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 have seen in the previous notes that those samples can be used to efficiently approximate integrals with respect to π. We discuss here the standard techniques used in most software packages.

1 Pseudo-random numbers

To start with, consider the problem of obtaining realisations of a random variable U distributed uniformly among the integers $\{0, \ldots, M-1\}$, for some $M \in \mathbb{N}$ (typically $M = 2^{32}$, yielding 32-bits integers). A computer is an entirely deterministic machine, and thus cannot by itself generate random numbers in the probabilistic sense. However it can generate deterministic sequences that behave just like random sequences, in a statistical sense. Those numbers are called pseudo-random numbers, because they try to mimick randomness. The most simple example of a pseudo-random number generation (pseudo-RNG) scheme is the linear congruential generator. Starting from a "seed" or starting value X_0, the next element of the sequence is obtained at step $n \geq 1$ as:

$$X_n = (aX_{n-1} + c) \mod M$$

for some integers a, c, M. For instance, some pseudo-RNG schemes implemented in standard C++ libraries use $M = 2^{32}$, $a = 1103515245$, $c = 12345$. According to the Hull-Dobell theorem these values guarantee that the sequence is of period m for every starting value X_0, thus it takes successively each value in $\{0, \ldots, M-1\}$. If we define

$$U_n = X_n/M$$

then we obtain values between 0 and 1 that behave like realisations of a uniform random variable $U \sim \mathcal{U}_{[0,1]}$. The first $10,000$ values of U_n obtained with the above pseudo-RNG scheme are plotted on Figure 1. We can see on the figure that, although the sequence is entirely deterministic, it seems to fill the interval $[0,1]$ in the manner of random draws from a uniform distribution.

Many RNG schemes exist, and improve upon linear congruential operators by generating sequences that look more "random" in some sense. The quality of a pseudo-RNG generator can be tested in several ways, and in particular the "die hard" series of tests is standard in the assessment of this quality. The default pseudo-RNG scheme used in R (as well as in `Python`) is the "Mersenne-Twister", but other schemes can be used. The R documentation on this topic can be accessed by typing `?.Random.seed`. In the following, we assume that we know how to sample exactly from the uniform distribution $\mathcal{U}_{[0,1]}$, keeping in mind that in practice, we will rely on a pseudo-RNG scheme.

2 Inversion Method

The inversion method allows to sample from a distribution π using draws from a uniform distribution $\mathcal{U}_{[0,1]}$, and the quantile function associated with π. Consider a real-valued random variable X and its associated cumulative distribution function (cdf)

$$\forall x \in \mathbb{R} \quad F(x) = \mathbb{P}(X \leq x).$$

The cdf $F : \mathbb{R} \to [0,1]$ is
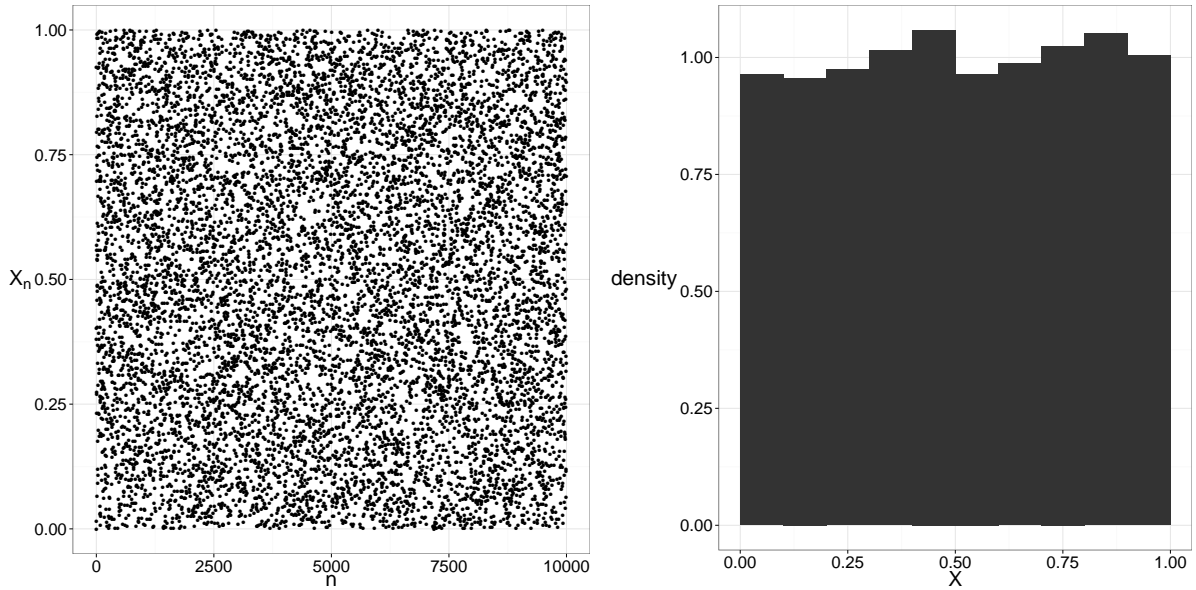
- increasing; i.e. if $x \leq y$ then $F(x) \leq F(y)$,

