Skoltech

Skolkovo Institute of Science and Technology

Stochastic methods in Mathematical Modelling

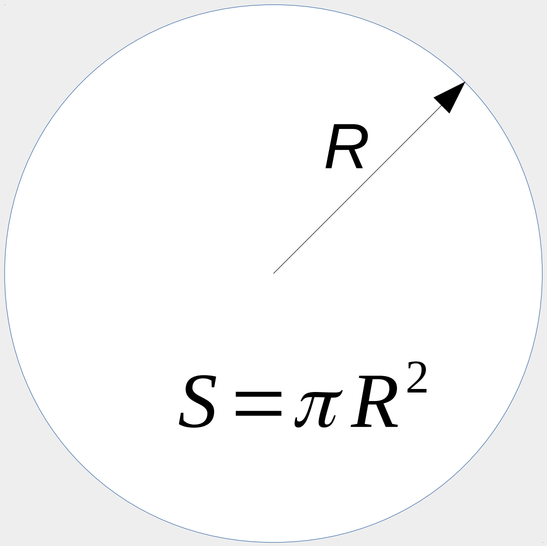
Lecture 16. Monte Carlo methods.

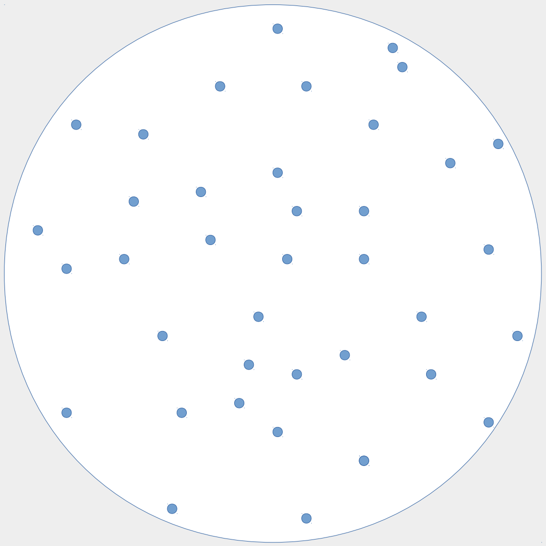
Direct sampling

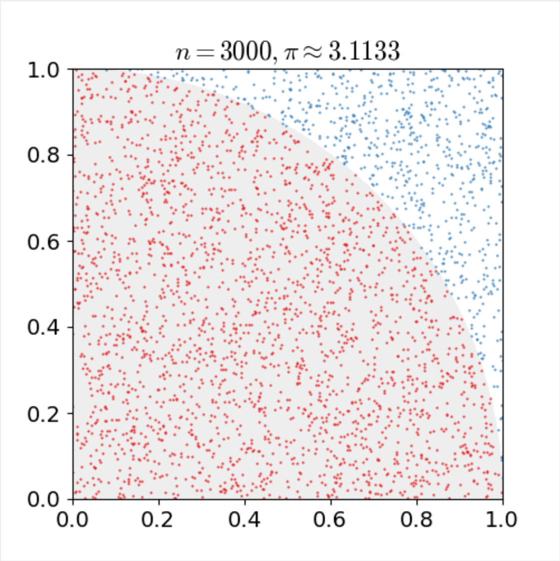
Monte Carlo. The main idea

Random sampling instead of exact solution

Malvin Kalos:
Monte Carlo means using random numbers to estimate something that is not random



