Abstract

Introduction, the basics of Monte Carlo Integration, and the elements of statistical physics (part 1).

Introduction

add the intro from linux

Integration by Simulation

The Basic Idea

Let $x \in R^p, g(x) > 0$ with

$$\int g(x)dx = 1.$$

The integral of any real valued integrable function f can be written as:

$$I = \int f(x)dx = \int \frac{f(x)}{g(x)} g(x)dx$$

The function g is a probability density. Thus, the last integral can be read as an expectation with respect to the pdf g, i.e.,

$$I = E_g \left[\frac{f(X)}{g(X)} \right]$$

The classical Strong Law of Large Numbers (SLLN) then assures that for n sufficiently large,

$$I \approx \frac{1}{n} \sum_{t=0}^{n-1} \frac{f(X_t)}{g(X_t)} = I_n$$

with probability close to one, when $X_0, X_1, \ldots$ are iid as X with pdf g. This is just the standard: "sample mean goes to the expected value" kind of statement. Moreover, by the Central Limit Theorem (CLT),

$$Z_n = \frac{n^{1/2}(I_n - I)}{\sigma}$$

goes (in Law) to the standard Gaussian distribution, provided that σ (which is the standard deviation of the r.v. $Y = f(X)/g(X)$), is finite and not zero. The size of σ controls the accuracy of the approximation. Notice, that I_n give or take,

$$1.96 \frac{\sigma}{n^{1/2}}$$

is a random interval with about 95% chance of covering the value of I. The smaller the value of σ, the more accurate is I_n as an estimator of I.

Best g

The best g is then the one that makes σ as little as possible. We have,

$$\sigma^2 = E_g \left[\frac{f(X)}{g(X)}\right]^2 - I^2$$

The integral of f, I, is independent of g so the best g is obtained by solving the variational problem,

$$\min \int \frac{f^2(x)}{g(x)} dx$$

over all positive pdfs g, i.e. functions such that, $g(x) > 0$ and $\int g = 1$. Using a Lagrange multiplier λ for the normalization constraint, we see that we need to find g and λ solving,

$$\min \int \left[\frac{f^2(x)}{g(x)} + \lambda g(x)\right] dx.$$

The Euler-Lagrange equation for the Lagrangian:

$$\mathcal{L}(g, \lambda) = \frac{f^2}{g} + \lambda g$$

is just the derivative of $\mathcal{L}$ w.r.t. g equal 0. This gives,

$$-\frac{f^2}{g^2} + \lambda = 0$$

from where we deduce that the best g is,

$$g^*(x) = \frac{1}{c}|f(x)|$$

where c is a normalization constant (the square root of λ actually). This is clearly a global minimum for σ (at least when f is positive) since,

$$(\sigma^*)^2 = c \int |f(x)| dx - I^2$$

and this is zero, due to the fact that $c = I$ for positive functions f.

Example

Take,

$$I = \int_{-1}^{1} x^2 dx = 0.666\ldots$$

Best $g(x) = 1.5x^2$ for x in the interval $[-1, 1]$. So for this g, $I_n = I$ for all n and the Monte Carlo algorithm is really not useful. We need to know I to implement the algorithm but the purpose of the algorithm is to compute I itself!. This is the problem with the optimal g when f is every where positive. Nevertheless, by knowing that the optimal g must follow the shape of $|f(x)|$ we can tune up g to gain efficiency. For example, $g(x) = |x|$, that goes down and up with x^2, will be better than the uniform $g(x) = 1/2$ in the interval $[-1, 1]$. In fact, to obtain a given accuracy we need to generate more than 6 times as many iterations with the uniform than with $|x|$.

- Look at the first 200 iterations using $g(x) = 1/2$

  (*http://omega.stat.psu.edu:8008/summer99/lecture1/x2a-js.html*) Look at the Javascript for this problem.

  (*http://omega.stat.psu.edu:8008/summer99/lecture1/x2a-js.txt*)

- Look at the first 200 iterations using $g(x) = |x|$

  (*http://omega.stat.psu.edu:8008/summer99/lecture1/x2b-js.html*) Look at the Javascript for this problem.

