define, for all $a \in \boldsymbol{R}$,

$$W(X_1, \ldots, X_k) = \\ \phi(X_1, \ldots, X_k) - a(\psi(X_1, \ldots, X_k) - \mu).$$

$$\begin{aligned} E[W(X_1, \ldots, X_k)] &= \theta \\ V[W(X_1, \ldots, X_k)] &= V[\phi(X)] + a^2 V[\psi(X)] \\ &- 2a\text{Cov}(\phi(X), \psi(X)), \end{aligned}$$

where $X = (X_1, \ldots, X_k)$.

Corollary. $V[W]$ is minimized in

$$a = \frac{\text{Cov}(\phi(X), \psi(X))}{V[\psi(X)]},$$

and for this value of a we get

$$V[W] = V[\phi(X)] - \frac{\text{Cov}(\phi(X), \psi(X))^2}{V[\psi(X)]}.$$

Definition. Let

$$X^{(1)} = (X_1^{(1)}, \ldots, X_k^{(1)})$$
$$\vdots$$
$$X^{(n)} = (X_1^{(n)}, \ldots, X_k^{(n)})$$

be a random sample simulated from $X = (X_1, \ldots, X_k)$.

The **control estimator** of θ is

$$\widehat{\theta}_C = \frac{1}{n}\sum_{i=1}^{n}\left[\phi(X^{(i)}) - a(\psi(X^{(i)}) - \mu)\right]$$

Definition. Given three control variates ψ_1, ψ_2 and ψ_3 such that $E[\psi_1] = \mu_1$, $E[\psi_2] = \mu_2$ and $E[\psi_3] = \mu_3$, the control estimator is

$$\widehat{\theta}_C = \frac{1}{n}\sum_{i=1}^{n}\Big[\phi(X^{(i)}) - a_1(\psi_1(X^{(i)}) - \mu_1)$$
$$-a_2(\psi_2(X^{(i)}) - \mu_2) - a_3(\psi_3(X^{(i)}) - \mu_3)\Big]$$

Corollary.

1. $E[\widehat{\theta}_C] = \theta$,
2. $V[\widehat{\theta}_C] = \frac{V[\phi(X)]}{n} - \frac{\text{Cov}(\phi(X),\psi(X))^2}{nV[\psi(X)]}, \quad V[\widehat{\theta}_C] \leq V[\widehat{\theta}]$.

Remark. Optimal value of a cannot be calculated because Cov$(\phi(X), \psi(X))$ is unknown.

This problem can be solved as follows:

1. Do previously a pilot simulation to estimate Cov$(\phi(X), \psi(X))$ and use the estimated value of "a".
2. Estimate "a" directly from the simulated data.

- First method has the disadvantage of being more slow.
- Second method has the disadvantage that "a" is not a constant any more to became a function of $X = (X_1, \ldots, X_k)$, so that $\widehat{\theta}_C$ is not unbiased.

- We want to estimate by simulation

$$\theta = \int_a^b \phi(x)\, dx,$$

where $\phi : \boldsymbol{R} \longrightarrow \boldsymbol{R}$ is measurable and positive.

- We want to estimate the area under the curve $y = \phi(x)$, between $x = a$ and $x = b$.

Procedure:

1. Divide the interval (a, b) in k disjoint subintervals

$$(\alpha_0, \alpha_1], (\alpha_1, \alpha_2], \ldots, (\alpha_{k-1}, \alpha_k),$$

with $\alpha_0 = a$ and $\alpha_k = b$.

2. Variability of $\phi(x)$ within each subinterval is lower than in the interval (a, b).

Idea: Estimate by pure Monte Carlo method

$$\theta_j = \int_{\alpha_{j-1}}^{\alpha_j} \phi(x)dx,$$

to obtain $\theta = \sum_{j=1}^{k} \theta_j$.

