Advanced Monte Carlo Methods

Non-Uniform Variate Generation

Recall CDF

$$f(x) = P(\xi \leq x)$$

$0 \leq f(x) \leq 1$, nondecreasing
right-continuous

Not all random variables have a simple CDF!

General Assumption:

can cheaply generate $x_i \sim U[0,1)$

Given x_i solve for y_i

$$f(y_i) = x_i$$

$$0 \leq \uparrow \leq 1$$

so $y_i = f^{-1}(x_i)$, $y \sim \eta$

with $f(\cdot)$ as CDF

This method is <u>inversion.</u>

Example:

$$p(x) = \lambda e^{-\lambda x} \qquad 0 \leq x < +\infty$$

$$f(x) = 0 \qquad x < 0$$

$$= \int_0^x \lambda e^{-\lambda y} dy \qquad x \geq 0$$

use $u = -\lambda y$; $du = -\lambda dy$

$$= \int_0^{-\lambda x} -e^u du = \int_{-\lambda x}^0 e^u du$$

$$= e^u \Big|_{-\lambda x}^0 = 1 - e^{-\lambda x}$$

given $x_i \sim U[0,1)$ solve

$$1 - e^{-\lambda y_i} = x_i$$

$$\underbrace{1 - x_i}_{U[0,1)} = e^{-\lambda y_i}$$

$$\tilde{X}_i = e^{-\lambda y_i}$$

$$-\frac{1}{\lambda}\ln(\tilde{x}_i) = y_i \sim \exp(\lambda)$$

Example: Cauchy distr.

$$p(x) = \frac{1}{\pi}\frac{b}{(x-m)^2+b^2} \quad -\infty \le x \le +\infty$$

Cauchy(m, b)

median $\quad$ half width

$$f(x) = \int_{-\infty}^{x} \frac{1}{\pi}\frac{b}{(y-m)+b^2}\,dy$$

let $u = \frac{y-m}{b} \Rightarrow (ub)^2 = (y-m)$

$$\Rightarrow b\,du = dy$$

$$= \int_{-\infty}^{\frac{x-m}{b}} \frac{1}{\pi}\frac{b}{u^2b^2+b^2}\,b\,du = \frac{1}{\pi}\int_{-\infty}^{\frac{x-m}{b}} \frac{du}{u^2+1}$$

$$= \frac{1}{\pi}\arctan\Big|_{-\infty}^{\frac{x-m}{b}} = \frac{1}{\pi}\left(\arctan\left(\frac{x-m}{b}\right) - -\frac{\pi}{2}\right)$$

$$= \frac{1}{2} + \arctan\left(\frac{x-m}{b}\right), \text{ solve}$$

$$x_i = \frac{1}{2} + \arctan\left(\frac{y_i-m}{b}\right)$$

$$x_i - \frac{1}{2} = \arctan\left(\frac{y_i - m}{b}\right)$$

$$\tan\left(x_i - \frac{1}{2}\right) = \frac{y_i - m}{b}$$

$$y_i = m + b\tan\left(x_i - \frac{1}{2}\right)$$

$$\sim \text{Cauchy}(m, b)$$

Example:

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

note; CDF is represented in terms of the "error function" and is not too easy to invert.

$$x_1 \sim N(0,1) \sim x_2$$

$$p(x_1, x_2) = p(x_1) \cdot p(x_2)$$

$$= \frac{1}{2\pi} e^{-\left(\frac{x_1^2 + x_2^2}{2}\right)}$$

Show $\int_{-\infty}^{\infty} p(x)\,dx = 1$, but

$$\int_{-\infty}^{\infty} dx_1\, p(x_1) \int_{-\infty}^{\infty} dx_2\, p(x_2) = N^2 =$$

$$\int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2 \; p(x_1, x_2) = \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2 \frac{1}{2\pi} e^{-\frac{(x_1^2 + x_2^2)}{2}}$$

$\rightarrow$ polar : $dx_1 dx_2 = r\,dr\,d\theta$

$x_1^2 + x_2^2 = r^2$

$$= \underbrace{\int_0^{2\pi} \frac{1}{2\pi} d\theta}_{=1} \underbrace{\int_0^{\infty} e^{-\frac{r^2}{2}} r\,dr}$$