Figure 1: Pseudo-random sequence of $10,000$ numbers X_n imitating uniform random variables in $[0, 1]$. Left: X_n plotted against n; right: histogram of X_n.

- 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" or "inverse cdf". Its definition is illustrated by Figure 1. Note that $F^-(u) = F^{-1}(u)$ if F is continuous.
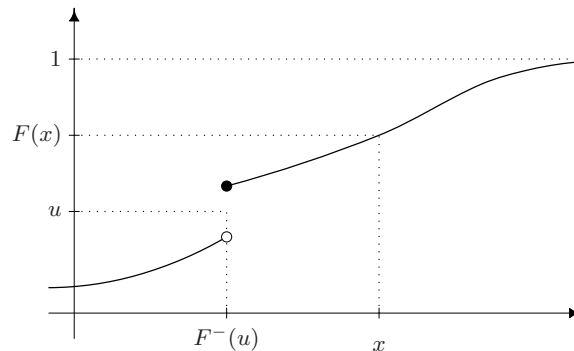

Figure 2: Illustration of the definition of the generalised inverse F^-.

Proposition 2.1. *(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

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

i.e. F is the cdf of $F^-(U)$. $\qquad\qquad\square$

In other words, if $U \sim \mathcal{U}_{[0,1]}$ and F^- is the quantile function associated with a distribution π, then $F^-(U)$ is distributed according to π.

Example 2.1. (Exponential distribution). *If $F(x) = 1 - e^{-\lambda x}$ for all $x \in \mathbb{R}^+$, 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 2.2. (Cauchy distribution). *The Cauchy distribution has density $\pi(x)$ and cdf $F(x)$ for all $x \in \mathbb{R}$ given by*

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

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

Example 2.3. (Discrete distribution). *Assume X takes the values $x_1 < x_2 < \cdots$ with probability $p_1, p_2, ...$. In this case, both F and F^- are step functions:*

$$F(x) = \sum_{k: x_k \leq x} p_k$$

and

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

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

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

The smallest $x \in \mathbb{N}$ giving $F(x) \geq u$ is the smallest $x \geq 1$ satisfying $x \geq \log(1-u)/\log(q)$ 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.

The inversion method can also be used to generate random variables with values in any countable set. The limitation of this method is that one needs to be able to evaluate the generalised inverse of the cumulative distribution function, F^-, which is typically unknown for the distributions encountered in real applications.

3 Transformation Methods

Suppose we have a $\mathbb{Y}$-valued random variable $Y \sim q$ which we can simulate (e.g. by the inversion method) and some other $\mathbb{X}$-valued random variable $X \sim \pi$ 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, in fact, a special case of this idea where Y is uniformly distributed and φ is the generalised inverse of the cumulative distribution function.

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

Example 3.1. (Gamma distribution). *Let Y_i, $i = 1, 2, ..., \alpha$, for $\alpha \in \mathbb{N}$, be i.i.d. random variable with $Y_i \sim \mathcal{E}xp(1)$ (e.g. simulated by inversion as above) and $X = \beta^{-1} \sum_{i=1}^{\alpha} Y_i$ with $\beta \in \mathbb{R}^+$ then $X \sim \mathcal{G}a(\alpha, \beta)$. Indeed the moment generating function of X is*

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

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

Example 3.2. (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 $\pi(x) \propto x^{\alpha-1}(1-x)^{\beta-1}$.

For continuous random variables, another useful tool is the transformation/change of variables formula for probability density function, as illustrated in the following example.

Example 3.3. (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

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

The Box-Muller can be used to generate vectors of independent Gaussian random variables. To generate multivariate (correlated) Gaussian random variables, we can use the following reasoning.