  (*http://omega.stat.psu.edu:8008/summer99/lecture1/x2b-js.txt*)

In the implementation of the previous examples we have used the following fundamental property of random variables for generating samples from g:

Theorem 1 *Let $U_1, U_2, \ldots, U_n$ be iid uniform on $[0, 1]$ and let F be a distribution function. Then,*

$$F^{-1}(U_1), F^{-1}(U_2), \ldots, F^{-1}(U_n)$$

are iid with cdf F

Proof

$$\begin{aligned} P[F^{-1}(U_1) \leq y_1, F^{-1}(U_2) \leq y_2, \ldots, F^{-1}(U_n) \leq y_n] &= P[U_1 \leq F(y_1), \ldots, U_n \leq F(y_n)] \\ &= F(y_1)F(y_2)\ldots F(y_n) \end{aligned}$$

The first equality follows from the fact that F^{-1} is always non-decreasing and the last equality is just the assumed hypothesis of iid uniform U_j.

The Elements of Statistical Physics

The first Markov Chain Monte Carlo algorithm, (The Metropolis algorithm) appeared in the statistical physics literature. The aim was to simulate the evolution of a solid in a heat bath towards thermal equilibrium.

In this section we introduce the main ideas, and notation from statistical physics that will be used in the rest of the course. Statistical Mechanics has been a continuous source of innovative ideas in mathematics and statistics but it is often not included as a required course for graduate students in mathematical statistics. This one-hour introduction will help to fill the gap.

From Newton to Hamilton

The formulation of classical mechanics evolved from the original laws of Newton, the most famous (and computationally most useful) being the second law:

$$F = ma$$

i.e. Force equals mass times acceleration. The acceleration being the second derivative with respect to time of the position q, denoted by,

$$a = \ddot{q}$$

For a single particle, q denotes its position vector (e.g. (x, y, z) in Cartesian coordinates) and for a system of particles q denotes the long vector with all the position coordinates for all the particles. General field forces are arbitrary vector functions of position and velocity but all the four fundamental forces of nature (gravitational, electro magnetic, weak and strong) preserve energy and are therefore conservative and coming from gradients w.r.t. q of potential functions V. Thus, we are only interested in field forces such that,

$$F = -\frac{\partial V}{\partial q}$$

for some scalar potential function,

$$V = V(q, \dot{q})$$

Newton's laws tell us that the evolution of a system of particles is governed by a second order system of differential equations:

$$F(q, \dot{q}) = m\ddot{q}$$

For a system of N particles without constraints, there are $3N$ second order differential equations to be solved. The theory of differential equations (developed in great part to understand mechanics) shows that under mild regularity conditions on the functions F and $q(t)$, the system has a unique solution for a given set of initial conditions (for example for initial values of positions and velocities for each particle). A second order system can always be reduced to a first order system by duplicating the number of equations. By introducing the momentum p,

$$p = m\dot{q}$$

or equivalently,

$$\dot{q} = \frac{p}{m}$$

we can now replace the original system of $3N$ second order differential equations by an equivalent first order system of $6N$ equations in the variables p and q. Just augment the previous equations with the original ones (but now written with only first derivatives in terms of q and p),

$$\dot{p} = F(q, p/m).$$

Hamiltonian Formulation

By introducing the function $H(q,p)$ representing the total energy of the system, i.e. the sum of kinetic and potential energies,

$$\begin{aligned} H(q,p) &= Energy = Kinetic + Potential \\ &= \frac{1}{2}m\dot{q}^2 + V \\ &= \frac{p^2}{2m} + V \end{aligned}$$

we obtain Hamilton's equations by replacing the right hand side of the system above with the derivatives of H:

$$\begin{aligned} \dot{p} &= -\frac{\partial H}{\partial q} \\ \dot{q} &= \frac{\partial H}{\partial p} \end{aligned}$$

The Hamiltonian formulation of classical mechanics has proven extremely useful for the development of modern physics but it contains as much information as the original laws of Newton. As before, given the initial conditions of position $q(0)$ and momentum $p(0)$ at time $t = 0$, Hamilton's equations predict the past and future of the system with complete certainty. The problem is that for macroscopic systems the number of particles N is of the order of Avogadro's number,

$$N \sim 10^{24}$$

and we need to provide of the order of $6 * 10^{24}$ initial conditions and to solve the same number of first order differential equations to be able to "see" the truth implied by the equations of classical mechanics. The tremendous size of the complexity of this task is the origin of statistical physics.