Sampling Random Variables

Introduction

Sampling a random variable X means generating a domain value $x \in X$ in such a way that the probability of generating x is in accordance with $p(x)$ (respectively, $f(x)$), the probability distribution (respectively, probability density) function associated with X. In this lecture we show how being able to sample a continuous uniform random variable U over the interval $(0, 1)$ allows one to sample any other distribution of interest. Moreover, an algorithm for sampling from a $U \sim \mathcal{U}(0, 1)$ is referred to as a **pseudorandom number generator (png)**. The development of good png's is both an art and science, and relies heavily on developing a sequence of operations on one or more binary words in order to produce the next random number between $(0, 1)$ (actually, a positive integer x is generated, and then divided by a large constant $y \geq x$ to produce $x/y \in (0, 1)$). These operations include arithmetic modulo a prime number, register shifts, register feedback techniques, and logical operations, such as and, or, and xor. Once a set of operations has been developed to form a png, the number sequences generated by the png are tested using several different statistical tests. The tests are used to confirm different properties that should be found in a sequence of numbers, had that sequence been drawn independently and uniformly over $(0, 1)$.

In this lecture we assume that we have access to a good png for generating independent samples of random variable $U \in \mathcal{U}(0, 1)$. Throughout the remaining lectures, assume that variable U represents a $\mathcal{U}(0, 1)$ random variable.

Sampling Finite and Discrete Random Variables

Sampling a Bernoulli random variable

If

$$X = \begin{cases} 1 & \text{if } U \leq p \\ 0 & \text{otherwise} \end{cases} .$$

then $X \sim Be(p)$ since 1 will be sampled with probability p, and 0 will be sampled with probability $1 - p$.

Discrete inverse transform technique

Let $X = \{x_1, \ldots, x_n\}$ be a random variable with probability distribution p, and where $x_1 \leq \cdots \leq x_n$. Define

$$q_i = P(X \leq x_i) = \sum_{j=1}^{i} p(x_j).$$

Then the following is a sampling formula for X.

$$X = \begin{cases} x_1 & \text{if } U < q_1 \\ x_2 & \text{if } q_1 \leq U < q_2 \\ \vdots & \vdots \\ x_{n-1} & \text{if } q_{n-2} \leq U < q_{n-1} \\ x_n & \text{otherwise} \end{cases}.$$

Indeed $X = x_i$ in the event that $q_{i-1} \leq U < q_i$, which has probability $p = q_i - q_{i-1} = p(x_i)$. This technique is referred to as the **discrete inverse transform** technique, since it involves computing $F^{-1}(U)$, where F is the CDF of X. Of course, since F is not one-to-one in the case that X is finite, here $F^{-1}(U)$ is defined as the least element $x \in X$ for which $U < F(x)$.

The Cutpoint method

This inverse-transform method has the advantage of having an optimal $O(n)$ setup time. However, the average number of steps required to sample X is not optimal, and if several samples of X are needed, then the **cutpoint method** offers an average number of two comparison steps needed to sample an observation, yet still has an $O(n)$ initial setup time.

Without loss of generality, we can assume that $X = [1, n]$. Also, let $q_i = P(X \leq i)$. Then the idea behind the cutpoint method is to choose $m \geq n$, and define sets $Q_1, \ldots, Q_m$ for which

$$Q_i = \{q_j | j = F^{-1}(U) \text{ for some } U \in [\frac{i-1}{m}, \frac{i}{m})\},$$

for all $i = 1, \ldots, m$. In words, the unit interval $[0, 1]$ is partitioned into m equal sub-intervals of the form $[\frac{i-1}{m}, \frac{i}{m})$, $i = 1, \ldots, m$. And when U falls into the i th sub-interval, then Q_i contains all the possible q_j values for which $F^{-1}(U) = j$. That way, instead of searching through all of the q values, we save time by only examining the q_j values in Q_i, since these are the only possible values for which $F^{-1}(U) = j$.

The algorithm is now described as follows. sample $U \sim U(0, 1)$, and let $i = \lceil mU \rceil$. Then (assuming Q_i is sorted) find the first $q_j \in Q_i$ for which $U < q_j$. Return j.

Example 1. Given the distribution $(.2, .05, .02, .03, .3, .25, .1, .05)$ and using $m = 8$, compute the sets $Q_1, \ldots, Q_8$.