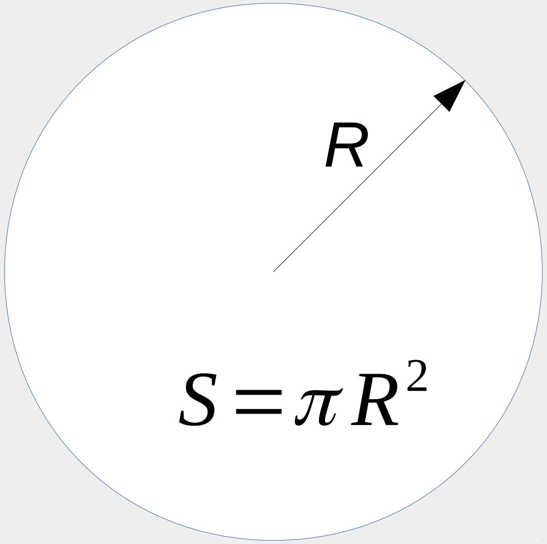


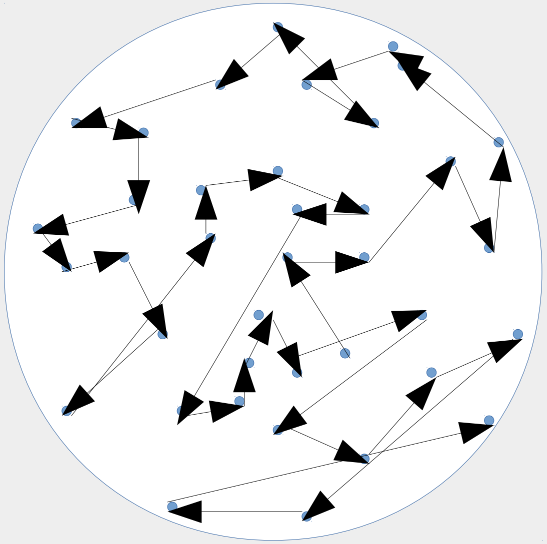


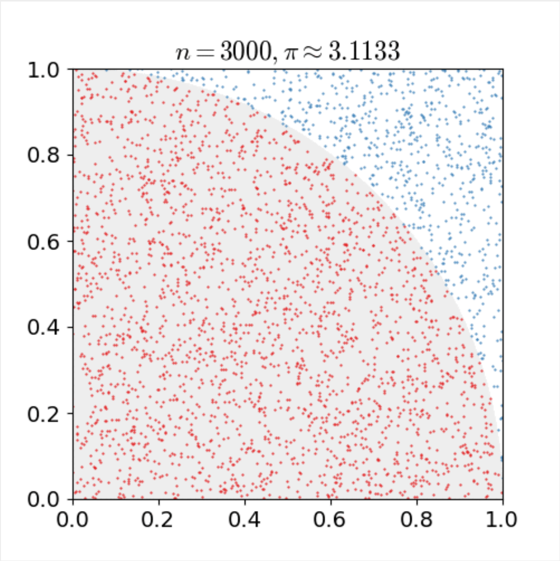
Monte Carlo. The main idea

Random sampling instead of exact solution

Alternatively we could do a random walk and sample the space in a correlated way





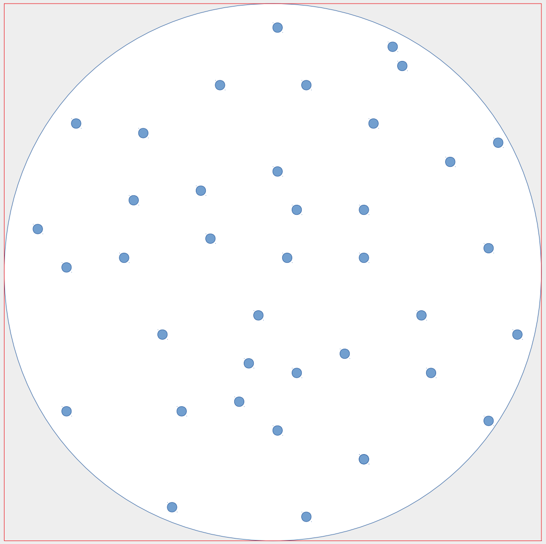


Monte Carlo. The main idea

Random sampling instead of exact solution

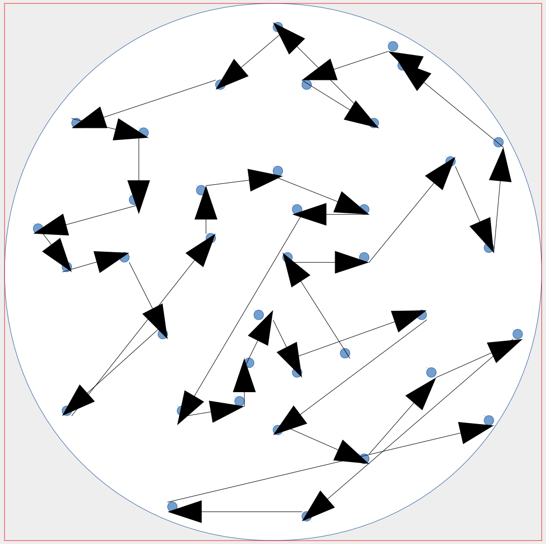
Direct sampling

Independent samples from a distribution



Markov Chain Monte Carlo (MCMC)