Monte Carlo method to estimate θ_j

$$\begin{aligned} \theta_j &= (\alpha_j - \alpha_{j-1}) \int_{\alpha_{j-1}}^{\alpha_j} \phi(x) \frac{1}{\alpha_j - \alpha_{j-1}} dx \\ &= (\alpha_j - \alpha_{j-1}) \int_{\alpha_{j-1}}^{\alpha_j} \phi(x) f_{\mathcal{U}(\alpha_{j-1},\alpha_j)}(x)\, dx. \end{aligned}$$

The estimator of θ_j is

$$\widehat{\theta}_j = (\alpha_j - \alpha_{j-1})\frac{1}{n_j}\sum_{i=1}^{n_j}\phi\left(\alpha_{j-1} + (\alpha_j - \alpha_{j-1})U_{ij}\right)$$

where U_{ij} i.i.d. $\mathcal{U}(0,1)$

Definition: $\widehat{\theta}_E = \sum_{j=1}^{k}\widehat{\theta}_j$.

$$E\left[\widehat{\theta}_E\right] = \sum_{j=1}^{k} E\left[\widehat{\theta}_j\right] = \sum_{j=1}^{k}\theta_j = \theta$$

$$V\left[\widehat{\theta}_E\right] = \sum_{j=1}^{k} V\left[\widehat{\theta}_j\right] = \sum_{j=1}^{k}\frac{(\alpha_j - \alpha_{j-1})^2}{n_j}V\left[\phi(X_{ij})\right],$$

whre $X_{ij} \sim \mathcal{U}(\alpha_{j-1}, \alpha_j)$, and because of

$$V\left[\phi(X_{ij})\right] = \frac{1}{\alpha_j - \alpha_{j-1}} \int_{\alpha_{j-1}}^{\alpha_j} \phi^2(x)dx - \left(\frac{1}{\alpha_j - \alpha_{j-1}} \int_{\alpha_{j-1}}^{\alpha_j} \phi(x)dx\right)^2,$$

it holds

$$V\left[\widehat{\theta}_E\right] = \sum_{j=1}^{k} \frac{1}{n_j} \left\{ (\alpha_j - \alpha_{j-1}) \int_{\alpha_{j-1}}^{\alpha_j} \phi^2(x)dx - \left(\int_{\alpha_{j-1}}^{\alpha_j} \phi(x)dx\right)^2 \right\}$$

Remark. When using stratified sampling we have to choose k, $\{\alpha_j\}$ and $\{n_j\}$.

Problem: Given n, k and $\{\alpha_j\}$. Find $\{n_j\}$ such that $V\left[\widehat{\theta}_E\right]$ be minimized.

This is to say

$$\begin{cases} \text{minimize} & V\left[\widehat{\theta}_E\right] = \sum_{j=1}^{k} \frac{a_j}{n_j} \\ \text{Restricted to} & \displaystyle\sum_{j=1}^{k} n_j = n \end{cases}$$

where $a_j = (\alpha_j - \alpha_{j-1})^2 V\left[\phi(X_{ij})\right]$.

Solution:

$$n_j \propto +\sqrt{a_j} = (\alpha_j - \alpha_{j-1})\sqrt{V\left[\phi(X_{ij})\right]},$$

where $X_j \sim \mathcal{U}(\alpha_{j-1}, \alpha_j)$

Remark. $V\left[\phi(X_{ij})\right]$ is unknown; however the conclusion is clear: n_j *should be large if size of stratum is large or if variability within the stratum is large.*

How to proceed:

1. First, select the $\{\alpha_j\}$ in such a way that the curve $y = \phi(x)$ be approximately constant in the intervals (α_{j-1}, α_j). This is to say, in such a way that $V\left[\phi(X_{ij})\right] \cong$ constant. Second, select n_j proportionally to the length of the interval $(\alpha_j - \alpha_{j-1})$.
2. Estimate $V\left[\phi(X_j)\right]$ with a pilot simulation. Choose n_j according to the formula $n_j \propto (\alpha_j - \alpha_{j-1})\sqrt{\widehat{V}\left[\phi(X_{ij})\right]}$.

