Advanced Simulation Methods

Chapter 2 - Inversion Method, Transformation Methods and Rejection Sampling

We consider here the following generic problem. Given a "target" distribution of probability density or probability mass function π, we want to find an algorithm which produces random samples from this distribution. We discuss here the standard techniques which are used in most software packages.

1 Inversion Method

Consider a real-valued random variable X and its associated cumulative distribution function (cdf)

$$F(x) = \mathbb{P}(X \leq x) = F(x)$$

The cdf $F : \mathbb{R} \to [0, 1]$ is
- increasing; i.e. if $x \leq y$ then $F(x) \leq F(y)$
- right continuous; i.e. $F(x + \epsilon) \to F(x)$ as $\epsilon \to 0$ ($\epsilon > 0$)
- $F(x) \to 0$ as $x \to -\infty$ and $F(x) \to 1$ as $x \to +\infty$.

We define the generalised inverse

$$F^{-}(u) = \inf \{x \in \mathbb{R}; F(x) \geq u\}$$

also known as the quantile function. Its definition is illustrated by Figure 1. Note that $F^{-}(u) = F^{-1}(u)$ if F is continuous.

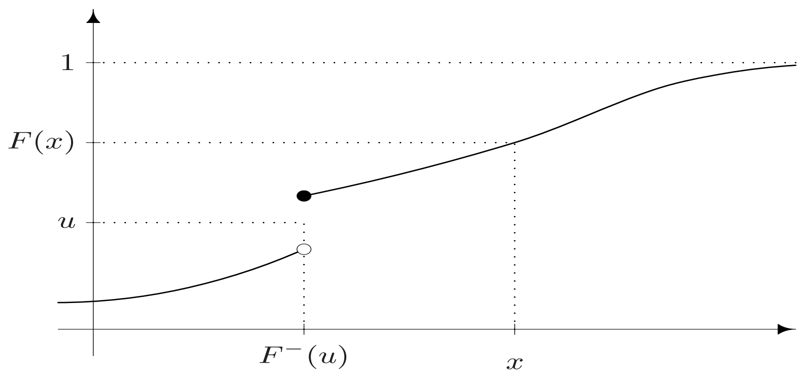

Figure 1: Illustration of the definition of the generalised inverse F^{-}

Proposition 1 ***(Inversion method).*** *Let F be a cdf and $U \sim \mathcal{U}_{[0,1]}$. Then $X = F^{-}(U)$ has cdf F.*

Proof. It is easy to see (e.g. Figure 1) that $F^{-}(u) \leq x$ is equivalent to $u \leq F(x)$. Thus for $U \sim \mathcal{U}_{[0,1]}$, we have

$$\mathbb{P}\left(F^{-}(U) \leq x\right) = \mathbb{P}(U \leq F(x)) = F(x);$$

i.e. F is the cdf of $F^{-}(U)$.

Example 1 ***(Exponential distribution).*** *If $F(x) = 1 - e^{-\lambda x}$, then $F^{-}(u) = F^{-1}(u) = -\log(1-u)/\lambda$. Hence $-\log(1-U)/\lambda$ and $-\log(U)/\lambda$ where $U \sim \mathcal{U}_{[0,1]}$ are distributed according to an exponential distribution $\mathcal{E}xp(\lambda)$.*

Example 2 ***(Cauchy distribution)***. *The Cauchy distribution has density* $\pi(x)$ *and cdf* $F(x)$ *given by*

$$\pi(x) = \frac{1}{\pi(1+x^2)}, \; F(x) = \frac{1}{2} + \frac{arc\tan x}{\pi}$$

Hence we have $F^-(u) = F^{-1}(u) = \tan\left(\pi\left(u - \frac{1}{2}\right)\right)$.

Example 3 ***(Discrete distribution)***. *Assume* X *takes the values* $x_1 < x_2 < \cdots$ *with probability* $p_1, p_2, \ldots$. *In this case, both* F *and* F^- *are step functions*

$$F(x) = \sum_{x_k \le x} p_k$$

and

$$F^-(u) = x_k \text{ for } p_1 + \cdots + p_{k-1} < u \le p_1 + \cdots + p_k.$$

For example, if $0 < p < 1$ *and* $q = 1 - p$ *and we want to simulate* $X \sim \mathcal{G}eo(p)$ *then*

$$\pi(x) = pq^{x-1}, \; F(x) = 1 - q^x \qquad x = 1, 2, 3...$$

The smallest $x \in \mathbb{N}$ *giving* $F(x) \ge u$ *is the smallest* $x \ge 1$ *satisfying* $x \ge \log(1-u)/\log(q)$ *and this is given by*

$$x = F^-(u) = \left\lceil \frac{\log(1-u)}{\log(q)} \right\rceil$$

where $\lceil x \rceil$ *rounds up and we could replace* $1 - u$ *with* u.