Draws are correlated according to
a Markov chain



Monte Carlo.
Direct Sampling
Brute force sampling

Checking all possibilities

Monte Carlo.
Direct Sampling by mapping

Skoltech
Skolkovo Institute of Science and Technology

Application of deterministic function to samples from a distribution
you already can sample efficiently. We are already familiar with it from Lecture 3 and HW1,3!

Inverse transform sampling

Example: Exponential distribution $P(x)=\lambda e^{-\lambda x}$

y is uniform random on [0,1]

$$x=F_X^{-1}(y)=\frac{-1}{\lambda}\ln(y)$$

Box-Muller transform

U_1 and U_2 are uniformly distributed RVs on [0,1]

$$Z_0=\sqrt{-2\ln U_1}\sin(2\pi U_2)$$

$$Z_1=\sqrt{-2\ln U_1}\cos(2\pi U_2)$$

Alpha-stable distributions S (α,β,c=1,μ=0;x) (Zolotarev formula)

$$S(\alpha,\beta)=\frac{\sin\alpha(\Phi-\Phi_0)}{(\cos\Phi)^{1/\alpha}}\cdot\left(\frac{\cos(\Phi-\alpha(\Phi-\Phi_0))}{W}\right)^{(1-\alpha)/\alpha},\quad \alpha\neq 1,$$

$$S(1,\beta)=\frac{2}{\pi}\left(\left(\tfrac{1}{2}\pi+\beta\Phi\right)\tan(\Phi)-\beta\ln\left(\frac{\frac{1}{2}\pi W\cos\Phi}{\frac{1}{2}\pi+\beta\Phi}\right)\right),$$

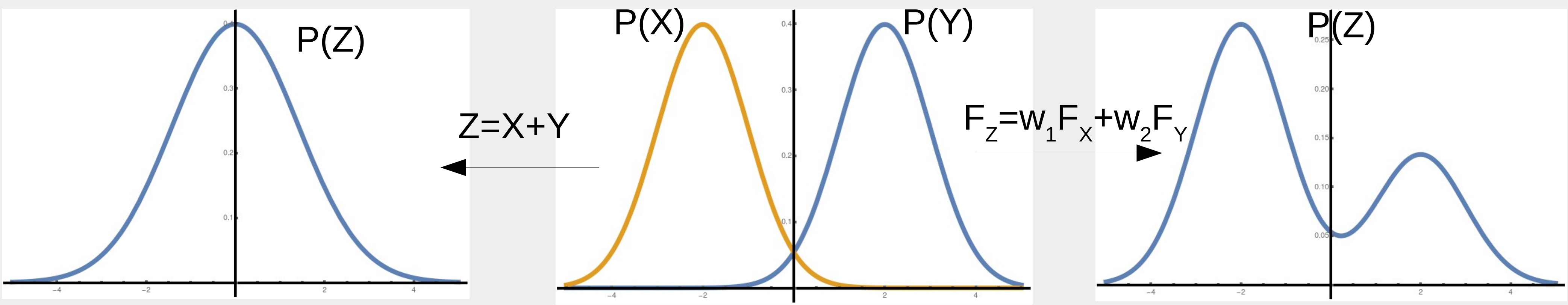
where W is exponentially distributed and $\Phi=U[-\pi/2,\pi/2]$

