Sampling methods

Importance sampling

$$I = \int g(x)\, f(x)\, dx \tag{1}$$

$X_1, \ldots, X_N$ i.i.d from $f(X)$

$$I_N = \frac{1}{N} \sum_{i=1}^{N} g(X_i) \tag{2}$$

$$E(I_N) = I \tag{3}$$

$$\text{Var}(I_N) = \frac{1}{N} \text{Var}(g(X)) \tag{4}$$

Rejection sampling

Consider

$$f(x) = l(x)\, h(x) \tag{5}$$

where $h(x)$ can be easily sampled and $l(x)$ is bounded by M, i.e. $l(x) \leq M$ for all x. The joint density is

$$f(x, y) = M\, \mathbf{1}\left(y \leq \frac{l(x)}{M}\right) h(x) \tag{6}$$

and

$$f(x) = \int f(x, y)\, dy = \int M\, \mathbf{1}\left(y \leq \frac{l(x)}{M}\right) h(x)\, dy = l(x)\, h(x) \tag{7}$$

So we can sample x from $h(x)$ and y from $U(0, 1)$ independently and if $y \leq \frac{l(x)}{M}$ then we accept X as coming from $f(x) = l(x)\, h(x)$. If it does not, we try again. It is worth calculating the acceptance probability

$$P\left(y \leq \frac{l(x)}{M}\right) = \int P\left(y \leq \frac{l(x)}{M} \,\middle|\, x\right) h(x)\, dx = \int \frac{l(x)}{M} h(x)\, dx = \frac{1}{M} \int f(x)\, dx = \frac{1}{M} \tag{8}$$

Even if $f(x)$ is unnormalized, we can still compute the acceptance probability as

$$P\left(y \leq \frac{l(x)}{M}\right) = \frac{N}{M} \tag{9}$$

Adaptive rejection sampling

To sample $f(x)$, we can use lines that are upper bound of $f(x)$. In other words, we need to find $h(x)$ that is the envelope function of $f(x)$ such that $h(x) \geq f(x)$ everywhere and the gap between them is as small as possible. So we can make $h(x)$ piecewise linear. For such $h(x)$, we can apply rejection sampling as

$$f(x) = l(x)\, h(x) \tag{10}$$

How we construct the piecewise $h(x)$ is by adaptive rejection sampling: Suppose $f(x)$ is log concave:

$$\frac{d^2 \log f(x)}{dx^2} \geq 0 \tag{11}$$

We will be able to find an upperbound for $\log f(x)$ using lines. Suppose at some stage we have 3 lines as upper bound for $f(x)$, we take a proposal x^* from $h(x)$. If we accept x^*, we resample with the same envelope. If we reject x^*, we adjust the envelope by evaluating $\frac{d}{dx} \log f(x)\, |_{x^*}$, which will give us the tangent at x^*, we can form a new piecewise linear upper bound for $\log(f(x))$ by connecting up previous lines with the tangent.

Ratio of uniforms

To sample $f(x)$, let $a = \max_x \sqrt{f(x)}$ and $b = \max_x x\sqrt{f(x)}$, then take $u_1 \sim U(0, a)$, $u_2 \sim U\left(\min_x x\sqrt{f(x)}, b\right)$. We accept $X = \frac{U_2}{U_1}$ as coming from $f(X)$ if $U_1 \leq \sqrt{f(U_2/U_1)}$. To show that this algorithm is correct, we examine the joint density

$$f(x, y) \propto 2\, y\, \mathbf{1}\left(y < \sqrt{f(x)}\right) \tag{12}$$

So

$$f(x) = \int f(x, y)\, dy \tag{13}$$

Now we use the transformation

$$y = U_1,\ x = U_2/U_1 \tag{14}$$

Then

$$f(U_1, U_2) \propto 2\, U_1\, \mathbf{1}\left(U_1 \leq \sqrt{f(U_2/U_1)}\right) \left|J\right| \tag{15}$$

where

$$|J| = \left|\begin{pmatrix} \frac{\partial x}{\partial u_1} & \frac{\partial x}{\partial u_2} \\ \frac{\partial y}{\partial u_1} & \frac{\partial y}{\partial u_2} \end{pmatrix}\right| = \left|\begin{pmatrix} -U_2/U_1^2 & 1/U_1 \\ 1 & 0 \end{pmatrix}\right| = 1/U_1 \tag{16}$$

So if we can sample uniformly from the interval

$$S = \left\{(U_1, U_2) \,\middle|\, U_1 \leq \sqrt{f(U_2/U_1)}\right\} \tag{17}$$

then we can take $X = U_2/U_1$ as coming from $f(x)$. Obviously, the region is bounded by a rectangle. To find the suitable rectangle, we notice that

- $u_1 \leq \max_x \sqrt{f(x)}$, so we sample u_1 from Uniform$\left(0, \max_x \sqrt{f(x)}\right)$.
- If $x \geq 0$, we have $U_2 = Y\,X$ and $Y = U_1 \leq \sqrt{f(x)}$, so $U_2 \leq \max_x x\sqrt{f(x)}$
- If x can be negative, then $U_2 \geq \min_x x\sqrt{f(x)}$

Overall, $U_2 \sim \left(\min_x x\sqrt{f(x)}, \max_x x\sqrt{f(x)}\right)$. So we can implement a ratio of uniforms if $f(x)$ has an upper bound and $\left|x\right|\sqrt{f(x)}$ has an upper limit. Clearly, the probability of acceptance is given by $\frac{S}{R}$.

