```
n = 1000;  q = 3/4
u = rand(n)
x = ceil.(log.(u)/log(q))
```

In fact, both the continuous and the discrete case follow as special cases from the following general inverse c.d.f. result.

Theorem 2.5 (*)**.** *Assume* $U \sim \mathcal{U}(0,1)$ *and let* $F : \mathbb{R} \to [0,1]$ *be a c.d.f.*[4]*. Define*

$$X := F^{-1}(U) \qquad \text{where}$$
$$F^{-1}(u) := \min\{x \in \mathbb{R} \ : \ F(x) \geq u\} \quad \text{for } 0 < u < 1.^{6}$$

Then, $X \sim F$*, that is,* X *has c.d.f.* F*.*

Proof. Recall that a c.d.f. F is increasing and right-continuous (which implies that the the min above is well-defined). The proof follows as in the proof of Theorem 2.1, by noticing that

$$F^{-1}(u) \leq x \iff u \leq F(x) \qquad \text{for all } x \in \mathbb{R} \text{ and } u \in (0,1).$$

Namely, suppose $F^{-1}(u) \leq x$ and denote $x_u := F^{-1}(u) \leq x$, then $F(x) \geq F(x_u) \geq u$. Conversely, if $u \leq F(x)$, then $F^{-1}(u) = \min\{y \in \mathbb{R} : F(y) \geq u\} \leq x$, because x is included in the set which is minimised. □

Example 2.6 (*). Consider the following c.d.f.:

$$F(x) := \left(\frac{1}{2} + \frac{1}{2}\big(1 - \exp(-x)\big)\right)\mathbf{1}\,(x \geq 0)\,.$$

Its generalised inverse is

$$F^{-1}(u) = -\log\big(2(1-u)\big)\mathbf{1}\,(u > 1/2)\,.$$

We may replace U with $1 - U$ again, resulting in the following:

```
u = rand(1000)
x = -log.(2u) .* (u .<= 1/2)
```

2.2 Distribution of transformed random variables

The inverse c.d.f. method provides a general result to transform $\mathcal{U}(0,1)$ random variables into scalar random variables, provided that the (inverse) c.d.f. is accessible. In a multivariate setting, or when c.d.f. is inaccessible, other transformations can be useful.

4. Recall that F is a c.d.f. if it is increasing[5], right-continuous, $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$.

6. The function F^{-1} is called the *generalised inverse c.d.f.*.

Suppose that $X \sim p_X$, the function $g : \mathbb{R} \to \mathbb{R}$ is strictly increasing and continuously differentiable. Let $Y = g(X)$, then

$$F_Y(y) := \mathbb{P}(Y \le y) = \mathbb{P}\big(g(X) \le y\big) = F_X\big(g^{-1}(y)\big),$$

where $F_X(x) := P(X \le x)$ is the c.d.f. of X. Now, the p.d.f. of Y is

$$p_Y(y) = F_Y'(y) = F_X'\big(g^{-1}(y)\big)(g^{-1})'(y) = \frac{p_X\big(g^{-1}(y)\big)}{g'(g^{-1}(y))},$$

because $(g^{-1})'(y) = 1/g'\big(g^{-1}(y)\big)$.

Recall the following multivariate generalisation of the above, which we use without proof.

Theorem 2.7. *Suppose $X \sim p_X$ and $S := \operatorname{supp}(p) := \{x \in \mathbb{R}^d : p_X(x) > 0\}$ is an open set. If $g : S \to \mathbb{R}^d$ is one-to-one and continuously differentiable such that its Jacobian Dg is invertible, $\det(Dg(x)) \neq 0$ for all $x \in S$, then $Y = g(X)$ has density p_Y given as follows,*

$$p_Y(y) = \begin{cases} p_X(g^{-1}(y))|\det(Dg^{-1})(y)|, & y \in g(S) \\ 0, & y \notin g(S), \end{cases}$$

where Dg^{-1} stands for the Jacobian of g^{-1}.