Theorem 1. Assuming $m \geq n$, the expected number of q values that must be compared with U during the cutpoint algorithm is bounded by two. Therefore, sampling X can be performed in O(1) steps.

Proof of Theorem 1. Upon sampling U, let E_i, $i = 1, \ldots, m$ denote the event that $U \in [\frac{i-1}{m}, \frac{i}{m})$. Also, denote by r the number of Q sets for which $|Q| \geq 2$. Moreover, if R denotes the set of indices i for which $|Q_i| \geq 2$, then we claim that

$$\sum_{i \in R} |Q_i| \leq n + r.$$

To see this, first notice that each such Q_i must contain at least one q value for which $q \notin Q_j$, for all $j = 1, \ldots, i-1$. Moreover, there can be at most r instances where an element of Q_i, $i \in R$, also appears in Q_{i+1}, $i+1 \in R$. In other words, in the worst case all n elements are contained in some Q_i, $i \in R$, and there can be at most r elements that are double counted.

Now, let C be a random variable that counts the number of comparisons of U with a q value. Then,

$$E[C] = \sum_{i=1}^{n} E[C|E_i]P(E_i) \leq \frac{1}{m}\sum_{i=1}^{m} |Q_i| = \frac{1}{m}\left(\sum_{i \in R} |Q_i| + \sum_{i \in \overline{R}} |Q_i|\right)$$

$$\leq \frac{1}{m}\left[(n+r) + (m-r)\right] = \frac{1}{m}(n+m) \leq \frac{2m}{m} = 2.$$

Here we are using the facts that i) $|\overline{R}| = m - r$ and ii) $|Q_i| = 1$ for all $i \in \overline{R}$.

Theorem 2: Geometric Random Variables. If $U \sim U(0,1)$, then

$$X = \lfloor \frac{\ln U}{\ln q} \rfloor + 1.$$

has a geometric distribution with parameter $p = 1 - q$; i.e. $X \sim G(p)$.

Proof. First sample $U \sim U(0,1)$. Then return k, where

$$\sum_{n=1}^{k-1} (1-p)^{n-1} p \leq U < \sum_{n=1}^{k} (1-p)^{n-1} p. \tag{1}$$

Then using the formula for geometric series

$$\sum_{n=1}^{k} ar^{n-1} = a \frac{r^k - 1}{r - 1},$$

some algebra shows that Equation 1 implies

$$1 - (1-p)^{k-1} \leq U < 1 - (1-p)^k \Rightarrow$$

$$(1-p)^k < 1 - U \leq (1-p)^{k-1}.$$

Taking logs of all sides and dividing by the negative number $\ln(1-p)$ then yields

$$k - 1 \leq \frac{\ln(1-U)}{\ln(1-p)} < k \Rightarrow$$

$$k = \lfloor \frac{\ln(1-U)}{\ln(1-p)} \rfloor + 1.$$

Finally, letting $q = 1 - p$, and noting that $1 - U$ is also uniformly distributed over $[0,1]$, we have

$$k = \lfloor \frac{\ln U}{\ln q} \rfloor + 1.$$

QED

Binomial $B(n,p)$**.** If $X \sim B(n,p)$ then an observation of X can be sampled by summing n indepenent Bernoulli random variables $X_1, \ldots, X_n$. Note that the generating cost is $O(n)$. Also, the cutpoint method may also be used. Or if $q = \min(p, 1-p)$ is very small, then one can use a sum of geometric random variables with the expected number of steps equal to $O(qn)$.

Poisson $P(\lambda)$**.** Similar to a binomial random variable, an observation for a Poisson random variable can be sampled by simulating the arrival of customers over a unit time interval for which their interarrival distribution is $E(\lambda)$. The sampled value equals the number of arrivals. Also, a modified version of the cutpoint method may be used in which the cumulative probabilities q_i are computed so long as $q_i \leq 1 - 1/n$, where n is large and equal to the number of desired samples. Then, should $U > 1 - 1/n$ occur, one may compute additional q_i values as needed.

Negative Binomial $NB(r,p)$**.** If $X \sim NB(r,p)$ then an observation of X can be sampled by summing r geometric random variables $X_1, \ldots, X_r$.