- We want to estimate by simulation

$$\theta = \int_a^b \phi(x)\,dx,$$

where $\phi\colon \boldsymbol{R} \longrightarrow \boldsymbol{R}^+$ is measurable and positive.

- Let $f(\cdot)$ be a density function with support in the interval (a,b), then

$$\theta = \int_a^b \phi(x)dx = \int_a^b \frac{\phi(x)}{f(x)} f(x)dx = E\left[\frac{\phi(X)}{f(X)}\right],$$

where $X \sim f(\cdot)$.

- If $X_1, \ldots, X_n$ is a simulated random sample from $f(\cdot)$, we can estimate θ by the pure Monte Carlo method. We have

$$\widehat{\theta}_I = \frac{1}{n}\sum_{i=1}^{n} \frac{\phi(X_i)}{f(X_i)}$$

$$(1)\quad E\left[\widehat{\theta}_I\right] = \theta$$

$$(2)\quad V\left[\widehat{\theta}_I\right] = \frac{1}{n}\left\{\int_a^b \frac{\phi^2(x)}{f(x)}dx - \theta^2\right\}$$

Remarks.

- If $f(x) = \frac{\phi(x)}{\theta} \Rightarrow V\left[\widehat{\theta}_I\right] = 0.$
  $f(x) = \frac{\phi(x)}{\theta} I_{(a,b)}(x)$ is the optimal selection.
  As θ is unknown, we can choose $f(x)$ with a shape similar to $\phi(x)$.
- The "importance sampling method" is so called because $\phi(\cdot)$ is evaluated more frequently in the places where it is more beneficial.

Remark.

- Stratified sampling estimation method is a particular case of the importance sampling method, when f is a mixture of uniform distributions

$$f(x) = \sum_{j=1}^{k} \frac{n_j}{n} \frac{1}{(\alpha_j - \alpha_{j-1})} I_{(\alpha_{j-1},\alpha_j)}(x)$$

Example (Estimation of $P(\mathcal{N}(0,1) < a)$)

We want to estimate

$$\theta = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx = \int_{-\infty}^{a} \phi(x)\, dx.$$

A probability density function with shape similar to $\phi(\cdot)$ is the logistic p.d.f. with mean 0 and variance 1; this is to say

$$f(x) = \frac{\pi e^{-\frac{\pi x}{\sqrt{3}}}}{\sqrt{3}\left(1 + e^{-\frac{\pi x}{\sqrt{3}}}\right)^2}.$$

We can write

$$\theta = \int_{-\infty}^{a} \frac{k\phi(x)}{f(x)} \frac{f(x)}{k}\, dx,$$

where $k \in \boldsymbol{R}$ is such that $g(x) = \frac{f(x)}{k}$ is a p.d.f. in $(-\infty, a)$. Therefore

$$k = \int_{-\infty}^{a} \frac{\pi e^{-\frac{\pi x}{\sqrt{3}}}}{\sqrt{3}\left(1 + e^{-\frac{\pi x}{\sqrt{3}}}\right)^2} dx = \frac{1}{1 + e^{-\frac{\pi a}{\sqrt{3}}}}.$$

Finally, the estimator of θ is

$$\widehat{\theta} = \frac{1}{n\left(1 + e^{-\frac{\pi x}{\sqrt{3}}}\right)} \sum_{i=1}^{n} \frac{\phi(x_i)}{f(x_i)},$$

where $x_1, \ldots, x_n$ are random numbers from the p.d.f. $g(\cdot)$.

Generation of $X \sim g(\cdot)$.

To generate X we can combine the *rejection* and the *inversion* methods.