Example 3.4. (Multivariate Gaussian distribution). *Let $Z = (Z_1,...,Z_d)$ be a collection of d independent standard normal random variables. 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 $q(z) = (2\pi)^{-d/2}\exp\left(-\frac{1}{2}z^T z\right)$ 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 $\det(L)^2 = \det(\Sigma)$, and $\det\left(L^{-1}\right) = 1/\det(L)$ so $\det\left(L^{-1}\right) = \det(\Sigma)^{-1/2}$ and

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

In practice we typically use a Cholesky factorization $\Sigma = L\,L^T$ where L is a lower triangular matrix.

Example 3.5. (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*

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

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 $S_n \leq t < S_{n+1}$ 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

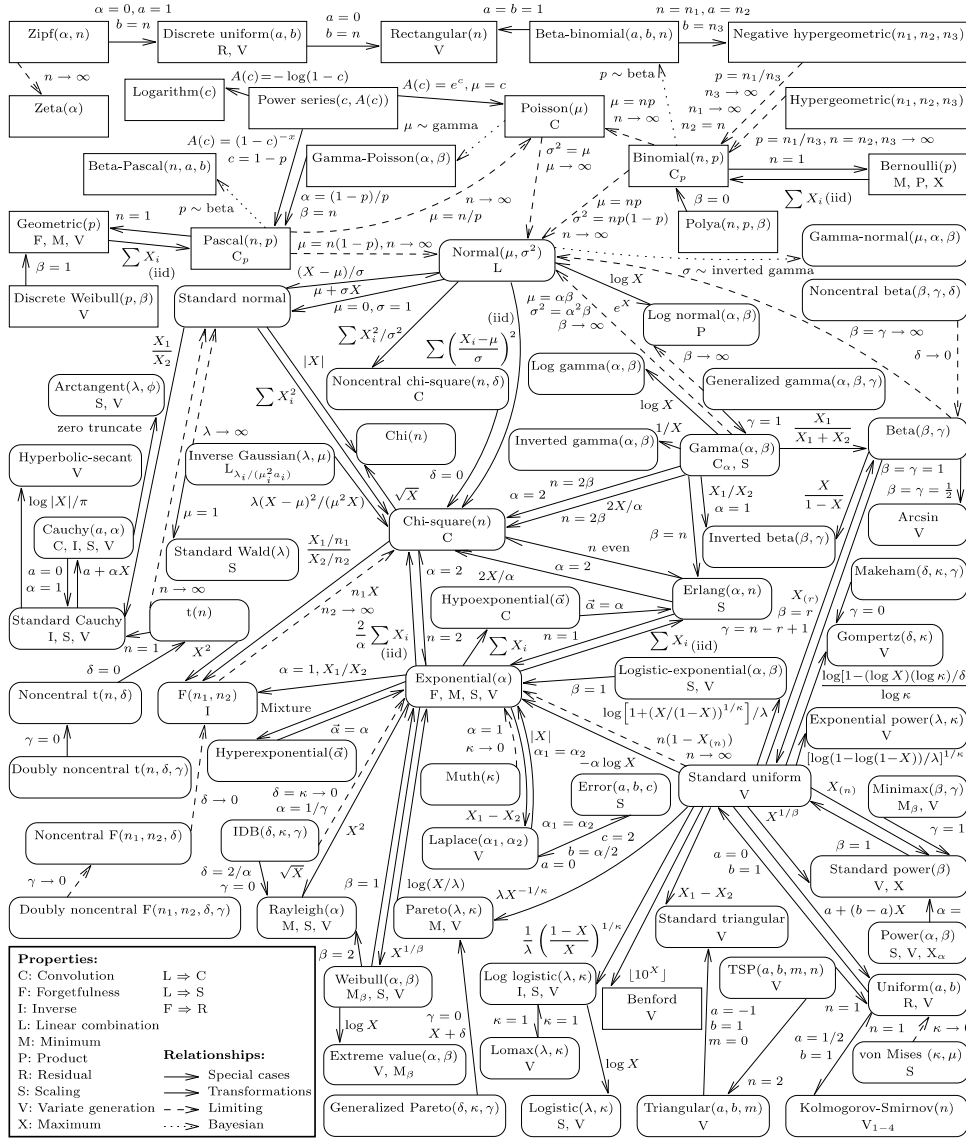$$X = \min\{n : S_n > t\} - 1 \sim \mathcal{P}oi(t).$$

Figure 3: Relationships between univariate probability distributions, taken from Lawrence M. Leemis and Jacquelyn T. McQueston, "univariate distribution relationships", Teacher's corner, 2008.