This algorithm can also be used to generate random variables with values in any countable set.

2 Transformation Methods

Suppose we have a $\mathbb{Y}$-valued random variable (rv) $Y \sim q$ which we can simulate (eg, by inversion) and some other $\mathbb{X}$-valued rv $X \sim \pi$ which we wish to simulate. It may be that we can find a function $\varphi : \mathbb{Y} \to \mathbb{X}$ with the property that if we simulate $Y \sim q$ and then set $X = \varphi(Y)$ then we get $X \sim \pi$. Inversion is a special case of this idea.

We may generalize this idea to take functions of collections of rv with different distributions.

Example 4 ***(Gamma distribution)***. *Let* Y_i, $i = 1, 2, ..., \alpha$, *be iid rv with* $Y_i \sim \mathcal{E}xp(1)$ *(we can simulate these as above) and* $X = \beta^{-1} \sum_{i=1}^{\alpha} Y_i$ *then* $X \sim \mathcal{G}a(\alpha, \beta)$. *Indeed the moment generating function of* X *is*

$$\mathbb{E}\left(e^{tX}\right) = \prod_{i=1}^{\alpha} \mathbb{E}\left(e^{\beta^{-1} t Y_i}\right) = (1 - t/\beta)^{-\alpha}$$

which is the moment generating function of the gamma density $\pi(x) \propto x^{\alpha-1}\exp(-\beta x)$ *of parameters* α, β.

For continuous random variables, a useful tool is the transformation/change of variables formula for probability density function.

Example 5 ***(Beta distribution)***. *Let* $X_1 \sim \mathcal{G}a(\alpha, 1)$ *and* $X_2 \sim \mathcal{G}a(\beta, 1)$ *then*

$$\frac{X_1}{X_1 + X_2} \sim \mathcal{B}eta(\alpha, \beta)$$

where $\mathcal{B}eta(\alpha, \beta)$ *is the Beta distribution of parameter* α, β *of density* $\pi(x) \propto x^{\alpha-1}(1-x)^{\beta-1}$.

Example 6 ***(Gaussian distribution, Box-Muller Algorithm)***. *Let* $U_1 \sim \mathcal{U}_{[0,1]}$ *and* $U_2 \sim \mathcal{U}_{[0,1]}$ *be independent and set*

$$R = \sqrt{-2\log(U_1)},$$
$$\vartheta = 2\pi U_2.$$

We have

$$X = R\cos\vartheta \sim \mathcal{N}(0,1),$$
$$Y = R\sin\vartheta \sim \mathcal{N}(0,1).$$

Indeed $R^2 \sim \mathcal{E}xp\left(\frac{1}{2}\right)$ *and* $\vartheta \sim \mathcal{U}_{[0,2\pi]}$ *and their joint density is* $q\left(r^2,\theta\right) = \frac{1}{2}\exp\left(-r^2/2\right)\frac{1}{2\pi}$. *By the change of variables formula,*

$$\pi(x,y) = q\left(r^2,\theta\right)\left|\det\frac{\partial\left(r^2,\theta\right)}{\partial(x,y)}\right|$$

where

$$\left|\det\frac{\partial\left(r^2,\theta\right)}{\partial(x,y)}\right|^{-1} = \left|\det\begin{pmatrix}\frac{\partial x}{\partial r^2} & \frac{\partial x}{\partial\theta}\\ \frac{\partial y}{\partial r^2} & \frac{\partial y}{\partial\theta}\end{pmatrix}\right| = \left|\det\begin{pmatrix}\frac{\cos\theta}{2r} & -r\sin\theta\\ \frac{\sin\theta}{2r} & r\cos\theta\end{pmatrix}\right| = \frac{1}{2}.$$

that is

$$\pi(x,y) = \frac{1}{2\pi}\exp\left(-\frac{x^2+y^2}{2}\right).$$

Example 7 ***(Multivariate Gaussian distribution)***. *Let* $Z = (Z_1,...,Z_d)$ *be a collection of* d *independent standard normal rv. Let* L *be a real invertible* $d\times d$ *matrix satisfying* $L\ L^T = \Sigma$, *and* $X = LZ + \mu$. *Then* $X \sim \mathcal{N}(\mu,\Sigma)$. *We have indeed* $q(z) = (2\pi)^{-d/2}\exp\left(-\frac{1}{2}z^Tz\right)$ *and*