1. We apply the inversion method to generate logistic random numbers with mean 0 and variance 1.

$$u = F(x) = \frac{1}{1 + e^{-\frac{\pi x}{\sqrt{3}}}} \Leftrightarrow \frac{\pi x}{\sqrt{3}} = -\log \frac{1-u}{u}.$$

   Generation formula is

$$x = F^{-1}(u) = -\frac{\sqrt{3}}{\pi} \log \frac{1-u}{u}.$$

2. We apply a rejection criterium to obtain values of X with logistic distribution truncated at $(-\infty, a)$.

Algorithm:

1. Generate $U \sim \mathcal{U}(0,1)$
2. Do $X = -\frac{\sqrt{3}}{\pi} \ln \frac{1-u}{u}$
3. If $X > a$, go to 1.
4. Output: X

Problem: estimate

$$\theta = E[Y] = E[\phi(X)] = E[\phi(X_1, \ldots, X_k)],$$

where $\phi : (\boldsymbol{R}^k, B_k) \longrightarrow (\boldsymbol{R}, B)$ is measurable

Basic solution:

- $X^{(1)} = (X_1^{(1)}, \ldots, X_k^{(1)}), \ldots, X^{(n)} = (X_1^{(n)}, \ldots, X_k^{(n)})$ is a random sample from X.
- *Monte Carlo Estimator*

$$\widehat{\theta} = \frac{1}{n} \sum_{i=1}^{n} \phi(X^{(i)}) = \frac{1}{n} \sum_{i=1}^{n} Y^{(i)}.$$

Target: Reduce $V[\widehat{\theta}]$.

Idea: Conditioning.

$$\begin{aligned} E[Y] &= E\left[E[Y \,/\, Z]\right], \\ V[Y] &= E\left[V[Y \,/\, Z]\right] + V\left[E[Y \,/\, Z]\right], \end{aligned}$$

Let $Z^{(1)}, \ldots, Z^{(n)}$ be random variables such that $E[Y^{(i)} \,/\, Z^{(i)}]$ can be easily calculated. Then:

$$\begin{aligned} E\left[E[Y^{(i)} \,/\, Z^{(i)}]\right] &= E[Y^{(i)}] = \theta, \\ V\left[E[Y^{(i)} \,/\, Z^{(i)}]\right] &= V[Y^{(i)}] - E\left[V[Y^{(i)} \,/\, Z^{(i)}]\right] \\ &\leq V[Y^{(i)}]. \end{aligned}$$

The *conditional Monte Carlo estimator* of θ is:

$$\widehat{\theta}_Z = \frac{1}{n}\sum_{i=1}^{n} E[Y^{(i)} \,/\, Z^{(i)}].$$

For the conditional estimator, it holds

$$V[\widehat{\theta}_Z] \leq V[\widehat{\theta}].$$

Example 8.1.

$W \sim \mathcal{P}o(\lambda)$ and $X \sim \mathcal{B}e(W, W^2+1)$.

The algorithm to calculate the *Monte Carlo estimator* is:

1. Generate n pairs $(W^{(i)}, X^{(i)}), i = 1, \ldots, n$;

$$W^{(i)} \sim \mathcal{P}o(\lambda), X^{(i)} \sim \mathcal{B}e(W^{(i)}, W^{(i)2}+1).$$

2. Calculate $\widehat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X^{(i)}$.

Conditional Monte Carlo method:

$$E[X \,/\, W = w] = \frac{w}{w^2 + w + 1}$$

$$\theta = \sum_{w=0}^{\infty} E[X \,/\, W = w] \frac{e^{-\lambda}\lambda^w}{w!}.$$

The algorithm to calculate the *conditional Monte Carlo estimator* is:

1. Generate n values $W^{(i)} \sim \mathcal{P}o(\lambda)$.
2. Calculate $\widehat{\theta}_W = \frac{1}{n}\sum_{i=1}^{n} \frac{W^{(i)}}{W^{(i)2} + W^{(i)} + 1}$.