Skoltech
Skolkovo Institute of Science and Technology

Monte Carlo. Direct Sampling. Sums versus mixtures

Sums of random variables

Mixtures of random variables



X from f(x), y from g(y).
Then for z = x + y the PDF P(z)

$$P(z)=\int_{-\infty}^{\infty} f(x)\, g(z-x)\, dx$$

$$\tilde{P}(\omega)=\tilde{f}(\omega)\tilde{g}(\omega)$$

For CDFs

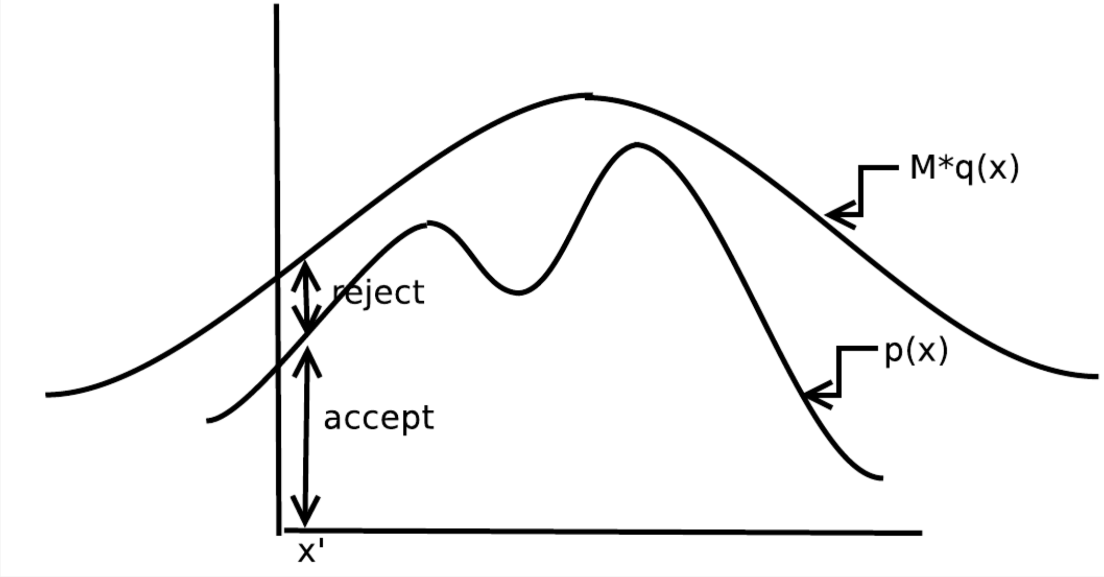
$$F_X(x)=\sum_{i=1}^{n} w_i F_{X_i}(x)$$

If all X_i are absolutely continuous then also

$$P_X(x)=\sum_{i=1}^{n} w_i P_{X_i}(x)$$

Monte Carlo.
Sampling by rejection

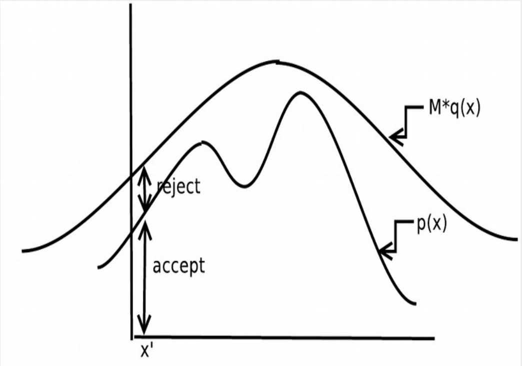
generates values from a target distribution with probability density p(x) by using a proposal distribution with probability density q(x). The idea is to sample p(x) by sampling q(x) and then accepting the values with probability p(x)/(M*q(x)), M any number which satisfies p(x)≤M*q(x) for all values of x



Monte Carlo.
Sampling by rejection. An exercise

How to get positive part of zero-mean Gaussian random variable from an exponential random variable

1. One samples from the exponential distribution $p_0(x)=e^{-x}$
2. We want to sample the positive half of a Gaussian $p_+(x)=\sqrt{2/\pi}\exp\left(-x^2/2\right)$