$$\pi(x) = q(z)\left|\det\partial z/\partial x\right|$$

where $\partial z/\partial x = L^{-1}$ *and* $\det(L) = \det\left(L^T\right)$ *so* $\det\left(L^2\right) = \det(\Sigma)$, *and* $\det\left(L^{-1}\right) = 1/\det(L)$ *so* $\det\left(L^{-1}\right) = \det(\Sigma)^{-1/2}$ *and*

$$z^Tz = (x-\mu)^T\left(L^{-1}\right)^T L^{-1}(x-\mu)$$
$$= (x-\mu)^T\Sigma^{-1}(x-\mu).$$

Practically we typically use a Cholesky factorization $\Sigma = L\ L^T$ *where* L *is a lower triangular matrix.*

Example 8 ***(Poisson distribution)***. *Let* (X_i) *be i.i.d.* $\mathcal{E}xp(1)$ *and* $S_n = \sum_{i=1}^n X_i$ *with* $S_0 = 0$. *Then* $S_n \sim \mathcal{G}a(n,1)$ *and*

$$\begin{aligned}\mathbb{P}(S_n \le t \le S_{n+1}) &= \mathbb{P}(S_n \le t) - \mathbb{P}(S_{n+1}\le t)\\ &= \int_0^t e^{-x}\left(\frac{x^{n-1}}{(n-1)!} - \frac{x^n}{n!}\right)dx\\ &= e^{-t}\frac{t^n}{n!}.\end{aligned}$$

If X_i *correspond to the interarrival time between customers in a queueing system, then* S_n *is the arrival time of the* n*-customer and* $S_n \le t < S_{n+1}$ *means that the number of customers that have arrived up to time* t *is equal to* n. *This number has a Poisson distribution with parameter* t *so*

$$X = \min\{n : S_n > t\} - 1 \sim \mathcal{P}oi(t).$$

Practically this can be simulated using

$$\begin{aligned}X &= \min\left\{n : -\sum_{i=1}^n \log U_i > t\right\} - 1\\ &= \min\left\{n : \prod_{i=1}^n U_i > e^{-t}\right\} - 1.\end{aligned}$$

3 Sampling via Composition

Assume we have a joint pdf $\overline{\pi}$ with marginal π; i.e.

$$\pi(x) = \int \overline{\pi}_{X,Y}(x,y)\,dy \tag{1}$$

where $\overline{\pi}(x,y)$ can always be decomposed as

$$\overline{\pi}_{X,Y}(x,y) = \overline{\pi}_Y(y)\,\overline{\pi}_{X|Y}(x|\,y).$$

It might be easy to sample from $\overline{\pi}(x,y)$ whereas it is difficult/impossible to compute $\pi(x)$. In this case, it is sufficient to sample

$$Y \sim \overline{\pi}_Y \text{ then } X|\,Y \sim \overline{\pi}_{X|Y}(\cdot|\,Y)$$

so $(X,Y) \sim \overline{\pi}_{X,Y}$ and hence $X \sim \pi$ as (1) holds.

Example 9 ***(Scale mixture of Gaussians)***. *A very useful application of the composition method is for scale mixture of Gaussians; i.e.*

$$\pi(x) = \int \underbrace{\mathcal{N}(x;0,1/y)}_{\overline{\pi}_{X|Y}(x|y)}\overline{\pi}_Y(y)\,dy.$$

For various choices of the mixing distributions $\overline{\pi}_Y(y)$, we obtain distributions $\pi(x)$ which are t-student, $\alpha-$stable, Laplace, logistic.

Example 10 ***(Finite mixture of distributions)*** *Assume one wants to sample from*

$$\pi(x) = \sum_{i=1}^{p} \alpha_i.\pi_i(x)$$

where $\alpha_i > 0$, $\sum_{i=1}^{p} \alpha_i = 1$ and $\pi_i(x) \geq 0$, $\int \pi_i(x)\,dx = 1$. We can introduce $Y \in \{1,...,p\}$ and introduce

$$\overline{\pi}_{X,Y}(x,y) = \alpha_y \times \pi_y(x).$$

To sample from $\pi(x)$, then sample Y from a discrete distribution such that $\mathbb{P}(Y=k) = \alpha_k$ then

$$X|\,(Y=y) \sim \pi_y.$$

4 Rejection Sampling

The basic idea of rejection sampling is to sample from a proposal distribution q different from the target π and then to correct through a rejection step to obtain a sample from π. The method proceeds as follows.