In practice a Poisson random variable can be simulated using uniform random variables (U_i), by defining

$$X = \min\left\{n : -\sum_{i=1}^{n} \log U_i > t\right\} - 1$$

$$= \min\left\{n : \prod_{i=1}^{n} U_i > e^{-t}\right\} - 1.$$

Some of the known transformations linking standard probability distributions are shown in Figure 3. They prove useful to generate realisations from a standard distribution given realisations of another; but ultimately many of the distributions of interest in Bayesian statistics will be outside any well-known family of distributions.

4 Sampling via composition

Assume we have a joint probability density function $\bar{\pi}$ with marginal distribution π; i.e.

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

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

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

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

$$Y \sim \bar{\pi}_Y \ \text{ then } \ X|Y \sim \bar{\pi}_{X|Y}\left(\cdot|Y\right)$$

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

Example 4.1. *(Scale mixture of Gaussians). A very useful application of the composition method is for scale mixture of Gaussians; i.e.*

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

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

Example 4.2. *(Finite mixture of distributions) Assume one wants to sample from the mixture distribution with density*

$$\forall x \in \mathbb{R} \quad \pi\left(x\right) = \sum_{i=1}^{p} \alpha_i \pi_i\left(x\right)$$

where $\alpha_i > 0$, $\sum_{i=1}^{p} \alpha_i = 1$ and $\pi_i\left(x\right) \geq 0$, $\int \pi_i\left(x\right) dx = 1$. Assume that we can sample from each π_i, $i \in \{1,\ldots,p\}$. We can introduce a random variable $Y \in \{1,...,p\}$ and the joint distribution

$$\forall x \in \mathbb{R} \quad \forall y \in \{1,\ldots,p\} \quad \bar{\pi}_{X,Y}\left(x,y\right) = \alpha_y \times \pi_y\left(x\right).$$

To sample from π, first sample Y from a discrete distribution with probabilities $\mathbb{P}\left(Y = k\right) = \alpha_k$, then sample from π_Y:

$$X|Y = y \sim \pi_y.$$

Then X follows π.

The methods described thus far rely on some specific knowledge available on the target distribution π of interest. Inversion requires point-wise evaluations of the quantile function, while transformations and compositions require a representation of π in terms of other, more simple distributions that can be readily sampled. Rejection sampling, as described in the next section, only requires the ability to compute point-wise evaluations of the probability density function of π, and to have access to an upper bound of π.

5 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 on $\mathbb{X}$ with $\pi\left(x\right) \leq Mq\left(x\right)$ for all $x \in \mathbb{X}$ and some $M < \infty$, we can generate a sample from π as follows:

1. Draw $X \sim q$.

2. Accept $X = x$ as a sample from π with probability

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

otherwise go back to step 1.

Note that to implement rejection sampling, we need to implement a mechanism to either "do this" or "do that" according to some probability $\alpha \in [0,1]$. The standard way to implement this goes as follows.

1. Draw $U \sim \mathcal{U}_{[0,1]}$.

2. If $U \leq \alpha$ then "do this", otherwise "do that".

Indeed $\mathbb{P}(U \leq \alpha) = \alpha$ when $U \sim \mathcal{U}_{[0,1]}$, so that the above scheme will "do this" with probability α.

Proposition 5.1. (Rejection sampling). *The distribution of the samples accepted by rejection sampling is π.*

Proof. We have for any (measurable) set $A \subset \mathbb{X}$,

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

where

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

Hence

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

Thus the distribution of the accepted values is precisely π. $\square$

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

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

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

Note that if the algorithm works for $M \geq \sup_x \pi(x)/q(x)$, then it is also valid for any $M' \geq M$. Does it matter which M we choose? The answer (yes) is given by the following lemma that indicates how much time we are expected to wait until the algorithm returns a sample from π, as a function of M.

Lemma 5.1. *Let T denote the waiting time, that is, the number of pairs (X, U) that have to be generated until $U \leq \pi(X)/(Mq(X))$ for the first time, where $M \geq \sup_x \pi(x)/q(x)$. Then T is geometrically distributed with parameter $1/M$ and in particular $\mathbb{E}(T) = M$.*

According to the lemma, the smaller M, the faster it takes to generate samples from π. Denoting

$$M^\star = \sup_{x \in \mathbb{X}} \frac{\pi(x)}{q(x)}$$

then the smallest choice of M for a given proposal distribution q is $M^\star$. In practice we first have to choose a proposal distribution q. In order for $M^\star$ to be as small as possible, we see that q should be as close as possible to π; in the limiting case, if $q = \pi$ then $M^\star = 1$ and of course each draw from q is accepted. On the other hand, if we do not know how to bound well $\sup_x \pi(x)/q(x)$, then picking M to be very large (in order "to be safe") would result in a very slow algorithm to generate samples from π.