Another case:

- For $j = 1, 2, \ldots$, there exist random variables $Z_j^{(1)}, \ldots, Z_j^{(n)}$ such that $E[Y^{(i)} \,/\, Z_j^{(i)}]$ be calculable.
- Let $p_1, p_2, \ldots$ be such that $p_j \geq 0$, $\sum_{j=1}^{\infty} p_j = 1$.
- The *conditional Monte Carlo estimator* of θ is:

$$\widehat{\theta}_{p,Z} = \sum_{j=1}^{\infty} p_j \frac{1}{n} \sum_{i=1}^{n} E[Y^{(i)} \,/\, Z_j^{(i)}]$$

Proposition.

$$E\left[\sum_{j=1}^{\infty} p_j E[y \,/\, Z_j]\right] = E[Y], \qquad V\left[\sum_{j=1}^{\infty} p_j E[y \,/\, Z_j]\right] \leq V[Y].$$

Example 8.2. Let $X_1, X_2 \ldots$ be independent exponential random variables with parameters $\lambda_1, \lambda_2, \ldots$ respectively. Estimate

$$\theta = P(X_1 + \ldots + X_k \leq t) = E[\phi(X_1, \ldots, X_k)],$$

where

$$Y = \phi(X_1, \ldots, X_k) = \begin{cases} 1 & \text{if} \quad \sum_{\ell=1}^{k} X_\ell \leq t \\ 0 & \text{otherwise} \end{cases}$$

Solution: Let the random sample

$$\begin{aligned} X^{(1)} &= (X_1^{(1)}, \ldots, X_k^{(1)}) \\ &\vdots \\ X^{(n)} &= (X_1^{(n)}, \ldots, X_k^{(n)}) \end{aligned}$$

simulated from $X = (X_1, \ldots, X_k)$.

Pure Monte Carlo estimator is:

$$\widehat{\theta}_1 = \frac{1}{n}\sum_{i=1}^{n}\phi(X^{(i)}) = \frac{1}{n}\sum_{i=1}^{n}Y_i$$
$$= \frac{\text{number of samples such that } \sum_{\ell=1}^{n} X_\ell \le t}{n}$$

We are interested in finding the *Monte Carlo estimator conditioned to*

$$\varepsilon_j(X) = (X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_k)$$

Solution:

$$\begin{aligned} E[Y \,/\, \varepsilon_j(X)] &= P\left(\sum_{\ell=1}^{k} X_\ell \leq t \,/\, \varepsilon_j(X)\right) \\ &= F_j\left(t - \sum_{\ell \neq j}^{k} X_\ell\right), \end{aligned}$$

where

$$F_j(s) = 1 - \exp\{-\lambda_j s\}, \quad s > 0.$$

The conditional Monte Carlo estimator is:

$$\widehat{\theta}_2 = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{k} p_j F_j \left(t - \sum_{\ell \neq j}^{k} X_\ell^{(i)}\right)$$

where $p_j \geq 0$, $j = 1, \ldots, k$ y $\sum_{j=1}^{k} p_j = 1$.

- Generate n i.i.d. random variables, $Y^{(1)}, \ldots, Y^{(n)}$, with mean μ and variance σ^2.
- Use $\bar{Y}_n = \frac{1}{n}\left(Y^{(1)} + \ldots + Y^{(n)}\right)$ to estimate μ.
- The precision of the estimator can be measure with its variance

$$V[\bar{Y}_n] = E\left[(\bar{Y}_n - \mu)^2\right] = \frac{\sigma^2}{n}.$$

Problem: Find the sample size n such that $V[\bar{Y}_n]$ be sufficiently small.

Solution: If random variables $Y^{(i)}$ are normal, the problem can be solved with a 95% confidence interval for μ

$$\bar{Y}_n \pm \frac{1.96\sigma}{\sqrt{n}}.$$

“find n such that $\frac{1.96\sigma}{\sqrt{n}} \leq \varepsilon$".

Procedure:
As σ^2 is unknown, simulate initially m variables $Y^{(1)}, \ldots, Y^{(m)}$, where $m \geq 30$ (in order to make the Central Limit Theorem applicable) and estimate σ^2 with

$$\widehat{\sigma}_m^2 = \frac{1}{m-1} \sum_{i=1}^{m} \left(Y^{(i)} - \bar{Y}_m \right)^2 .$$

If ε the desired precision level, then

$$\frac{1.96^2 \widehat{\sigma}_m^2}{n} \leq \varepsilon^2 \Longleftrightarrow n \geq \left(\frac{1.96 \widehat{\sigma}_m}{\varepsilon} \right)^2 ;$$

this is to say,

$$n = \left[\left(\frac{1.96 \widehat{\sigma}_m}{\varepsilon} \right)^2 \right] + 1.$$

Finally,

- If $n > m$, then generate $n - m$ new random variables $Y^{(i)}$.
- If $n \leq m$, then do not generate new random variables $Y^{(i)}$.