Normalized Importance Sampling

Suppose we want to estimate

$$E_f\, g(x) = \int g(x)\, f(x)\, dx \tag{18}$$

but we cannot sample f directly and we only know f^* that is proportional to f up to normalization, then in order to do importance sampling, we could use $h(x)$ which is easy to sample:

$$E_f\, g(x) = \int g(x)\, f(x)\, dx = \int g(x) \frac{f(x)}{h(x)} h(x)\, dx = \frac{\int g(x) \frac{f^*(x)}{h(x)} h(x)\, dx}{\int \frac{f^*(x)}{h(x)} h(x)\, dx} \tag{19}$$

$$I_N = \frac{\frac{1}{N}\sum_{i=1}^N \frac{g(x_i) f^*(x_i)}{h(x_i)}}{\frac{1}{N}\sum_{i=1}^N \frac{f^*(x_i)}{h(x_i)}} \tag{20}$$

Markov Chain Monte Carlo Sampling

Stationary condition:

$$\pi(y) = \int p(y \mid x)\, \pi(x)\, dx \tag{21}$$

where $\pi(x)$ is the target density. Our goal is to find $p(x_{n+1} \mid x_n)$.

Conditional probability

We introduce a conditional density $\pi(u \mid x)$ so that we have a joint pdf

$$\pi(u, x) = \pi(u \mid x)\, \pi(x) \tag{22}$$

Then

$$p(x_{n+1} \mid x_n) = \int \pi(x_{n+1} \mid u)\, \pi(u \mid x_n)\, du \tag{23}$$

To sample from $p(x_{n+1} \mid x_n)$, we sample u from $\pi(u \mid x_n)$ and then sample x_{n+1} from $\pi(x_{n+1} \mid u)$. The remaining question is how to construct $\pi(u, x)$. We can use similar idea as in rejection sampling:

$$\pi(u, x) = \mathbf{1}\,(u \leq \pi(x)) \tag{24}$$

Then sampling $\pi(u \mid x_n)$ can be done by sampling u from Uniform$(0, \pi(x_n))$, i.e. sampling V_n from Uniform$(0, 1)$ and set $u = \pi(x_n)\, V_n$; On the other hand, we can sample x_{n+1} from $\pi(x_{n+1} \mid u)$ by sampling from Uniform$(\pi(x_{n+1}) \geq u)$, i.e. solve for the range of X_{n+1}, sampling W from Uniform$(0, 1)$ and set

$$X_{n+1} = \psi(X_n, W_n, V_n) \tag{25}$$

where W_n and V_n are i.i.d. Uniform$(0, 1)$

Metropolis-Hastings algorithm

We construct a transition density as follows

$$p(x' \mid x) = \alpha(x', x)\, q(x' \mid x) + (1 - r(x))\, \mathbf{1}\,(x' = x) \tag{26}$$

where

- $q(x' \mid x)$ is the proposal density.
- $\alpha(x', x) = \min\left\{1, \frac{\pi(x')\, q(x|x')}{\pi(x)\, q(x'|x)}\right\}$
- $r(x) = \int \alpha(x', x)\, q(x' \mid x)\, dx'$

It is easy to show that

$$p(x' \mid x)\, \pi(x) = p(x \mid x')\, \pi(x') \tag{27}$$

The MH sampling algorithm is

- Take x^* from $q(x' \mid x)$
- Take u from $U(0, 1)$
- If $u < \alpha(x^*, x)$, then $x' = x^*$ else $x' = x$

To trigger off the chain, x_0 can be anything that $\pi(x_0) > 0$.

Barker algorithm

Gibbs Sampler

Suppose $\boldsymbol{x}$ is p dimensional, i.e. $\boldsymbol{x} = (x_1, \ldots, x_p)$ and $\pi(\boldsymbol{x})$ is the target stationary density. A MH algorithm might struggle as hard to find a suitable proposal for all of $\boldsymbol{x}$. In such a case the common structure for Markov chain is the Gibbs sampler. It works by finding the **full conditional densities** of $\pi(\boldsymbol{x})$, i.e. for each j, $\pi_j(x_j \mid x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_p)$. Then the transition density is

$$\begin{aligned} p(\mathbf{x}' \mid \mathbf{x}) = {} & \pi_p\left(x_p' \mid x_1', \ldots, x_{p-1}'\right) \pi_{p-1}\left(x_{p-1}' \mid x_1', \ldots, x_{p-2}', x_p\right) \\ & \pi_{p-2}\left(x_{p-2}' \mid x_1', \ldots, x_{p-3}', x_{p-1}, x_p\right) \ldots \pi_2(x_2' \mid x_1', x_3, \ldots, x_p)\, \pi_1(x_1' \mid x_2, \ldots, x_p) \end{aligned} \tag{28}$$

which satisfies

$$\pi(\mathbf{x}') = \int p(\mathbf{x}' \mid \mathbf{x})\, \pi(\mathbf{x})\, d\mathbf{x} \tag{29}$$

In particular, $p = 2$

$$p(x_1', x_2' \mid x_1, x_2) = \pi(x_2' \mid x_1')\, \pi(x_1' \mid x_2) \tag{30}$$

Then

$$\int p(x_1', x_2' \mid x_1, x_2)\, \pi(x_1, x_2)\, dx_1\, dx_2 = \int \pi(x_2' \mid x_1')\, \pi(x_1' \mid x_2)\, \pi(x_1, x_2)\, dx_1\, dx_2$$

$$= \int \pi(x_2' \mid x_1')\, \pi(x_1' \mid x_2)\, \pi(x_2)\, dx_2$$

$$= \pi(x_2' \mid x_1')\, \pi(x_1')$$

$$= \pi(x_1', x_2')$$