Remark 2.8. By the inverse function theorem, for all $y \in g(S)$,

$$(Dg^{-1})(y) = [(Dg)(x)]^{-1},$$

where $y = g(x)$ (or $x = g^{-1}(y)$). Also, $\det(A^{-1}) = 1/\det(A)$, so we have

$$|\det(Dg^{-1})(y)| = \frac{1}{|\det(Dg)(x)|}.$$

Remark 2.9 (*). If $\operatorname{supp}(p)$ can be partitioned (up to set of volume (measure) zero) into disjoint open sets $S_1, S_2, \ldots$ such that g satisfies the conditions required in Theorem 2.7, then Theorem 2.7 can be applied piecewise, leading into

$$p_Y(y) = \sum_i p_X(g^{-1}(y))|\det(Dg^{-1}(y)|\mathbf{1}\left(y \in g(S_i)\right).$$

2.3 (Multivariate) normal random variables

Normal distribution is, of course, particularly important in applications. The inverse c.d.f. method is not (directly) applicable because the c.d.f. is not available in a closed form. However, it is possible to generate normal random variables by a simple bivariate transformation.

Recall that the standard normal $N(0,1)$ p.d.f. is

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

and the general multivariate normal $N(\mu, \Sigma)$ p.d.f. is

$$p(x) = \frac{1}{(2\pi)^{d/2}\sqrt{\det(\Sigma)}} \exp\Big(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\Big).$$

The random vector $X := (X_1, \ldots, X_d)^T$, where $(X_k) \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ is distributed by $\mathcal{N}(0, I_d)$, that is, the standard multivariate Gaussian distribution with zero mean vector and identity covariance matrix.

Theorem 2.10 (Box-Muller transform)**.** *Let $U_1, U_2 \overset{i.i.d.}{\sim} \mathcal{U}(0,1)$ and define*

$$\begin{aligned} X_1 &:= R\cos(T) \\ X_2 &:= R\sin(T), \end{aligned} \qquad \textit{where} \qquad \begin{aligned} R &:= \sqrt{-2\ln U_1} \\ T &:= 2\pi U_2. \end{aligned}$$

Then, $X_1, X_2 \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$.

Proof. The density of (R,T) is (exercise)

$$p_{R,T}(r,t) = \begin{cases} \frac{1}{2\pi} r e^{-r^2/2}, & 0 < t < 2\pi, 0 < r < \infty, \\ 0, & \text{otherwise.} \end{cases}$$

Now, $(X,Y) = h(R,T)$ with $h(r,t) := (r\cos t, r\sin t)$ (polar-to-Cartesian transform), with

$$|\det(Dh)(r,t)| = \left| \det \begin{pmatrix} \cos t, & -r\sin t \\ \sin t, & r\cos t \end{pmatrix} \right| = r.$$

Now we may apply Theorem 2.7 and Remark 2.8 to deduce that

$$p_{X,Y}(x,y) = p_{R,T}(r(x,y), t(x,y)) \frac{1}{r(x,y)} = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}, \quad (x,y) \neq 0,$$

where $r(x,y) := \sqrt{x^2+y^2}$ and $t(x,y) := \operatorname{atan2}(y,x)$. □

Proposition 2.11 (Generic multivariate Gaussian distribution)**.** *Let $\mu \in \mathbb{R}^d$ and $\Sigma \in \mathbb{R}^{d\times d}$ be a positive definite matrix, and let $L \in \mathbb{R}^{d\times d}$ be the Cholesky factor of Σ (lower-triangular matrix satisfying $LL^T = \Sigma$). Then, if $Z \sim N(0, I_d)$,*

$$X := \mu + LZ \qquad \textit{satisfies} \qquad X \sim \mathcal{N}(\mu, \Sigma). \tag{3}$$

Proof. The Jacobian of $g(z) = \mu + Lz$ is $|\det(L)| = \sqrt{\det(\Sigma)} > 0$ and the inverse $g^{-1}(x) = L^{-1}(x-\mu)$.

$$\begin{aligned} p_X(x) &= p_Z(L^{-1}(x-\mu))/\sqrt{\det(\Sigma)} \\ &= \frac{1}{(2\pi)^{d/2}\sqrt{\det(\Sigma)}} \exp\Big(-\frac{1}{2}(x-\mu)^T (L^{-1})^T L^{-1}(x-\mu)\Big), \end{aligned}$$

and $(L^{-1})^T L^{-1} = (L^T)^{-1}L^{-1} = (LL^T)^{-1} = \Sigma^{-1}$. □

Remark 2.12. We could use, of course, any matrix $L \in \mathbb{R}^{d\times d}$ satisfying $LL^T = \Sigma$, but the Cholesky factor is both easy to compute and the lower-triagular structure allows for some savings when computing the transform (3).