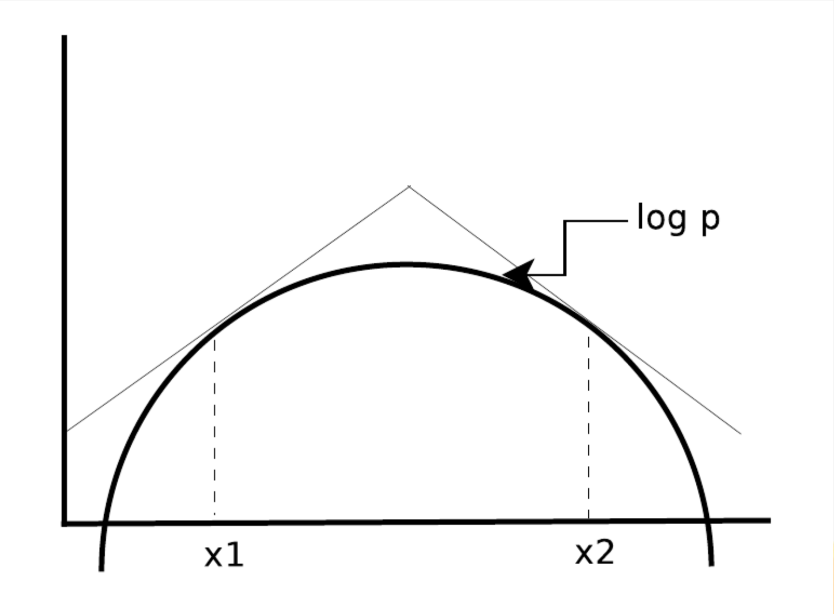
Algorithm

1. obtain a sample X from distribution $p_0(x)$ and a sample u from U[0,1]
2. Check whether or not $u < p_+(X)/(Mp_0(X))$.
3. If this holds, accept X as a sample drawn from $p_+(x)$ if not, reject the value of X and return to the sampling step.

The algorithm will take an average of M iterations to obtain a sample.

Monte Carlo.
Adaptive rejection sampling

1. works for logarithmically concave functions (concave functions which are non-negative on their domains are log-concave)
2. instead of a single envelope distribution a piece-wise linear envelope is constructed
3. If we choose a point that is rejected we tighten the proposal distribution with a line segment that is tangent to the curve at the point with the same x-coordinate as the chosen point.

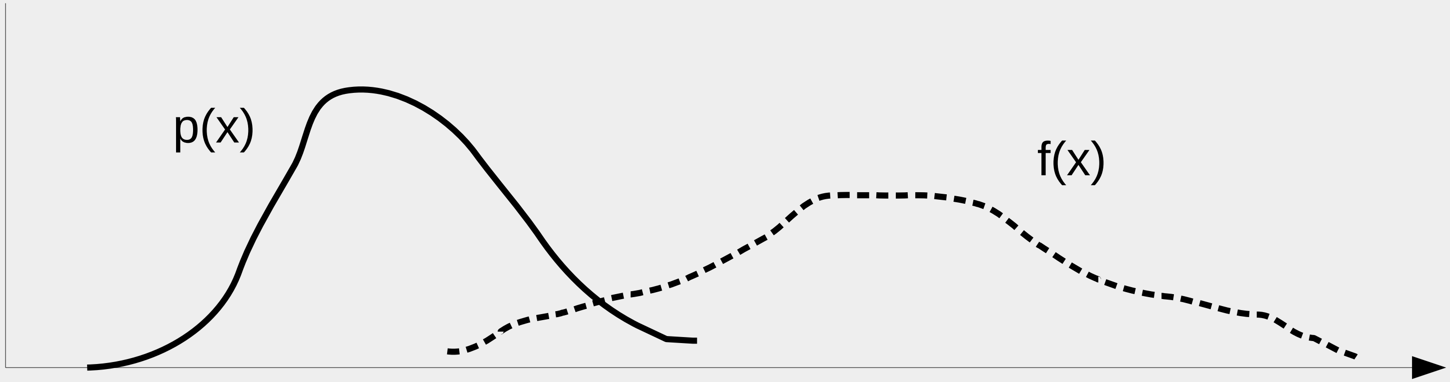


Problems: we have find the intersection of hyperplanes many times and usually the approach does not work for high dimensions

