Introduction to Monte Carlo methods

1 Monte Carlo method for integrals. Modeling of distributions.

1.1 Notations

We will present a general stochastic approach to compute integrals, to solve linear integral equations and to solve non-stationary kinetic equations with possible source in chemistry and in statistical physics. Elementary introduction to the master equation will be also given.

0. For stochastic variable Q expectation is $M(Q)$, variance is $D(Q) = M(Q^2) - (MQ)^2$.

If Q has density $q(x)$ the probability that Q is in a subset $D \subset G$ is by the definition integral of the density q over D:

$$P(Q \in D) = \int_D q(x)dx \tag{1}$$

1.2 Computing integrals

We consider an integral over a domain G in R^n and stochastic variable Q uniformly distributed over G. It follows directly from the definition of the density that

$$I_1 = \int_G f(x) = M(f(Q)) \approx \frac{1}{N}\sum_i^N f(Q_i) \tag{2}$$

where $\{Q_i\}_i^N$ is a sampling of the stochastic variable Q of the size N.

Consider an integral of the function f with a weight $p(x)$ such that $p(x)$ it is a probability density $\int_G p(x)dx = 1$ for a stochastic variable P with values in G. The similar arguments to ones above lead to

$$I_2 = \int_G f(x)p(x)dx = M(f(P)) \approx \frac{1}{N}\sum_i^N f(P_i) \tag{3}$$

where $\{P_i\}_i^N$ is a sampling of the stochastic variable P of the size N.

Example.

We consider a simple example with one-dimensional integral and will introduce a Monte Carlo scheme to compute it.

$$I = \int_0^\infty f(x)e^{-kx}dx \approx \frac{1}{Nk}\sum_i^N f(\xi_i) \tag{4}$$

where ξ is the stochastic variable with density $p(x) = ke^{-kx}$ and $\{\xi_i\}_1^N$ is a sampling for it.

For using these approximations in practice as also in previous general cases one needs to sample ξ.

1.3 Modeling stochastic variables with given density

a) Discrete stochastic variable.

$$\begin{pmatrix} x_1, x_2, ..., x_n \\ p_1, p_2, ..., p_n \end{pmatrix}; \qquad \sum_{i=1}^{n} p_i = 1 \tag{5}$$

Notation γ will be used here and later for a "standard" uniformly distributed on the unit interval $[0, 1]$ stochastic variable that is available in any programming language usually as a function with name **rand**.

Cut the interval $[0, 1]$ in intervals Δ_1, Δ_2,..., Δ_n. Define a stochastic variable λ by the following rule:

$$\lambda = x_i \quad if \quad \gamma \in \Delta_i$$

Then λ has the distribution above.

Proof. $P(\lambda = x_i) = P(\gamma \in \Delta_i) = length(\Delta_i) = p_i$ for any i.

b) Continuous 1-dim stochastic variables.

If $\xi \in [a, b]$ and has density $p(x) > 0$ (a and b we infinite) we consider the distribution function

$$F(x) = P(a < \xi < x) = \int_a^x p(u)du \tag{6}$$

Theorem.

Stochastic variable ξ such that $F(\xi) = \gamma$ has density $p(x)$

Proof.

F is monotone and therefore $F(\xi) = \gamma$ has exactly one solution for any $\gamma \in [0, 1]$. Simple calculation shows the statement of the theorem:

$P(x < \xi < x + \Delta x) = P(F(x) < \gamma < F(x + \Delta x)) = \int_{F(x)}^{F(x+\Delta x)} = F(x + \Delta x) - F(x) = \int_x^{x+\Delta x} p(u)du$

Example.

We want to model the variable ξ with the density $p(x) = ae^{-a(x-x_0)}$;

$x_0 < x < +\infty$.

$F(x) = \int_{x0}^{x} ae^{-au-x_0}du = 1 - e^{-a(x-x_0)}$

Therefore the equation for ξ is $\gamma = 1 - e^{-a(\xi-x_0)}$ and

$$\xi = x_0 - 1/a \ln(1 - \gamma)$$

Variables γ and $1 - \gamma$ have the same distribution, therefore also the simpler formula can be used:

$$\xi = x_0 - 1/a \ln(\gamma)$$

Therefore the integral in the earlier example can be computed by the formula

$$\int_0^{\infty} f(x)e^{-kx}dx \approx \frac{1}{kN} \sum_{i=1}^{N} f\left(-\frac{1}{k} \ln \gamma_i\right)$$

where γ is uniformly distributed over $[0, 1]$.