$u = \frac{r^2}{2}$

$du = r\,dr$

uniform in θ, $U[0, 2\pi)$

$$= \underbrace{\int_0^{\infty} e^{-u} du}_{\text{radius}^2/2 \text{ is exponential}} = -e^{-u}\Big|_0^{\infty} = 0 - -e^0 = 1$$

so $N^2 = 1$

$N = \pm 1$

So $x_1, x_2 \sim N(0,1)$

$\Longleftrightarrow$ $\theta \sim U[0, 2\pi)$

$\frac{r^2}{2} \sim \exp(1)$

$x_1 = r\cos\theta$

$x_2 = r\sin\theta$

$\theta = 2\pi y_1$

$r = \sqrt{-2\ln y_2}$

$y_1, y_2 \sim U[0,1)$

Given $y_1, y_2 \sim U[0,1)$

$$x_1 = \sqrt{-2 \ln y_1} \cos(2\pi y_2)$$
$$x_2 = \sqrt{-2 \ln y_1} \sin(2\pi y_2)$$

will be $N(0,1)$, this is the famous Box-Muller transformation, and is derivable as an inversion

Acceptance - Rejection Method

General method # 2

We want to sample $x \sim p(x)$, assume you have the ability to sample $y \sim g(x)$.

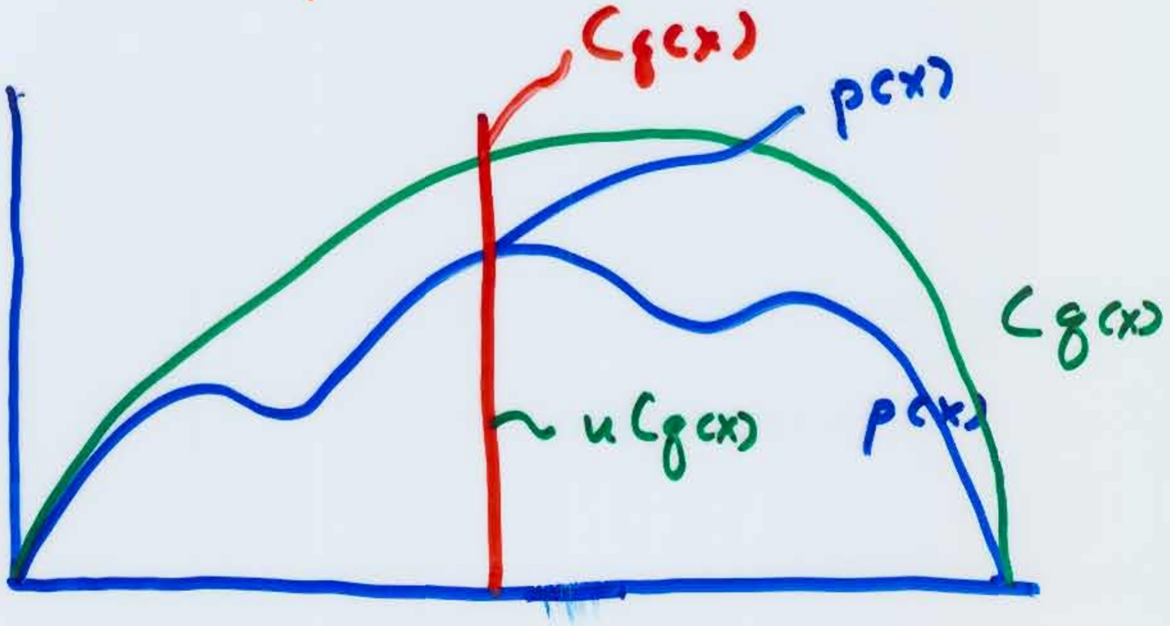
Assume $\forall x \quad p(x) \leq C g(x)$, $C > 1$.

to generate from $p(x)$, generate $x \sim g(x)$, compute $p(x)$ & test if $u C h(x) \leq f(x)$, where

$u \sim U[0,1)$

If $\circledast$ holds $x \sim p(x)$

If $\circledast$ does not hold, reject and resample



$Cg(x)$ generates a point that should be accepted only $\dfrac{p(x)}{Cg(x)}$, so

when $\dfrac{p(x)}{Cg(x)} \geq u \sim U[0,1)$ we

can accept.

Examples of Generation with Acceptance - Rejection:

Gamma:

$$p(x;a) = \frac{x^{a-1} e^{-x}}{\Gamma(a)} \qquad a \in \mathbb{Z}$$

waiting time for a^{th} Poisson event with unit rate, i.e. a^{th} exponential event.

a small — $-\ln(x_i)$ is $\exp(1)$

$$\sum_{i=1}^{a} -\ln(x_i) = -\ln\left(\prod_{i=1}^{a} x_i\right)$$

a big — $p(x;a)$ is more "Bell shaped"

Knuth mentions majoriser of (majorant)

$$g(x) = \frac{1}{\pi} \frac{b}{(x-m)^2+b^2}$$

$$\sim \text{Cauchy}(m,b)$$

Example: Poisson Distribution

$$p(x=j) = \frac{x^j e^{-x}}{j!} \qquad \text{(discrete)}$$

$$p(x) = \frac{x^{\lfloor j \rfloor} e^{-x}}{\lfloor j \rfloor !}$$ is now continuous

use as a majorant a well-known distribution, Cauchy, Normal.