Hypergeometric $HG(m,n,r)$**.** $X \sim HG(m,n,r)$ can be sampled by creating an array a_0 of length $m+n$ in which m cells are marked as `blue`, and the remaining cells are marked as `red`. Then array a_i, $i = 1, \ldots, r$, is obtained by considering a_{i-1} and swapping the marking of cell i with the marking of a randomly selected cell from $i, i+1, \ldots, m+n$. Then X equals the number of the first r cells of a_r that are marked as `blue`.

Inverse Transform Technique

Theorem 3. Let X be a continuous random variable with cdf $F(x)$ which possesses an inverse F^{-1}. Let $U \sim U(0,1)$ and $Y = F^{-1}(U)$, then $F(x)$ is the cdf for Y. In other words, Y has the same distribution as X.

Proof. It suffices to show that Y has the same cdf as X. Letting F and F_Y denote the respective cdf's of X and Y respectively.Then

$$F_Y(x) = P(Y \leq x) = P(F^{-1}(U) \leq x) = P(F(F^{-1}(U)) \leq F(x)) =$$

$$P(U \leq F(x)) = F(x),$$

where the last equality follows from the fact that $U \sim U(0,1)$, and the third-to-last equality follows from the fact that F is strictly increasing.

Corollary. Let $U \sim \mathcal{U}(0,1)$ be a uniform random variable. Then the following random-variables have the indicated distributions.

Uniform $X \sim U(a,b)$ $X = a + U(b-a)$

Exponential $X \sim E(\lambda)$ $X = -\ln(U)/\lambda$

Weibull $X \sim We(\alpha,\beta,\nu)$ $X = \nu + \alpha[-\ln(U)]^{1/\beta}$

Triangular $X \sim T(a,b,c)$

$$X = \begin{cases} a + \sqrt{U(b-a)(c-a)} & \text{if } U \leq \frac{b-a}{c-a} \\ c - \sqrt{(1-U)(c-a)(c-b)} & \text{otherwise} \end{cases}$$

Cauchy $X \sim C(\mu,\sigma^2)$ $X = \mu + \sigma \tan \pi(U - \frac{1}{2})$

Example 2. Prove the corollary for the uniform, exponential, and Cauchy cases.

Example 2 Continued.

Empirical Cumulative Distribution Functions With Linear Interpolation

Empirical cdf's are used to model continuous distributions. Let $x_1 \leq x_2 \leq \ldots \leq x_n$ be a sorted collection of n data points where each $x_i \in [a, \infty)$ for some real number a. Then the empirical cdf $F(x)$ with linear interpolation is defined in the following steps.

1. Given $x \in \{x_1, x_2, \ldots, x_n\}$, let i be the largest index for which $x = x_i$ then $F(x) = \frac{i}{n}$
2. $F(x) = 0$ for all $x \leq a$
3. $F(x) = 1$ for all $x \geq x_n$
4. if $x \in (a, x_1)$, then $F(x) = \frac{F(x_1)}{x_1 - a}(x - a)$
5. if $x \in (x_i, x_{i+1})$, then $F(x) = F(x_i) + \frac{(x - x_i)[F(x_{i+1}) - F(x_i)]}{(x_{i+1} - x_i)}$

Example 3. Let $a = 0$ and suppose $1, 1, 2, 5, 7$ are 5 data points. Sketch a graph of the empirical cdf $F(x)$ with linear interpolation with respect to this data. Compute the following: $F(-1)$, $F(.3)$, $F(2)$, $F(4)$, $F(8)$.

Sampling an empirical cdfs with linear interpolation. Let $F(x)$ be an empirical cdf with linear interpolation with respect to data $x_1 \leq x_2 \leq \ldots \leq x_n$, where each $x_i \in [a, \infty)$. Then the following procedure can be used sample a value for random variable X, where X has cdf $F(x)$.

1. sample random U where $U \sim U(0,1)$
2. if $U = 0$ return a.
3. else if $U = F(x_i)$ for some $1 \leq i \leq n$, then return x_i
4. else if $U < F(x_1)$ then return
$$a + (x_1 - a)\frac{U}{F(x_1)}$$
5. else $F(x_i) < U < F(x_{i+1})$, and return
$$x_i + (x_{i+1} - x_i)\frac{(U - F(x_i))}{(F(x_{i+1}) - F(x_i))}$$

Example 4. For the cdf of Example 3, what values for X get sampled for values of $U = .1, .5, .8$?