1.4 Estimates of the errors and importance sampling for integrals

Consider the integral $\int_G f(P)dP$.

Density $p(P)$ is acceptable for f if $p(P) > 0$ in points P where $f(P) \neq 0$.

If $p(P) > 0$ in the whole G then it is acceptable for any f.

Consider the integral

$$I_0 = \int_G \frac{f(P)}{p(P)}\, p(P)dP$$

where $p(P)$ is the density of a stochastic variable Q such that $p(P)$ is acceptable for f. Using the above method for computing integrals for I_0. Let G^+ be the subset of G where $f \neq 0$.

$$Z_0(P) = \left\{ \begin{matrix} \frac{f(P)}{p(P)},\ p \in g^+ \\ 0 \end{matrix} \right\}$$

$$I_0 = M(Z_0(Q)) = \int_G Z_0(P)p(P)dP = \int_{G^+} f(P)dP \approx \frac{1}{N}\sum_{i=1}^{N} Z_0(Q_i)$$

Error bounds depend on the variance $DZ_0(Q)$ of $Z_0(Q)$. Minimal $DZ_0 = 0$ will be if $p(P)$ is proportional to $f(P)$.

For the average $\overline{\xi_i}$ of similarly distributed stochastic variables ξ_i with expectation $M(\xi_i) = a$:

$$\overline{\xi_N} = \frac{1}{N}\sum_{i=1}^{N} \xi_i$$

we have:

$$P(\overline{\xi_N} - a > \varepsilon) \to 0, \quad N \to +\infty$$

for any $\varepsilon > 0$.

One can get useful estimates of errors by using the central limit theorem that for $N \to +\infty$ implies the following :

$$P\left(\left|\overline{\xi_N} - a\right| < x\sqrt{D\xi/N}\right) \approx \Phi(x)$$

where $\Phi(x) = (2/\sqrt{2\pi}\int_0^x e^{-t^2/2}dt$.

Choosing an acceptable probability β one can get corresponding constant $x = x_\beta$ in the estimate above from the equation $\Phi(x_\beta) = \beta$. For example the probability $\beta = 0.997$ corresponds to $x_\beta = 3$. The probability $\beta = 0.95$ corresponds to $x_\beta = 1.96$.

Approximate error r_N can be estimated from the same relation $\Phi(x_\beta) = \beta$ with $\beta = 0.5$. It gives $x_\beta = 0.6745$and the error $r_N \approx 0.6745\sqrt{D\xi/N}$. It shows that a smaller variance $D\xi$ for an estimate gives better precision.

2 Integral equations of second type.

We will consider integral equations of the second type:

$$z(P) = \int_G K(P, P')z(P')dP' + f(P)$$

or in operator form

$$z = \boldsymbol{K} z + f.$$

We will use the method with iterations to get an approximate solution by setting $z^{(0)} = \varphi(P)$; $z^{(i)} = \boldsymbol{K} z^{i-1} + f$.

Therefore

$$z^{(i)} = f + \boldsymbol{K} f + \boldsymbol{K}^2 f + \boldsymbol{K}^3 f ... + \boldsymbol{K}^{i-1} f + \boldsymbol{K}^i \varphi.$$

If this sequence of approximations converges then the solution is represented by the Neumann series:

$$z = \sum_{l=0}^{+\infty} \boldsymbol{K}^l f.$$

If $\iint\limits_{G\times G} |K(P,P')|^2 \, dP'dP < 1$ then the Neumann series converges in the mean square sense

$$\lim_{i\to\infty} \int_G \left| z^{(i)}(P) - z(P) \right|^2 dP = 0.$$

The proof is similar to one in the case we had for the integral equation corresponding to ODE.

Example.