Example 2.13. Generating bivariate Gaussians.

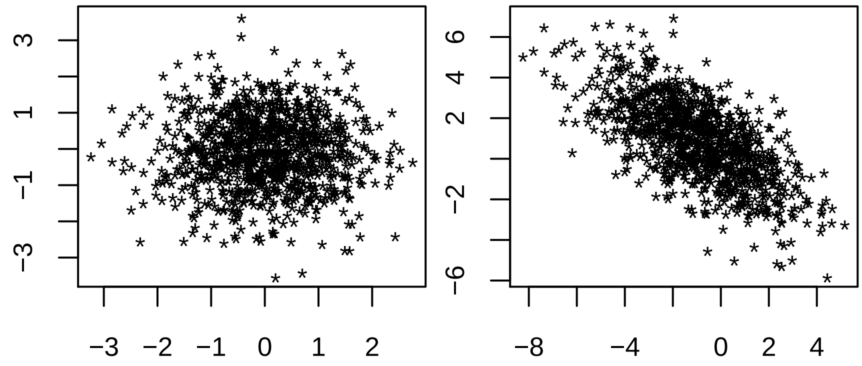

Figure 3: Standard bivariate samples $(Z_k)_{k\geq 1}$ (left) and $N(m,S)$ samples (X_k) generated in Example 2.13.

```
using LinearAlgebra
n = 1000; d = 2    # Number of samples & dimension
m = [-1,1]         # Mean vector
S = [5 -3; -3 4]   # Covariance matrix
L = cholesky(S).L # (Lower-triangular) Cholesky factor
X = zeros(d, n)    # Initialise output space
for k = 1:n
  X[:,k] = m + L*randn(d)
end
```

2.4 Relations of probability distributions (*)

Known relationships between probability distributions may yield useful transformations.

Example 2.14. [Gamma distribution with integer shape] Consider $\Gamma(\alpha,\beta)$ distribution with $\alpha \in \mathbb{N}$ and $\beta > 0$ with p.d.f.

$$p(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \mathbf{1}(x \geq 0).$$

Inverse c.d.f. method is not easily applicable. Instead,

(a) Simulate $Y_1,\ldots,Y_\alpha \overset{\text{i.i.d.}}{\sim} \mathrm{Exp}(1)$.

(b) Set $X := \frac{1}{\beta}\sum_{i=1}^\alpha Y_i$.

Then $X \sim p$.

Proof. We can check that $X \sim p$ by inspecting moment generating functions. The m.g.f. of $Y \sim \mathrm{Exp}(1)$ is

$$M_Y(t) = \mathbb{E}\left(e^{tY}\right) = \frac{1}{1-t}, \qquad t \in [0,1),$$

so the m.g.f. of X is

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

for $t \in [0, \beta)$, which is the m.g.f. of $\Gamma(\alpha, \beta)$. □

2.5 Spherically/elliptically symmetric distributions (*)

Example 2.15 (Uniform distribution on a $(d-1)$-sphere)*.* Suppose $X \sim N(0, I)$, a standard Gaussian distribution in $\mathbb{R}^d$. Then, $V = X/\|X\| \sim \mathcal{U}(S^{d-1})$, that is, V is uniformly distributed on the unit sphere $S^{d-1} := \{x \in \mathbb{R}^d : \|x\| = 1\}$, because the Gaussian distribution is spherically symmetric.

If p is a spherically symmetric distribution, then it is possible to draw independent 'direction vector' $V \sim \mathcal{U}(S^{d-1})$ and a 'radius' $R \geq 0$ so that $RV \sim p$. The density of the radius q can be found by polar integration.