Algorithm (Rejection Sampling). Given two densities π, q with $\pi(x) \leq M.q(x)$ for all x, we can generate a sample from π by

1. Draw $X \sim q$
2. Accept $X = x$ as a sample from π with probability

$$\frac{\pi(x)}{M.q(x)},$$

   otherwise go to step 1.

We establish here the validity of the rejection sampling algorithm.

Proposition 2 ***(Rejection sampling)***. *The distribution of the samples accepted by rejection sampling is π.*

Proof. We have for any (measurable) set A

$$\mathbb{P}\left(X \in A \mid X \text{ accepted}\right) = \frac{\mathbb{P}\left(X \in A, X \text{ accepted}\right)}{\mathbb{P}\left(X \text{ accepted}\right)}$$

where

$$\begin{aligned}
\mathbb{P}\left(X \in A, X \text{ accepted}\right) &= \int_{\mathbb{X}} \int_0^1 \mathbb{I}_A\left(x\right) \mathbb{I}\left(u \leq \frac{\pi\left(x\right)}{M.q\left(x\right)}\right) q\left(x\right) du dx \\
&= \int_{\mathbb{X}} \mathbb{I}_A\left(x\right) \frac{\pi\left(x\right)}{M.q\left(x\right)} q\left(x\right) dx \\
&= \int_{\mathbb{X}} \mathbb{I}_A\left(x\right) \frac{\pi\left(x\right)}{M} dx = \frac{\pi\left(A\right)}{M} \\
\mathbb{P}\left(X \text{ accepted}\right) = \mathbb{P}\left(X \in \mathbb{X}, X \text{ accepted}\right) &= \frac{\pi\left(\mathbb{X}\right)}{M} = \frac{1}{M}
\end{aligned}$$

so

$$\mathbb{P}\left(X \in A \mid X \text{ accepted}\right) = \pi\left(A\right).$$

Thus the distribution of the accepted values is π.

Important remark: In most practical scenarios, we only know π and q up to some normalising constants; i.e.

$$\pi = \widetilde{\pi}/Z_\pi \text{ and } q = \tilde{q}/Z_q$$

where $\widetilde{\pi}, \tilde{q}$ are known but $Z_\pi = \int_{\mathbb{X}} \widetilde{\pi}\left(x\right) dx$, $Z_q = \int_{\mathbb{X}} \tilde{q}\left(x\right) dx$ are unknown. We can still use rejection in this scenario as

$$\frac{\pi\left(x\right)}{q\left(x\right)} \leq M \Leftrightarrow \frac{\widetilde{\pi}\left(x\right)}{\widetilde{q}\left(x\right)} \leq M \frac{Z_\pi}{Z_q}.$$

Practically, this means we can throw the normalising constants out at the start: if we can find M' to bound $\widetilde{\pi}\left(x\right)/\widetilde{q}\left(x\right)$ then it is correct to accept with probability $\widetilde{\pi}\left(x\right)/\left(M'\widetilde{q}\left(x\right)\right)$ in the rejection algorithm.

Lemma 1 *Let T denote the number of pairs (X, U) that have to be generated until $U \leq \pi\left(X\right)/\left(Mq\left(X\right)\right)$ for the first time. Then T is geometrically distributed with parameter $1/M$ and in particular $\mathbb{E}\left(T\right) = M$.*

Example 11 ***(Uniform density on a bounded subset of $\mathbb{R}^p$)****. Consider $B \subset \mathbb{R}^p$ be a bounded subset of $\mathbb{R}^p$. We are interested in sampling from the uniform distribution on B*

$$\pi\left(x\right) \propto \mathbb{I}_B\left(x\right).$$

Assume we can find a rectangle R with $B \subset R$ then we can use for q the uniform distribution on R. Then using $\widetilde{\pi}\left(x\right) = \mathbb{I}_B\left(x\right)$, $\widetilde{q}\left(x\right) = \mathbb{I}_R\left(x\right)$, we can simply use $M' = 1$ and $\widetilde{\pi}\left(x\right)/\left(M'\widetilde{q}\left(x\right)\right) = \mathbb{I}_B\left(x\right)$.