We consider a very general model for stochastic scattering of particles. Let variable P describe coordinate in a possibly high-dimensional space including coordinates, velocities, energy levels of internal degrees of freedom,

chemical type etc. Let $f(P)$ denote the the density of first collisions per unit time. Density $f(P)$ is easy to calculate explicitly if the source of particles is given. Let $z(P)$ denote the unknown density of all collisions. Let $K_{\mathrm{col}\,l}(P', P)$ denote the probability that a particle after undergoing a collision in point P' will have after a unit time the next collision in the element dP of the phase space around the near P. Corresponding to this rule equation is easy to write down:

$$z(P) = \int K_{\mathrm{col}\,l}(P', P) z(P') dP' + f(P)$$

because a collision in the point P can be the first collision or collision that follows a collision in the point P'.The amount of such collisions per unit time is $z(P')dP'$. Integration over all possible P' we get the density of all collisions.

Iterations $\boldsymbol{K} f$, $\boldsymbol{K}^2 f$, $\boldsymbol{K}^3 f$, ..., $\boldsymbol{K}^i f$ have a simple sense: they are densities of second, third,... etc. collisions of particles.The Neumann series is therefore the sum of collisions of all possible multiplicities.

2.1 Monte Carlo computation of iterations of a linear integral operator.

We consider how to compute $\boldsymbol{K}^i \varphi$. Usually one computes functionals of $\boldsymbol{K}^i f$ that are mean values $\int_G \boldsymbol{K}^i \varphi(P) \psi(P) dP$ with some weight function $\psi(P)$. The Monte Carlo method is in fact effective only for such computations, not for computation of the whole function $\boldsymbol{K}^{i-1} \varphi(P)$.

It is easy to observe that

$$\int_G \boldsymbol{K}^i \varphi(P)\psi(P)dP = \int_G dP_0 \int_G dP_1 \int_G dP_2 ... \int_G dP_i \; K(P_0,P_1)K(P_1,P_2)...$$
$$...K(P_{i-1},P_i)\varphi(P_i)\psi(P_0)$$

The idea with computing such expression is essentially the same as the idea with importance sampling approach in the chapter on Monte Carlo computing of an integral. A particular idea in the present case is used for choosing sampling points in the high-dimensional domain $G \times G \times G \times ... \times G$.

The sampling is organized in the following way. We choose in G an arbitrary probability density $p(P)$ acceptable for ψ and an arbitrary probability density $p(P,P')$ in P' variable acceptable for the kernel $K(P,P')$.

$$\int_G p(P)dP = 1; \quad \int_G K(P,P')dP' = 1$$

We define in G a stochastic "trajectory" $T_i(Q_0 \to Q_1 \to Q_2 \to Q_3 ... \to Q_i)$ where::
a) Q_0 is chosen stochastically with density $p(P)$,
b) point Q_j for a known Q_{j-1} is chosen with density $p(Q_{j-1},Q_j)$.
The function $p(P,P')$ is therefore called transition probability from P, to P' and is denoted by $p(P,P') = p(P \to P')$. The function $p(P)$ is called the initial density.
The "trajectory" T_i can be interpreted as point in $G \times G \times G \times ... \times G$ and can be used for the numerical integration in $\int_G \boldsymbol{K}^i \varphi(P)\psi(P)dP$.
Density of this point is $p_i(\, Q_0, Q_1, Q_2, ..., Q_i) = p(Q_0,Q_1)p(Q_1,Q_2)p(Q_2,Q_3)...p(Q_{i-1},Q_i)$.

We introduce for shorter notations weights

$$W_j = \frac{K(Q_0,Q_1)K(Q_1,Q_2)...K(Q_{j-1},Q_j)}{p(Q_0,Q_1)p(Q_1,Q_2)...p(Q_{j-1},Q_j)}$$

defined for $j = 1, ..., i$.For $j = 0$ we put $W_0 = 1$. Then

$$W_j = W_{j-1}\frac{K(Q_{j-1},Q_j)}{p(Q_{j-1},Q_j)}.$$

Let $\theta_i\,[\psi]$ be the following stochastic variable:

$$\theta_i\,[\psi] = \left[\frac{\psi(Q_0)}{p(Q_0)}\right] W_i\,\varphi(Q_i).$$

Theorem.
The expectation $M\,(\theta_i\,[\psi])$ of $\theta_i\,[\psi]$ is equal to $\int_G \boldsymbol{K}^i \varphi(P)\psi(P)dP$:

$$M\,(\theta_i\,[\psi]) = \int_G \boldsymbol{K}^i \varphi(P)\psi(P)dP$$

The proof is straightforward consequence of the definition of $T_i(Q_0 \to Q_1 \to Q_2 \to Q_3 ... \to Q_i)$.