Example.

Estimate by simulation the parameter

$$\theta = P\left(\mathcal{N}(0,1) < 1\right),$$

whose value is known to be $\theta = 0.8413$.

Example 10.1.

$$\theta = \int_{-\infty}^{\infty} \phi(x) f_{\mathcal{N}(0,1)}(x) dx,$$

where $\phi(x) = I_{(-\infty,1)}(x)$.
In this case, Monte Carlo estimator is

$$\widehat{\theta}_1 = \frac{1}{n} \sum_{i=1}^{n} \phi(X_i) = \frac{\text{Núm. de } X_i < 1}{n},$$

where $X_1, \ldots, X_n$ are i.i.d. $\mathcal{N}(0,1)$. We have

$$V\left[\widehat{\theta}_1\right] = \frac{0,1335}{n}.$$

To obtain a precision of 10^{-3} with a 95% confidence we need $n = \left[(1.96)^2 0.1335 \cdot 10^6\right] + 1 = 512854$ simulated random variables.

Example 10.2.

$$\theta = 1 - \frac{1}{2} P\left(|\mathcal{N}(0,1)| > 1\right) = \int_{-\infty}^{\infty} \phi(x) f_{\mathcal{N}(0,1)}(x) dx,$$

where $\phi(x) = 1 - \frac{1}{2} I_{(-\infty,-1)\cup(1,\infty)}(x)$.

In this case, Monte Carlo estimator is

$$\widehat{\theta}_2 = \frac{1}{n} \sum_{i=1}^{n} \phi(x_i) = 1 - \frac{1}{2} \frac{R_2}{n},$$

where the x_i are random numbers from a $\mathcal{N}(0,1)$ pd.f. and R_2 is the number of x_i such that $|x_i| > 1$. We have

$$V\left[\widehat{\theta}_2\right] = \frac{0.05416}{n}$$

To obtain a precision of 10^{-3} with a 95% confidence we need $n = \left[(1.96)^2 0.05416 \cdot 10^6\right] + 1 = 208062$ simulated random variables.

Example 10.3. $\theta = \frac{1}{2} + I,$

$$I = P\left(0 < \mathcal{N}(0,1) < 1\right) = \int_{-\infty}^{\infty} \phi(x) f_{\mathcal{N}(0,1)}(x) dx$$

where $\phi(x) = I_{(0,1)}(x)$.

In this case, Monte Carlo estimator is $\widehat{\theta}_3 = \frac{1}{2} + \widehat{I}_3$, where

- $\widehat{I}_3 = \frac{1}{n} \sum_{i=1}^{n} \phi(x_i) = \frac{R_3}{n}$
- $x_1, \ldots, x_n$ are $\mathcal{N}(0,1)$ random numbers.
- R_3 is the number of $x_i \in (0,1)$.

We have

$$V\left[\widehat{\theta}_3\right] = \frac{0.2248}{n}$$

To obtain a precision of 10^{-3} with a 95% confidence we need $n = \left[(1.96)^2 0.2248 \cdot 10^6\right] + 1 = 863647$ simulated random variables.

Example 10.4. $\theta = \frac{1}{2} + I$,

$$I = P\left(0 < \mathcal{N}(0,1) < 1\right) = \int_{-\infty}^{\infty} \phi(x) f_{\mathcal{U}(0,1)}(x)dx$$

where $f_{\mathcal{U}(0,1)}(x) = I_{(0,1)}(x)$ y $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$.
Monte Carlo estimator of θ is

$$\widehat{\theta}_4 = \frac{1}{2} + \widehat{I}_4$$

where

$$\widehat{I}_4 = \frac{1}{n\sqrt{2\pi}} \sum_{i=1}^{n} e^{-\frac{1}{2}U_i^2}$$

and $U_1, \ldots, U_n$ are i.i.d. $\mathcal{U}(0,1)$.