Monte Carlo. Importance sampling

Suppose we want to compute an expectation of a function f(x) over a distribution p(x),

$$\int dx p(x) f(x)$$



If f(x) and p(x) are concentrated around different x the overlap of f(x) and p(x) is small and a lot of MC samples will be wasted. The idea is to reduce the variance of the estimator by sampling the "important" values. The way is to move from sampling p(x) to $p_a(x)$ which is closer to important values

Monte Carlo. Importance sampling

Skoltech
Skolkovo Institute of Science and Technology

Suppose we want to compute an expectation of a function f(x) over a distribution p(x),

$$\int dx p(x) f(x)$$

If f(x) and p(x) are concentrated around different x the overlap of f(x) and p(x) is small and a lot of MC samples will be wasted.

Importance sampling is the method which helps to fix the small-overlap problem by adjusting the distribution function from p(x) to $p_a(x)$ and then utilising the following formula

$$\int dx p(x) f(x) = \int dx\, p_a(x) \frac{f(x)\, p(x)}{p_a(x)} = E_{p_a}\left[\frac{f(x)\, p(x)}{p_a(x)}\right]$$

Problem: in real multidimensional cases one can hardly have a good guess for $p_a(x)$. Solution: search for $p_a(x)$ *adaptively*

A sophisticated toolkit for adaptive importance sampling can be found here: https://pypi.org/project/pypmc/1.0/

Skoltech

Skolkovo Institute of Science and Technology

Monte Carlo. Importance sampling. Example

Importance sampling is the method which helps to fix the small-overlap problem by adjusting the distribution function from p(x) to p_a(x) and then utilising the following formula

$$\int dx p(x) f(x) = \int dx\, p_a(x) \frac{f(x) p(x)}{p_a(x)} = E_{p_a}\left[\frac{f(x) p(x)}{p_a(x)}\right]$$

Exercise: Implement the importance sampling for the computation of the average of f(x)

$$p(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2), f(x) = \exp(-(x-4)^2/2)$$

Proposal distribution

$$p_a(x) = \frac{1}{\sqrt{\pi}} \exp(-(x-2)^2)$$

Direct Monte Carlo. Variance reduction

The error in a direct MC simulation scales as $\sigma/n^{1/2}$. Hence, one can reduce the error by taking more samples or find a different formulation of MC with smaller σ (*variance reduction*). For instance, one could use *antithetic* variables

If X and Y are not independent and denoting the means as μ_X and μ_Y

$$var(X+Y)=E[(X+Y)^2]-(\mu_X+\mu_Y)^2=E[X^2]-\mu_X^2+E[Y^2]-\mu_Y^2+2(E[XY]-\mu_X\mu_Y)$$
$$=var(X)+var(Y)+2cov(X,Y)$$

$$cov(X,Y)=\rho_{X,Y}\sigma_X\sigma_Y, -1\le\rho\le 1$$

Def. Random variables X, Y on the same probability space are called *antithetic* if they have the same distribution and their covariance is negative.

Direct Monte Carlo. Variance reduction

We want to compute μ=E[X]. Let's find another variable Y such that X,Y are antithetic (E[Y]=μ).

Algorithm: we generate n independent samples $w_1, \ldots, w_n$ from the probability space and let $X_i=X(w_i)$, $Y_i=Y(w_i)$. For the mean we get the estimator

$$\hat{\mu}_n=\frac{1}{2n}\sum_{i=1}^{n}(X_i+Y_i) \quad \Rightarrow \quad var\,\hat{\mu}_n=\frac{1}{(2n)^2}n\,var(X+Y)=\frac{\sigma^2}{2n}(1+\rho)$$

For direct MC we need the time to generate w and then evaluate X on it. For "antithetic" MC we need to do the evaluation twice (for X and for Y). The antithetic fares better if

$$\frac{\sigma^2}{2}(1+\rho)(\tau_w+2\tau_e)<\sigma^2(\tau_w+\tau_e)$$

Time to generate w

Time to evaluate X

Direct Monte Carlo. Variance reduction

We want to compute μ=E[X]. Let's find another variable Y such that X,Y are antithetic (E[Y]=μ).