Therefore the integral $\int_G \boldsymbol{K}^i\varphi(P)\psi(P)dP$ can be computed approximately by sampling N trajectories $\theta_i[\psi]_s$:

$$\int_G \boldsymbol{K}^i\varphi(P)\psi(P)dP \approx \frac{1}{N}\sum_{s=1}^{N}\theta_i[\psi]_s$$

where the index s means the number of the realization of $\theta_i[\psi]$.

Corollary.

The same set of N trajectories $T_i(Q_0 \to Q_1 \to Q_2 \to Q_3... \to Q_i)$ can be used for computing all $\int_G \boldsymbol{K}^l\varphi(P)\psi(P)dP$ for $1 \le l \le i$.

2.2 Trajectories with fixed initial point.

For computing $\boldsymbol{K}^i\varphi(P_0)$ in a particular point we need to take $\psi(P) = \delta(P - P_0)$. It means that the initial point in all trajectories T_i will be $Q_0 = P_0$. The disadvantage of this approach is that for computing values of $\boldsymbol{K}^i\varphi$ in another point the whole modeling of trajectories should repeated again. Taking as ψ the characteristic function of a subdomain $G' \subset G$ can give the mean value of $\boldsymbol{K}^i\varphi$ in this subdomain.

2.3 Trajectories with absorption.

One can define trajectories for computing multiple integrals $\int_G \boldsymbol{K}^i\varphi(P)\psi(P)dP$ in a slightly different way. Namely the modified algorithm is the following:

We define in G a stochastic "trajectory" $\widetilde{T}_\nu(Q_0 \to Q_1 \to Q_2 \to Q_3... \to Q_\nu)$ where::

a) Q_0 is chosen stochastically with density $p(P)$,

b) After each step Q_j the trajectory will be finished (absorbed) with the probability of "absorption" $a(Q_j)$ such that $0 < a(Q_j) < 1$.With probability $s(Q_j) = 1 - a(Q_j)$ the trajectory will continue.

c) point Q_j following the point Q_{j-1} is chosen with density $p(Q_{j-1}, P)$ with respect to P.

Therefore the length ν of the trajectory will be a stochastic number. We choose weights $\widetilde{W_j}$ for a trajectory $\widetilde{T}_\nu$ of the length ν in the following way:

$$\widetilde{W_0} = 1; \widetilde{W_j} = \widetilde{W}_{j-1}\left[\frac{K(Q_{j-1}, Q_j)}{s(Q_j)p(Q_{j-1}, Q_j)}\right], \quad j \le \nu.$$

Therefore $W_j(P_0, P_1, ..., P_j) = \widetilde{W}_j(P_0, P_1, ..., P_j)s(P_0)s(P_1), ..., s(P_{j-1})$. We introduce a stochastic variable $\widetilde{\theta}_i[\psi]$ dependent on trajectories $\widetilde{T}_i$ with absorption

$$\widetilde{\theta}_i[\psi] = \frac{\psi(Q_0)}{p(Q_0)}\widetilde{W_i}\left[\frac{\varphi(Q_i)}{a(Q_i)}\right].$$

In the same way as above for the expectation $M\left(\widetilde{\theta}_i[\psi]\,|\nu = i\right)$

$$M\left(\widetilde{\theta}_i[\psi]\,|\nu = i\right) = \int_G \boldsymbol{K}^i\varphi(P)\psi(P)dP\left[P\{\nu = i\}\right]^{-1}$$

where the probability $P\{\nu = i\}$ for a trajectory to have length ν. By modeling N such trajectories we will get N_i of those of the length i. Evidently

$$N_i/N \approx P\{\nu = i\}.$$

Therefore

$$\int_G \boldsymbol{K}^i \varphi(P)(P)\psi(P)dP \approx \frac{P\{\nu = i\}}{N_i} \sum_{s=1}^{N_i} \widetilde{\theta}_i [\psi]_s \approx \frac{1}{N} \sum_{s=1}^{N_i} \widetilde{\theta}_i [\psi]_s$$

where all trajectories of the length i are taken in the sum.