Proposition 2.16. *Assume that p is a spherically symmetric probability density on $\mathbb{R}^d$, that is,*

$$p(x) = c\hat{p}(\|x\|) \qquad \text{for all} \qquad x \in \mathbb{R}^d,$$

where $c > 0$ is a constant. Suppose q is a probability density on $[0, \infty)$ satisfying

$$q(r) = c' r^{d-1} \hat{p}(r) \qquad \text{for all} \qquad r \in [0, \infty),$$

for some constant $c' > 0$. Then, if $V \sim \mathcal{U}(S^{d-1})$ and $R \sim q$, the random variable $X := RV \sim p$.

Proof. Let $A \subset [0, \infty)$, then by polar integration

$$\int_{\|x\| \in A} p(x) \mathrm{d}x = cC_d \int_{r \in A} r^{d-1} \hat{p}(r) \mathrm{d}r,$$

where C_d is the surface area of the $(d-1)$-sphere. That is, we know that the right density q of R should satisfy

$$q(r) = c' r^{d-1} \hat{p}(r),$$

where $c' = cC_d$. The constant is unique, because q is a probability density. In fact,

$$c' = \bigg(\int_0^\infty r^{d-1} \hat{p}(r) \mathrm{d}r \bigg)^{-1}.$$

□

Example 2.17 (Uniform distribution on a d-ball)*.* If $V \sim \mathcal{U}(S^{d-1})$ and $U \sim \mathcal{U}(0, 1)$, then $Z = U^{1/d} V \sim \mathcal{U}(B^d)$, where $B^d := \{x \in \mathbb{R}^d : \|x\| \leq 1\}$.

```
n = 1000; d = 2
X = zeros(d, n)
for k = 1:n
    u = rand(); z = randn(d); v = z/sqrt(sum(z.^2))
    X[:,k] = u^(1/d) * v
end
```

Remark 2.18. Elliptically symmetric densities of the form $p(x) = c\hat{p}(\|L^{-1}(x - m)\|)$ with location $m \in \mathbb{R}^d$ and non-singular shape $LL^T \in \mathbb{R}^{d \times d}$ can be simulated by drawing X from the corresponding spherically symmetric distribution with radial decay $\hat{p}$ as in Proposition 2.16 and then transforming $Y = m + LX$; the argument is identical with Proposition 2.11.

3 Rejection sampling

When it is not possible (or efficient) to do transformations of variables to produce variables that are distributed according to a given distribution, rejection sampling (or the 'accept-reject' method) can make sampling possible (or more efficient).

Example 3.1 (Uniform distribution on a disc). Consider the raindrops in Example 1.4, and assume $(V_k)_{k\geq 1} \overset{\text{i.i.d.}}{\sim} \mathcal{U}([0,1]^2)$. Let $(\hat{V}_k)_{k\geq 1}$ consist of those V_k that fall within the unit disc $D := \{(w_1, w_2) \in \mathbb{R}^2 \,:\, w_1^2 + w_2^2 < 1\}$. Then, $(\hat{V}_k) \overset{\text{i.i.d.}}{\sim} \mathcal{U}(D)$, a uniform distribution on the unit disc D.

3.1 Rejection sampling algorithm

To give the general form of rejection sampling, assume that both p and q are p.d.f.s or p.m.f.s on a common space $\mathbb{X}$, and suppose that $M \in [1, \infty)$ is a constant such that

$$\boxed{\text{Assumption: } \frac{p(x)}{q(x)} \leq M \qquad \text{for all } x \in \mathbb{X},} \tag{4}$$

where by convention $0/0 = 0$ and $a/0 = \infty$ for $a > 0$.

Algorithm 3.2 (Rejection sampling)**.** Let $(Y_k)_{k\geq 1} \overset{\text{i.i.d.}}{\sim} q$ which are independent of $(U_k)_{k\geq 1} \overset{\text{i.i.d.}}{\sim} \mathcal{U}(0,1)$. Set $T = 1$ and

(A) If $U_T \leq \dfrac{p(Y_T)}{Mq(Y_T)}$, then output $X = Y_T$.

(R) Otherwise, increment $T = T + 1$ and retry (A).

Remark 3.3. The distribution q in rejection sampling is often called the *proposal distribution* (or the *instrumental distribution*).

Theorem 3.4. *Suppose* (4) *holds and consider* $X = Y_T$ *of Algorithm 3.2.*

(i) The running time $T \sim \text{Geometric}(1/M)$.