Example 12 ***(Beta density)****. We have for $\alpha, \beta > 0$*

$$\widetilde{\pi}\left(x\right) = x^{\alpha-1}\left(1-x\right)^{\beta-1}, \ 0 < x < 1.$$

For $\alpha, \beta \geq 1$, this is upper bounded on $[0, 1]$ so we can use $q\left(x\right) = \widetilde{q}\left(x\right) = \mathbb{I}_{(0,1)}\left(x\right)$ and

$$M = \sup_x \frac{\widetilde{\pi}\left(x\right)}{\widetilde{q}\left(x\right)} = \frac{\left(\alpha-1\right)^{\alpha-1}\left(\beta-1\right)^{\beta-1}}{\left(\alpha+\beta-2\right)^{\alpha+\beta-2}}.$$

For $\alpha < 1, \beta \geq 1$ we can use $q\left(x\right) = \widetilde{q}\left(x\right) = \alpha x^{\alpha-1}\mathbb{I}_{(0,1)}\left(x\right)$ thus

$$M = \sup_x \frac{\widetilde{\pi}\left(x\right)}{\widetilde{q}\left(x\right)} = \sup_x \frac{\left(1-x\right)^{\beta-1}}{\alpha} = \frac{1}{\alpha}.$$

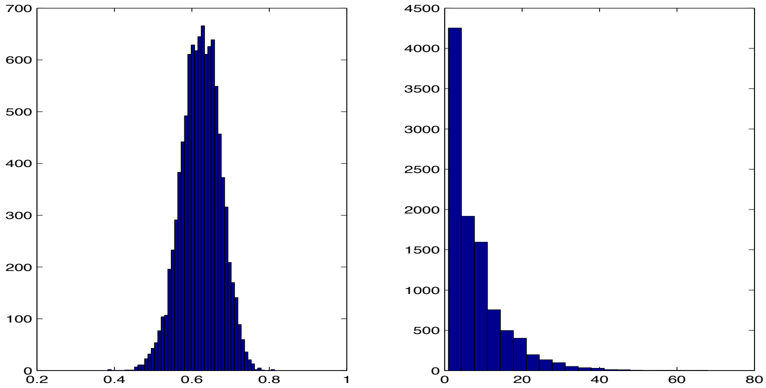

Figure 2: Histogram approximation of $p(\theta|\, y_1, ..., y_4)$ (left) and histogram approximation of waiting time distribution before acceptance (mean 7.8) (right)

Example 13 ***(Normal distribution)****. Let $\widetilde{\pi}(x) = \exp\left(-\frac{1}{2}x^2\right)$ and $\widetilde{q}(x) = 1/\left(1+x^2\right)$. We have*

$$\frac{\widetilde{\pi}(x)}{\widetilde{q}(x)} = \left(1+x^2\right)\exp\left(-\frac{1}{2}x^2\right) \leq 2/\sqrt{e} = M$$

which is attained at ± 1. Hence the probability of acceptance is

$$\mathbb{P}\left(U \leq \frac{\widetilde{\pi}(x)}{M\widetilde{q}(x)}\right) = \frac{Z_\pi}{MZ_q} = \frac{\sqrt{2\pi}}{\frac{2}{\sqrt{e}}\pi} = \sqrt{\frac{e}{2\pi}} \approx 0.66$$

and the mean number of trials to success is approximately $1/0.66 \approx 1.52$.

Example 14 ***(Genetic Linkage Model).*** *We observe*

$$(Y_1, Y_2, Y_3, Y_4) \sim \mathcal{M}\left(n; \frac{1}{2}+\frac{\theta}{4}, \frac{1}{4}(1-\theta), \frac{1}{4}(1-\theta), \frac{\theta}{4}\right)$$

where $\mathcal{M}$ is the multinomial distribution and $\theta \in (0,1)$. The likelihood of the observations is thus

$$p(y_1, ..., y_4|\, \theta) \propto (2+\theta)^{y_1}(1-\theta)^{y_2+y_3}\theta^{y_4}.$$

We follow here a Bayesian approach where we select a prior $p(\theta) = \mathbb{I}_{[0,1]}(\theta)$. Hence the resulting posterior is

$$p(\theta|\, y_1, ..., y_4) \propto (2+\theta)^{y_1}(1-\theta)^{y_2+y_3}\theta^{y_4}.$$

We propose to use rejection sampling using a proposal $q(\theta) = \widetilde{q}(\theta) = p(\theta)$ to sample from $p(\theta|\, y_1, ..., y_4)$. To use accept-reject, we need to upper bound

$$\widetilde{\pi}(\theta) = (2+\theta)^{y_1}(1-\theta)^{y_2+y_3}\theta^{y_4}$$

Using a simple optimization algorithm, we get

$$g(\theta) \leq g(\theta_{\max})$$

where $\theta_{\max} = 0.6268$ and $g(\theta_{\max}) \approx \exp(67.4)$. Hence we can use rejection sampling; see Figure 2.