Example 5.1. (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*

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

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

Note that in the above example, if we use the bound $\tilde{M}$ on the unnormalized probability density functions, instead of the bound M, we do not know the average waiting time (which is M and not $\tilde{M}$). To get M we would need to know the volume of the set B and the volume of the rectangle R.

Example 5.2. (Beta density). *We have for $\alpha, \beta > 0$*

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

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

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

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

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

Example 5.3. (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*

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

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

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

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

Example 5.4. (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

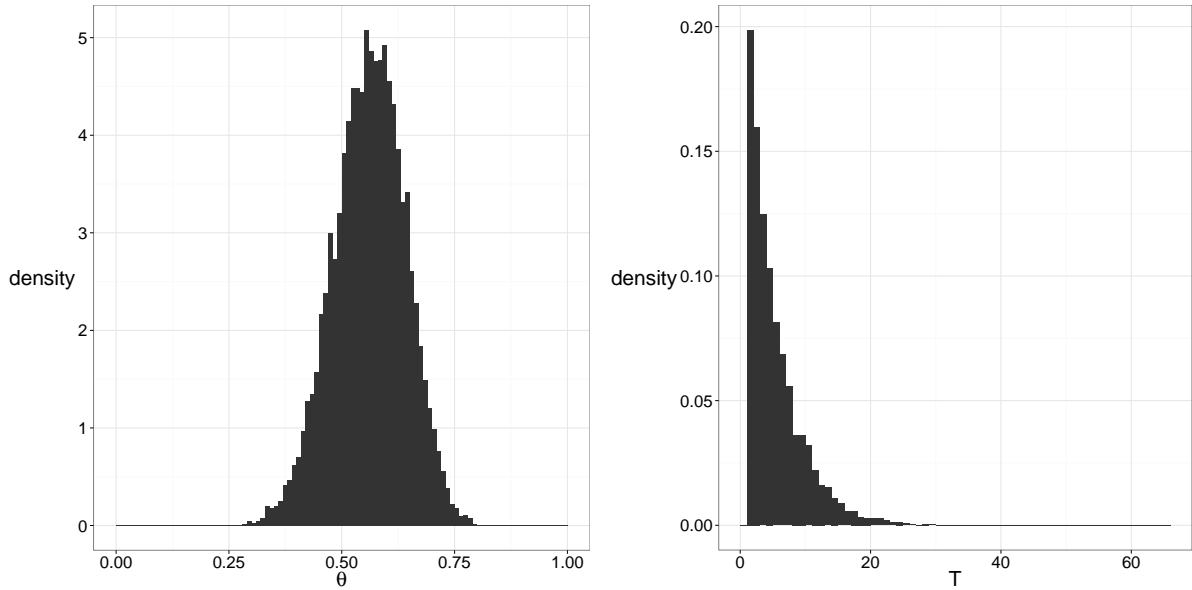$$p(\theta | y_1, ..., y_4) \propto (2+\theta)^{y_1}(1-\theta)^{y_2+y_3}\theta^{y_4}.$$

Figure 4: Histogram of the $10,000$ samples from $p\left(\theta\mid y_1,...,y_4\right)$ obtained by rejection sampling (left); and histogram of waiting time distribution before acceptance (right).

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

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

Using a simple optimization algorithm, we can find the maximum of $\tilde{\pi}(\theta)$ and we use $\tilde{M} = \sup_{\theta\in[0,1]}\tilde{\pi}(\theta)$ to perform rejection sampling. For a realisation of (Y_1, Y_2, Y_3, Y_4) equal to $(69, 9, 11, 11)$ obtained with $n = 100$ and $\theta^\star = 0.6$, a sample of $10,000$ realisations from π is shown on Figure 4, along with the distribution of the associated waiting times (T_i), that is, the distribution of the number of samples generated before one is accepted.

What if we are not capable of coming up with a distribution q and a (reasonably small) value M such that $\sup_x \pi(x)/q(x) \leq M$ or $\tilde{M}$ such that $\sup_x \tilde{\pi}(x)/\tilde{q}(x) \leq \tilde{M}$? In the next lectures we will see techniques to approximate integrals with respect to π, that do not necessarily require as much knowledge on π as the ones described above.