StatsM254 Statistical Methods in Computational Biology Lecture 10 - 05/06/2014

Lecture 10 Simulation Methods

Lecturer: Jingyi Jessica Li *Scribe: Huwenbo Shi*

1 Monte Carlo Method

1. Goal: To evaluate $\theta = E[f(X)]$ for $X \sim P$, where P is the target distribution.
2. Direct Monte Carlo: Sample x_i as i.i.d. from P and take the average $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} f(x_i)$.
   Examples: Can be used to evaluate $E[X^2]$, $P(X > c) = E[\mathbb{1}_{\{x>c\}}]$.
3. Monte Carlo methods are useful for:
   (a) Sampling from posterior in Bayesian inference
   (b) When dimension of the parameter space is high
   (c) When analytic form of the distribution is not available
4. Vanilla Monte Carlo:
   Question: Let $X \perp Y \sim Unif(0, 1)$, what is $P(X^2 + Y^2 \geqslant 1)$?
   Simulate N data points, count the number of data points that satisfy ${x_i}^2 + {y_i}^2 \geqslant 1$.
   Estimate the probability by: $\frac{|S|}{N}$, where $S = \{(x_i, y_i)|{x_i}^2 + {y_i}^2 \geqslant 1, i = 1, \ldots, n\}$.

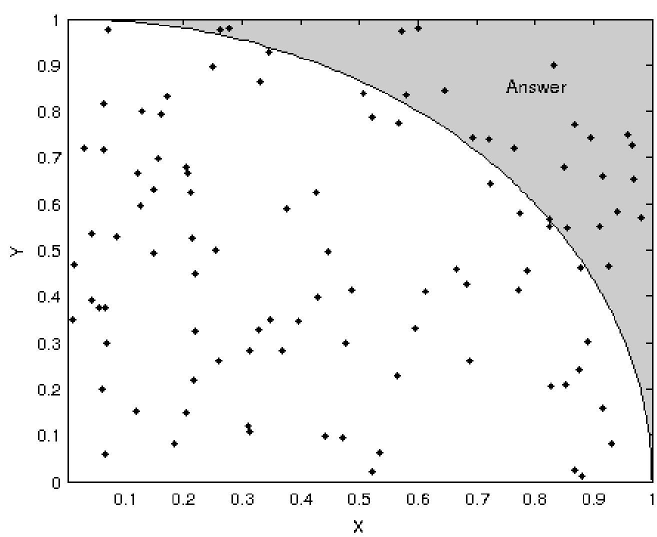

Figure 1: Vanilla Monte Carlo

2 Simulating from A Distribution

1. Suppose the distribution is known with CDF F.
   Theorem: Let $U \sim Unif(0, 1)$ and $X = F^{-1}(U)$, then $X \sim F$.

Proof: $P(X \leqslant x) = P(F^{-1}(U) \leqslant x) = P(U \leqslant F(x)) = F(x)$.

2. Example:
To sample $X \sim Exp(\lambda)$, in R: `x <- rexp(n, lambda)`
First sample $u_1, \dots, u_n \sim Unif(0,1)$.
CDF for exponential distribution is $F(x) = 1 - e^{-\lambda x}, x \geqslant 0$
Let $y = 1 - e^{-\lambda x}$, then $e^{-\lambda x} = 1 - y$, $x = -\frac{1}{\lambda} log(1-y), y \in [0,1)$
So $F^{-1}(x) = -\frac{1}{\lambda} log(1-x), x \in [0,1)$
Let $x_i = -\frac{1}{\lambda} log(1-u_i)$, then $x_1, \dots, x_n$ are samples from $Exp(\lambda)$.

3 Rejection Method (von Neumann 1951)

1. Setting:
   (a) Want to sample from a target distribution with density $\pi(x)$
   (b) The PDF of the target distribution is known up to a constant: $l(x) = c\pi(x)$, where $l(x)$ is known, c and $\pi(x)$ unknown.
   (c) Can construct:
      i. An envelope function $g(x)$
      ii. A constant M such that $Mg(x) \geqslant l(x), \forall x$