2.4 Monte Carlo computation of solutions to integral equations.

It is easy to observe that the integrals in the iterations

$$z^{(i)} = f + \boldsymbol{K}f + \boldsymbol{K}^2 f + \boldsymbol{K}^3 f ... + \boldsymbol{K}^{i-1} f + \boldsymbol{K}^i \varphi.$$

for approximate solutions to the integral equation of the second type

$$z(P) = \int_G K(P, P')z(P')dP' + f(P)$$

can be approximately computed using the same set of N stochastic trajectories $T_i(Q_0 \to Q_1 \to Q_2 \to Q_3 ... \to Q_i)$ as above.

We introduce a stochastic variable

$$\zeta_i [\psi] = \left[\frac{\psi(Q_0)}{p(Q_0)}\right] \left[\sum_{j=0}^{i-1} W_j f(Q_j) + W_i \varphi(Q_i)\right]$$

with

$$W_j = W_{j-1} \frac{K(Q_{j-1}, Q_j)}{p(Q_{j-1}, Q_j)}, \quad W_0 = 1$$

Theorem.

$$M\,\zeta_i [\psi] = \int_G \psi(P) z^{(i)}(P)dP \approx \frac{1}{N} \sum_{s=1}^{N} \zeta_i [\psi]_s$$

With the same trajectories one can compute $\int_G \psi(P) z^{(i)}(P)dP$ for different ψ and f. For different ψ with the same f even the computations of $\sum_{j=0}^{i-1} W_j f(Q_j)$ in expressions for $\zeta_i [\psi]_s$ will be the same.

The minimal variance and the theoretically best convergence for this method will be in the case if the integrals in the approximation are computed by the importance sampling approach. In the particular case with the equation it means that if $K(P, P') \geq 0$, $z(P) > 0$ the initial density $p(P)$ and the transition probability $p(P, P')$ are chosen as:

$$\begin{aligned} p(P) &= \frac{|\psi(P)|\,\varphi(P)}{\int_G |\psi(P)|\,\varphi(P)dP} \\ p(P, P') &= \frac{K(P, P')\varphi(P')}{\boldsymbol{K}\varphi(P)} \end{aligned}$$

and that $\varphi(P)$ is equal to the solution $z(P)$. This shows that the initial approximation $\varphi(P)$ is good to choose as close as possible to the actual solution $z(P)$.

2.5 Infinite trajectories.

One can also use "infinite" trajectories not prescribing in advance the length of the trajectories one wish to use in estimates.

$$\begin{aligned} \zeta\left[\psi\right] &= \left[\frac{\psi(Q_0)}{p(Q_0)}\right]\left[\sum_{j=0}^{+\infty} W_j\, f(Q_j)\right] \\ \int_G z(P)\psi(P)dP &\approx \frac{1}{N}\sum_{s=1}^{N}\zeta\left[\psi\right]_s \end{aligned}$$

In practice the infinite sum in the expression for $\zeta\left[\psi\right]$ must be catted using some thresholding.

2.6 Conjugate integral equation and the "forward" Monte Carlo method.

We considering here the equation

$$u(P) = \int_G K^*(P,P')u(P')dP' + \psi(P)$$

with kernel $K^*(P,P') = K(P',P)$, or generally

$$u = \boldsymbol{K}^* u + \psi$$

with the operator $\boldsymbol{K}^*$ where arguments P and P' in the kernel changed places in comparison with the original equation.

$$z = \boldsymbol{K} z + f.$$

This equation is useful for inventing alternative methods for computing functionals $\int_G \psi(P)z(P)dP$ of solutions z to the original equation. It is easy to observe that

$$\int_G u\, z\, dP = \int_G u\left[\boldsymbol{K} z + f\right] dP = \int_G z\, \boldsymbol{K}^* u + f\, u\, dP$$

$$\begin{aligned} \int_G u\, z\, dP &= \int_G z\left[\boldsymbol{K}^* u + \psi\right] dP = \int_G z\boldsymbol{K}^* u + \psi\, z dP \\ &\rightarrow \int_G \psi\, z dP = \int_G f\, u\, dP \end{aligned}$$

and therefore a functional of z can be computed by computing another functional for the solution u of the conjugate equation

$$u = \boldsymbol{K}^* u + \psi.$$

Introducing trajectories with probability densities $p(P)$ and $p(P,P')$ acceptable for $f(P)$ and $K^*(P,P') = K(P',P)$ we get a monte Carlo method for computing $\int_G f\,u\,dP = \int_G \psi\,z dP$.