(ii) The simulated sample $X \sim p$.

Proof (discrete case). Define

$$h(x) := \begin{cases} \frac{p(x)}{Mq(x)}, & \text{whenever } q(x) > 0, \\ 1, & \text{otherwise.} \end{cases}$$

Denote the 'acceptance indicators' $B_k := \mathbf{1}\left(U_k \leq h(Y_k)\right)$, then B_k are independent Bernoulli random variables, with

$$\begin{aligned} \mathbb{P}(B_k = 1) &= \sum_{y\in\mathbb{X}} \mathbb{P}(B_k = 1, Y_k = y) = \sum_{y\in\mathbb{X}} \mathbb{P}(B_k = 1 \mid Y_k = y)\mathbb{P}(Y_k = y) \\ &= \sum_{y\in\mathbb{X}} \mathbb{P}(U_k \leq h(y))q(y) = \frac{1}{M}\sum_{y\in\mathbb{X}} p(y) = \frac{1}{M}. \end{aligned}$$

That is, $\mathbb{P}(T = t) = \mathbb{P}(B_t = 1)\mathbb{P}(B_1 = 0)\cdots\mathbb{P}(B_{t-1} = 0)$ which proves (i).

Let then $x \in \mathbb{X}$, and calculate for any $t \in \mathbb{N}$,

$$\begin{aligned}\mathbb{P}(X = x \mid T = t) &= \mathbb{P}(Y_t = x \mid B_1 = 0, \dots, B_{t-1} = 0, B_t = 1) \\ &= \mathbb{P}(Y_t = x \mid B_t = 1) \\ &= \frac{\mathbb{P}(Y_t = x, B_t = 1)}{\mathbb{P}(B_t = 1)}.\end{aligned}$$

Similarly as above, for any $x \in \mathbb{X}$, we get

$$\mathbb{P}(Y_t = x, B_t = 1) = q(x)h(x) = p(x)/M.$$

We conclude that $\mathbb{P}(X = x \mid T = t) = p(x)$. □

Remark 3.5. Note that because $\mathbb{P}(X = x \mid T = t) = \mathbb{P}(X = x)$, the running time T and the sample X produced by Algorithm 3.2 are independent.

Because $T \sim \text{Geometric}(1/M)$, the expected running time (expected number of iterations before stopping) is $\mathbb{E}[T] = M$. Therefore, smaller M leads to a more efficient algorithm.

Remark 3.6. The proof in the continuous case is essentially identical, by considering $A \subset \mathbb{X}$ (or cylindrical sets) and calculating $\mathbb{P}(X \in A \mid T = t)$. In particular, notice that

$$\mathbb{P}(Y_t \in A, B_t = 1) = \int_A q(y)h(y)\mathrm{d}y = \frac{1}{M}\int_A p(y)\mathrm{d}y,$$

from which with $A = \mathbb{X}$ we also deduce that $\mathbb{P}(B_t = 1) = 1/M$.

Remark 3.7 (*). It is not difficult to see that the proof of rejection sampling generalises directly into general state spaces. A similar idea, called *thinning* is used in a point process context, in order to simulate a non-homogeneous Poisson process by discarding some points of a homogeneous Poisson process.

Example 3.8. Suppose we want to use rejection sampling to simulate from $N(0,1)$ using standard Cauchy proposals. We have

$$\frac{p(x)}{q(x)} = \sqrt{\frac{\pi}{2}}(1 + x^2)\exp\Big(-\frac{x^2}{2}\Big) \le \sqrt{\frac{2\pi}{e}} =: M,$$

because the ratio is maximised with $x = \pm 1$ (derivative zero also at $x = 0$).

```
using Distributions  # Package w/ all 'standard' distributions; install by:
                     # using Pkg; Pkg.add("Distributions")
function cauchy_normal(n)
    M = sqrt(2pi)*exp(-.5)
    x = zeros(n)
    while n>0
        y = rand(Cauchy())
        if M*rand() < exp(logpdf(Normal(),y) - logpdf(Cauchy(), y))
            x[n] = y;  n = n-1
        end
    end
    x
end
x = cauchy_normal(10_000)
```