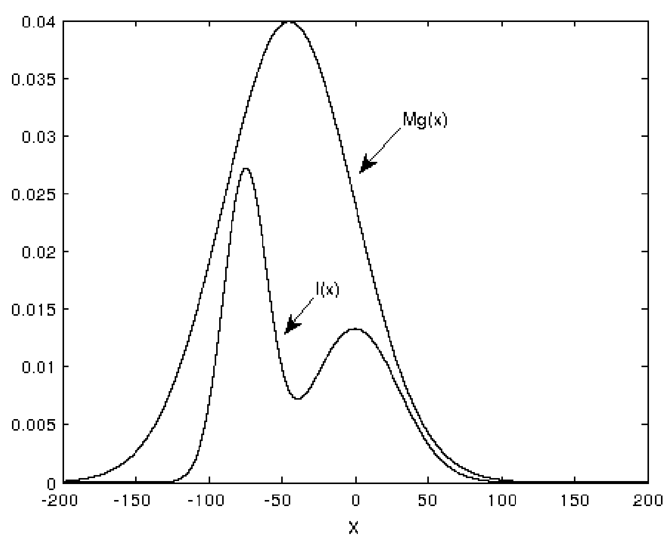

Figure 2: Envelope function

2. Procedure:
   (a) Draw a sample x from $g(x)$ and compute the ratio $r = \frac{l(x)}{Mg(x)} \in [0,1]$
   (b) Flip a coin with success probability r
      i. If head turns up, accept and keep x
      ii. Otherwise, discard x

(c) Go back to step (a) until the n^{th} sample is accepted

3. Why Does It Work?

   Proof: Let $I = \begin{cases} 1 & \text{if sample } x \sim g \text{ is accepted.} \\ 0 & \text{otherwise.} \end{cases}$

   $$P(I=1) = \int P(I=1|X=x)g(x)\,\mathrm{d}x = \int \frac{l(x)}{Mg(x)}g(x)\,\mathrm{d}x = \int \frac{l(x)}{M}\,\mathrm{d}x = \int \frac{c\pi(x)}{M}\,\mathrm{d}x = \frac{c}{M}$$

   $$P(X=x|I=1) = \frac{P(X=x, I=1)}{P(I=1)} = \frac{P(I=1|X=x)g(x)}{P(I=1)} = \frac{\frac{l(x)}{Mg(x)}g(x)}{\frac{c}{M}} = \frac{l(x)}{c} = \pi(x)$$

4. Example: Truncated Gaussian

   Target distribution with density $\pi(x) \propto \phi(x)\mathbb{1}_{\{x>c\}}$, where $\phi(x)$ is the PDF of $N(0,1)$.

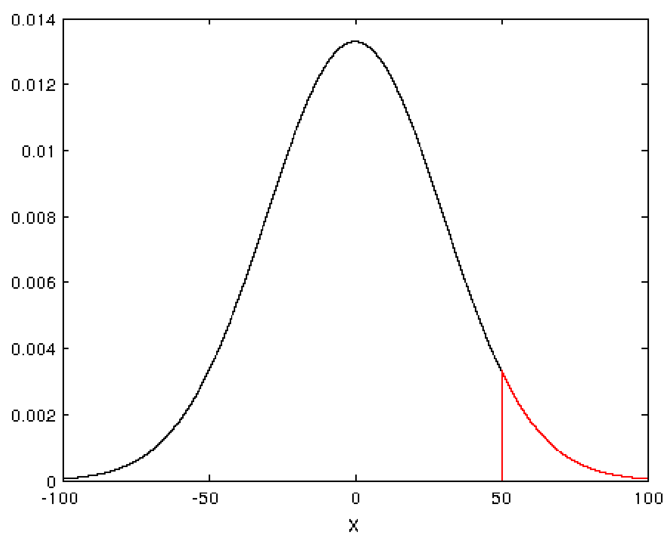

Figure 3: Truncated Gaussian, c=50

Envelope 1: $g(x) = \phi(x)$, $M = 1$
First, we notice that $Mg(x) \geqslant \phi(x), \forall x$.
$r = \frac{\phi(x)\mathbb{1}_{\{x>c\}}}{\phi(x)} = \mathbb{1}_{\{x>c\}}, E[r] = E[\mathbb{1}_{\{x>c\}}] = 1 - \Phi(c)$, where $\Phi(x)$ is the CDF of $N(0,1)$.

Envelope 1 procedure:
Draw x from $N(0,1)$, keep if $x > c$.
Repeat until the n^{th} sample is accepted.