$$\begin{aligned} \zeta^*[f] &= \frac{f(Q_0)}{p(Q_0)}\left[\sum_{j=0}^{+\infty} W_j^*\,\psi(Q_j)\right] \\ W_j^* &= W_{j-1}^*\left[\frac{K^*(Q_{j-1},Q_j)}{P(Q_{j-1},Q_j)}\right], \quad W_0^* = 1. \end{aligned}$$

$$M\,\zeta^*[f] = \int_G \psi\,z dP = \int_G f\,u\,dP \approx \frac{1}{N}\sum_{s=1}^{N} \zeta^*[f]_s$$

where $\zeta^*[f]_s$ is the value of $\zeta^*[f]$ on the s-th trajectory. Even for symmetric kernels $K(P,P') = K(P',P)$ when $W_j = W_j^*$ the last estimate differs from our previous estimates with $\zeta[\psi]$:

$$\begin{aligned} \zeta[\psi] &= \frac{\psi(Q_0)}{P(Q_0)}\left[\sum_{j=0}^{+\infty} W_j\,f(Q_j)\right], \\ \zeta^*[f] &= \frac{f(Q_0)}{p(Q_0)}\left[\sum_{j=0}^{+\infty} W_j\,\psi(Q_j)\right]. \end{aligned}$$

For computation of $\zeta[\psi]$ one calculates ψ one time and f many times. For computation of $\zeta^*[f]$ one calculates f one time and ψ many times. On the other hand the method using the conjugate equation with estimate of $\zeta^*[f]$ lets to solve equations with source f concentrated in a point P_0 : $f(P-P_0) = \delta(p-P_0)$. One just constructs $\zeta^*[f]$ with fixed initial point $Q_0 = P_0$.

2.7 Example with scattering of particles. Continuation.

The above example with scattering of particles:

$$z(P) = \int K_{\mathrm{col}\,l}(P',P)z(P')dP' + f(P)$$

can be solved by both methods considered above. On the other hand the method with conjugate equation has a simple physical interpretation as a method with "true trajectories" or the method of direct stochastic modeling.

The probability $s(P')$ that a prticle after a collision in P' will collide again is

$$s(P') = \int_G K_{\mathrm{col}\,l}(P',P)dP.$$

Therefore $s(P')$ is the probability of scattering in the point P'. $a(P') = 1 - s(P')$ is the probability that the particle will not collide more and disappears from the consideration, will be "absorbed", because we consider the density $z(P)$ of collisions.

"True" trajectories of paticles will be trajectories with absorption like $\widetilde{T}_\nu$ above with transition probability

$$p(P, P') = K_{\mathrm{col}\,l}(P, P')/s(P)$$

with "true" absorption probability $a(P) = 1 - s(P)$. This density is always acceptable for the kernel of the equation.Therefore is it usefult to construct an estimate for functional $\int_G f\ u\, dP$ instead of the functional $\int_G \psi\, z\ dP$.

Corresponding stochastic variable for computation is formally:

$$\widetilde{\zeta}^*_\nu[f] = \frac{f(Q_0)}{p(Q_0)} \widetilde{W}^*_\nu \left[\frac{\psi(Q_\nu)}{a(Q_\nu)}\right].$$

But for each i

$$\widetilde{W}^*_j = \prod_{j=1}^{i} \frac{K_{\mathrm{col}\,l}(Q_{j-1}, Q_j)}{s(Q_{j-1})\, p(Q_{j-1}, Q_j)} = 1 \quad !!!$$

Therefore in this case the stochastic variable looks particularly simple:

$$\widetilde{\zeta}^*_\nu[f] = \frac{f(Q_0)}{p(Q_0)} \left[\frac{\psi(Q_\nu)}{a(Q_\nu)}\right],$$

where ν is the (stochastic) length of the trajectory.

$$M\,\widetilde{\zeta}^*_\nu[f] = \int_G \psi\, z\ dP \approx \frac{1}{N_i} \sum_{s=1}^{N} \widetilde{\zeta}^*_\nu[f]_s\,.$$