Algorithm: we generate n independent samples $w_1, \ldots, w_n$ from the probability space and let $X_i=X(w_i)$, $Y_i=Y(w_i)$. For the mean we get the estimator

$$\hat{\mu}_n = \frac{1}{2n}\sum_{i=1}^{n}(X_i+Y_i) \quad => \quad var\,\hat{\mu}_n = \frac{1}{(2n)^2}\,n\,var(X+Y) = \frac{\sigma^2}{2n}(1+\rho)$$

For direct MC we need the time to generate w and then evaluate X on it. For "antithetic" MC we need to do the evaluation twice (for X and for Y). The antithetic fares better if

$$\frac{1}{2}(1+\rho)(\tau_w+2\tau_e)<\tau_w+\tau_e$$

The advantage is large if ρ is close to -1

Time to generate w

Time to evaluate X

Antithetic Variables Example 1

we would like to compute $\mu = \int_0^1 f(x)\,dx = E[f(U)],\ U \sim [0,1]$

we assume $f(x)$ grows monotonously

$X = F(U)$

$Y = F(1-U)$

map $R: \Omega \to \Omega$

$x \to 1-x$

$Y(\omega) = X(R\omega)$

$X_e(\omega) = \frac{1}{2}[X(\omega) + X(R\omega)]$

$X_o(\omega) = \frac{1}{2}[X(\omega) - X(R\omega)]$

$E[X_e] = \mu$

$E[X_o] = 0$

$E[X_o X_e] = 0$ $\Big\}\rightarrow$ X_o & X_e are not correlated

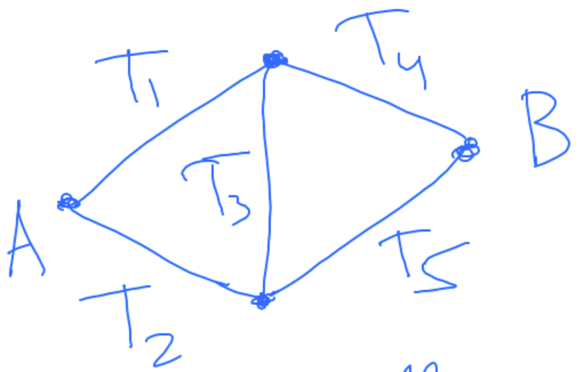
Correlation of X & Y

$\rho = \frac{\sigma_e^2 - \sigma_o^2}{\sigma_o^2 + \sigma_e^2}$

$\sigma^2 = \sigma_o^2 + \sigma_e^2$

if $\sigma_e << \sigma_o$ then Y is a good antithetic variable

Antithetic variables. Example 2.



$T_1 \in [0,1]$ $\quad T_3 \in [0,3]$ $\quad T_5 = [0,2]$

$T_2 \in [0,2]$ $\quad T_4 \in [0,1]$

What is the fastest path?

Analytically $\mu \simeq \frac{1339}{1440} \simeq 0.92986$

Estimator

$$\frac{1}{n}\sum_{i=1}^{n} z_i = \frac{1}{2n}\sum_{i=1}^{n}(X_i + Y_i)$$

$U_1, U_2, U_3, U_4, U_5 \in [0,1]$

$$\begin{cases} T_1 = U_1 \\ T_2 = 2U_2 \\ T_3 = 3U_3 \\ T_4 = U_4 \\ T_5 = 2U_5 \end{cases}$$

$X = h(U_1, U_2, \dots U_5)$

the quickest time

$$Z = \frac{X+Y}{2}$$

$$Y = h(1-U_1, 1-U_2, \dots, 1-U_5)$$

From simulation

$\operatorname{Var} X = 0.158$

$\operatorname{Var} Z = 0.0183$

The error reduction:

$$\sqrt{\frac{0.158}{0.0183}} \cdot \frac{1}{\sqrt{2}} \simeq 2$$

Skolkovo Institute of Science and Technology

Literature

1. https://math.nyu.edu/faculty/goodman/teaching/
2. https://people.eecs.berkeley.edu/~jordan/courses/260-spring10/
3. https://www.math.arizona.edu/~tgk/mc/book.pdf