Envelope 2:
Let $g(x) = \lambda \mathrm{e}^{-\lambda(x-c)}, x \geqslant c$.
An envelope function must satisfy $Mg(x) \geqslant \phi(x), x \geqslant c$.
What is the smallest M such that $M \geqslant \frac{\phi(x)}{g(x)}, \forall x \geqslant c$?
$M \geqslant \frac{\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2}}{\lambda \mathrm{e}^{-\lambda(x-c)}} = \frac{1}{\sqrt{2\pi}\lambda}\mathrm{e}^{-(x^2/2-\lambda(x-c))}$

So $M = \min_{x \geqslant c} \frac{1}{\sqrt{2\pi}\lambda} e^{-(x^2/2 - \lambda(x-c))} = \frac{1}{\sqrt{2\pi}\lambda} e^{(\lambda^2/2 - \lambda c)}$
$Mg(x) = \frac{1}{\sqrt{2\pi}} e^{(-\lambda^2/2 - \lambda x)}$
What λ to choose?
We want $\lambda^\star$ that maximizes the acceptance rate r.
maximize $r = \frac{\phi(x)}{Mg(x)} = \frac{\frac{1}{\sqrt{2\pi}} e^{-x^2/2}}{\frac{1}{\sqrt{2\pi}} e^{(-\lambda^2/2 - \lambda x)}}$, subject to $x \geqslant c$, which is equivalent to
maximize $\frac{-x^2}{2} + \lambda x$, subject to $x \geqslant c$, $\lambda^\star = c$, solved.

Envelope 2 procedure:

(a) Sample x from $Exp(c)$, let $y = x + c$

(b) Flip a coin with success rate $\frac{e^{-x^2/2}}{e^{(-c^2/2 - cx)}}$, if success keep y

(c) Go back to step (a) until we accept the n^{th} sample

Acceptance rate comparison for Envelope 1 and Envelope 2:

Envelope	$c = -1$	$c = 2$	$c = 3$
1	0.84	0.02	0.0009
2	0.57	0.88	0.96

Envelope 1 samples from the tail of a Gaussian distribution and has slow acceptance rate, whereas Envelope 2 samples from an exponential distribution and has high acceptance rate.

5. Good Envelope Function Properties:

   (a) Easy to construct

   (b) Easy to sample from

   (c) Close to the target distribution

   (d) Low rejection rate

4 Importance Sampling

1. Goal:
   To evaluate $E_\pi[h(X)], X \sim \pi$

2. Algorithm (Marshall 1956):

   (a) Draw $x_1, \ldots, x_n$ from a trial distribution g

   (b) Calculate the importance weight $w_i = \frac{\pi(x_i)}{g(x_i)}$

   (c) Estimate $E_\pi[h(X)]$ by $\frac{1}{n} \frac{\sum_{i=1}^n h(x_i) w_i}{\sum_{i=1}^n w_i}$

3. Proof:
   $E_\pi[h(X)] = \int h(x)\pi(x)\,dx = \int h(x)\frac{\pi(x)}{g(x)} g(x)\,dx = E_g[h(X)\frac{\pi(X)}{g(X)}], X \sim g$

4. Example:
   $\pi(x) = \frac{\phi(x)}{\int_0^1 \phi(x)\,dx} \mathbb{1}_{\{x \in [0,1]\}}$, where $\phi(x)$ is the PDF of $N(0,1)$.

Vanilla Simulation Approach:
Take draws from $N(0,1)$ and only keep draws in $[0,1]$.
This is inefficient because acceptance rate is $\Phi(1)-\Phi(0)$, where $\Phi(x)$ is the CDF of $N(0,1)$.
The good part: It gives you the actual draws.
The bad part: It rejects draws.

Importance Sampling Approach:
Draw $x_1,\ldots,x_n \sim Unif(0,1)$ so that $g(x)=1, \forall x \in [0,1]$.
The importance weight is $w_i = \frac{\pi(x_i)}{g(x_i)} = \frac{\phi(x_i)}{\int_0^1 \phi(x)dx}$.
The mean is $\frac{1}{n}\sum_{i=1}^n x_i w_i$.
The good part: It doesn't reject any draw.
The bad part: It doesn't give you the actual draws.

5. Another Example of Importance Sampling:
Target: $f(x,y) = 0.5\mathrm{e}^{-90(x-0.5)^2-10(y+0.1)^4}$
Proposal: $g(x,y) \propto 0.5\mathrm{e}^{-90(x-0.5)^2-10(y+0.1)^2}$
This is the density of $N\left(\begin{bmatrix}0.5\\-0.1\end{bmatrix},\begin{bmatrix}1/180 & 0\\0 & 1/20\end{bmatrix}\right)$.
To compute the weights, we use $w(x,y) = \frac{f(x,y)}{g(x,y)}$.

References

[1] J. von Neumann, "Various techniques used in connection with random digits. Monte Carlo methods", Nat. Bureau Standards, 12 (1951), pp. 3638.

[2] Marshall, A. W. (1956). The use of multi-stage sampling schemes in Monte Carlo computations. In M. Meyer (Ed.), Symposium on Monte Carlo Methods, pp. 123-140. New York: Wiley