We get

$$V[\widehat{\theta}_4] = \frac{0.00238931}{n}.$$

To obtain a precision of 10^{-3} with a 95% confidence we need $n = \left[(1.96)^2 0.0023893 \cdot 10^6\right] + 1 = 9179$ simulated random variables.

Note: In Examples 10.5 and 10.6 estimator $\widehat{\theta}_4$ has been improved.

Example 10.5. $\theta = \frac{1}{2} + I$,

$$I = P\left(0 < \mathcal{N}(0,1) < 1\right) = \int_{-\infty}^{\infty} \phi(x) f_{\mathcal{U}(0,1)}(x) dx$$

where $f_{\mathcal{U}(0,1)}(x) = I_{(0,1)}(x)$ y $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$.
Antithetic estimator of θ is

$$\widehat{\theta}_5 = \frac{1}{2} + \widehat{I}_5$$

where

$$\widehat{I}_5 = \frac{1}{2n\sqrt{2\pi}} \sum_{i=1}^{n} \left\{ e^{-\frac{1}{2}U_i^2} + e^{-\frac{1}{2}(1-U_i)^2} \right\}$$

and $U_1, \ldots, U_n$ are i.i.d. $\mathcal{U}(0,1)$.

Variance of estimator is

$$V[\widehat{\theta}_5] = V[\widehat{I}_5] = \frac{1}{2}V[\widehat{I}]$$
$$+\frac{1}{2n}\mathrm{Cov}\left(\frac{e^{-\frac{1}{2}U^2}}{\sqrt{2\pi}}, \frac{e^{-\frac{1}{2}(1-U)^2}}{\sqrt{2\pi}}\right) = \frac{0.0001278}{n}$$

To obtain a precision of 10^{-3} with a 95% confidence we need $n = \left[(1.96)^2 0.0001278 \cdot 10^6\right] + 1 = 491$ random variables simulated from a $\mathcal{U}(0,1)$ distribution.

Example 10.6. Antithetic estimator in Example 9.5 can be improved by using control variates.
Observe that

$$e^{-\frac{x^2}{2}} + e^{-\frac{(1-x)^2}{2}} \approx -x^2 + x + \frac{3}{2},$$

in a neighborhood of the origin.
Let

- $\psi(x) = -x^2 + x$ be a control function
- $\phi(x) = \phi_1(x) + \phi_2(x)$
- $\phi_1(x) = \frac{1}{2\sqrt{2\pi}} e^{-\frac{x^2}{2}}$
- $\phi_2(x) = \frac{1}{2\sqrt{2\pi}} e^{-\frac{(1-x)^2}{2}}$

The Control antithetic estimator is

$$\widehat{\theta}_6 = \frac{1}{2} + \frac{1}{n}\sum_{i=1}^{n}\left\{\phi(U_i) - a(\psi(U_i) - \mu)\right\},$$

where

- $U_1, \ldots, U_n$ are i.i.d. $\mathcal{U}(0,1)$
- $\mu = E[\psi(U_i)]$.

We have

$$\begin{aligned} \mu = E\left[\psi\right] &= -E\left[U^2\right] + E\left[U\right] = -\int_0^1 u^2\,du + \int_0^1 u\,du = \frac{1}{6} \\ V\left[\widehat{\theta}_6\right] &= \frac{V\left[\phi\right]}{n} - \frac{Cov\left(\phi,\psi\right)^2}{V\left[\psi\right]} = \frac{0.0000244}{n}, \end{aligned}$$

To obtain a precision of 10^{-3} with a 95% confidence we need $n = \left[(1.96)^2 0.0000244 \cdot 10^6\right] + 1 = 94$ random variables simulated from a $\mathcal{U}(0,1)$ distribution.