Efficiency of Acceptance - Rejection

Question: what fraction of the time is rejection chosen?

Answer:

Since $p(x) \leq C g(x)$, $C \geq 1$

$$R(x) = C g(x) - p(x) \geq 0$$

$\int_{-\infty}^{\infty} R(x)\,dx$ is "rejection mass"

$\int_{-\infty}^{\infty} C g(x)\,dx$ is "total mass"

Efficiency = fraction of time accepted

= 1 - fraction of time rejected

$$= 1 - \frac{\int_{-\infty}^{\infty} R(x)\,dx}{\int_{-\infty}^{\infty} C g(x)\,dx}$$

$$\int_{-\infty}^{\infty} R(x)\,dx = \int_{-\infty}^{\infty} [C g(x) - p(x)]\,dx = C - 1$$

$$\int_{-\infty}^{\infty} C g(x) = C$$

$$= 1 - \left(\frac{C-1}{C}\right) = 1 - 1 + \frac{1}{C}$$

$$= \frac{1}{C}, \quad C \geq 1$$

So $C = 1$ is perfectly efficient, but C must be chosen so that $C g(x) \geq p(x)$ $\forall x$. This is a trade-off.

Squeeze Principle in Acceptance-Rejection

For acceptance-rejection, need $Cg(x)$ as majorant. Must compare:

$Cg(x)$ to $p(x)$ at each step in a/r

If $p(x)$ is costly to evaluate, we pay a big price $O(C)$ $p(x)$ evaluations!

Let $n(x) \le p(x)$, be a minorant, and be simpler to evaluate than $p(x)$.
Then we modify acceptance-rejection as:

1. Compute $x \sim g(x)$
2. $u \sim U[0,1)$
3. if $u\,Cg(x) \le n(x)$ accept
4. test $u\,Cg(x) \lessgtr p(x)$ $\xrightarrow{y}$ accept, $\rightarrow$ reject

With majorant/minorant "squeeze" pair, one can very efficiently compute using acceptance-rejection, the problem of finding good "squeezes" then becomes an approximation issue.

Good example is UNU.RAN:

Universal Non-Uniform RANdom number generators:

http://statistik.wu-wien.ac.at/unuran

Metropolis Algorithm:

Consider the situation from Statistical Mechanics:

1. Very large phase (sample) space
2. $\forall q \in \mathcal{P}$ we have an energy function $H(q) : \mathcal{P} \to \mathbb{R}$
3. $P(q) = (e^{-H(q)/kT})/Z$

k – Boltzmann constant
T – temperature

$$Z = \int_{\mathcal{P}} e^{-\beta H(q)} dq$$ is the normalizing constant, called the partition function, note $\beta = (kT)^{-1}$, is a common notation

Fact Z is "impossible" to compute, so we have to work with unnormalized probabilities:

$$p(q_1) = Z^{-1} e^{-\beta H(q_1)}$$

$$p(q_2) = Z^{-1} e^{-\beta H(q_2)}$$, but relative probability can be computed:

$$\frac{p(q_1)}{p(q_2)} = e^{-\beta(H(q_1) - H(q_2))}$$

In Statistical Mechanics we want to compute time averages of variables to get macroscopic information of the system

$$\langle A \rangle = \frac{1}{T}\int_0^T A(t)\,dt \quad , \quad T \to \infty$$

dynamics of the system

Ergodic Hypothesis allows replacement of time averages with phase space averages w.r.t. the equilibrium (stationary) pdf:

$$\langle A \rangle = \int_{\mathcal{P}} A(g)\, dg_e$$

Metropolis algorithm allows one to sample from dg_e using a Markov Chain.

Suffices to define transition probabilities

$$P(g_i \to g_j) = p_{ij}$$

if $H(g_i) \geq H(g_j)$ move decreases energy, this move is always accepted; however, local energy minimization (Greedy) get caught in local energy minima, so must have non-zero probability of moving to higher energy state:

$$p_{ij} = \begin{cases} 1 & \text{if } H(g_j) \leq H(g_i) \\ e^{-\beta(H(g_j) - H(g_i))} & \text{otherwise} \end{cases}$$

(more in later sections)