Acceptance-Rejection Method

Theorem 4: Acceptance-Rejection (AR) Method. Let f and η be density functions over set $S \subseteq \mathcal{R}$ with property that

$$\kappa(f, \eta) = \max_{x \in S} \frac{f(x)}{\eta(x)}$$

is finite. Then if one repeatedly samples a value $x \in S$ using density η, followed by sampling $U \sim \mathcal{U}(0, 1)$, until it is true that

$$U \leq \frac{f(x)}{\kappa(f, \eta)\eta(x)}$$

(in which case we say that x has been *accepted*), Then the accepted value has density function $f(x)$.

Proof. Let A denote the event $U \leq \frac{f(x)}{\kappa(f,\eta)\eta(x)}$, and $k(x|A)$ denote the conditional density of x given A. Then using Baye's rule,

$$k(x|A) = \frac{P(A|x)\eta(x)}{P(A)}. \tag{2}$$

But

$$P(A|x) = \frac{f(x)}{\kappa(f, \eta)\eta(x)}.$$

Moreover,

$$P(A) = \int_S P(A|x)\eta(x)dx = \int_S \frac{f(x)}{\kappa(f, \eta)\eta(x)}\eta(x)dx = \int_S \frac{f(x)}{\kappa(f, \eta)}dx = \frac{1}{\kappa(f, \eta)},$$

where the last equality follows from the fact that $f(x)$ is a density function. Substituting for $P(A|x)$ and $P(A)$ in Equation 2 yields the desired result.

Example 5. Random variable X having density $f(x) = \sqrt{\frac{2}{\pi}}e^{-x^2/2}$ is said to have the **half normal distribution**, since the density function represents the positive half of the standard normal density. Using $\eta(x) = e^{-x}$, determine the average number of U samples that are needed in order to sample a value of X using the AR method.

Example 6. Recall that the gamma distribution $\text{Ga}(1, \alpha)$, for $0 < \alpha < 1$ has density function $e^{-x}x^{\alpha-1}/\Gamma(\alpha)$. Using $\eta(x)$ defined by

$$\eta(x) = \begin{cases} \frac{e\alpha x^{\alpha-1}}{\alpha+e} & \text{if } 0 \leq x \leq 1 \\ \frac{\alpha e^{-x+1}}{\alpha+e} & \text{if } x > 1 \end{cases}$$

can be used to sample an observation for $X \sim \text{Ga}(1, \alpha)$ using the AR method. Determine the average number of U samples that are needed in order to sample a value of X.

Sampling a Standard Normal Variable

Random variable transformations

Henceforth we use the notation $\overline{x}$ to denote the vector $(x_1, \ldots, x_n)$. Let $T(\overline{x}) = (T_1(\overline{x}), \ldots, T_n(\overline{x}))$ be a smooth (i.e. differentiable) transformation from R^n to R^n, then the **Jacobian** of the transformation, denoted $J_T(\overline{x})$ is defined as the determinant of the matrix whose (i, j) entry is $\frac{\partial T_i}{\partial x_j}(\overline{x})$.

Example 7. Consider the smooth transformation $T(r, \theta)$ defined by the equations $x = r\cos\theta$ and $y = r\sin\theta$. Compute $J_T(r, \theta)$.

Now suppose smooth transformation $\overline{y} = T(\overline{x})$ has an inverse $T^{-1}(\overline{y})$. Then it can be proved that

$$J_{T^{-1}}(\overline{y}) = 1/J_T(T^{-1}(\overline{y})).$$

More generally, the matrix of partial derivatives with entries $\frac{\partial T_i}{\partial x_j}(\overline{x})$ is invertible, and its inverse is the matrix of partial derivatives whose entries are $\frac{\partial T_i^{-1}}{\partial x_j}(\overline{y})$.

Example 8. For the transformation from Example 7, compute $J_{T^{-1}}(x, y)$ and verify that

$$J_{T^{-1}}(\overline{y}) = 1/J_T(T^{-1}(\overline{y})).$$

The following result is stated without proof.

Change of Variables Formula for Integration. Let $\overline{y} = T(\overline{x})$, $x \in S \subset \mathcal{R}^n$, be a smooth one-to-one transformation from S to $T(S)$. Then for Riemann integrable $f(\overline{x})$,

$$\int_S f(\overline{x})d\overline{x} = \int_{T^{-1}(S)} f(T^{-1}(\overline{y}))J_{T^{-1}}(\overline{y})d\overline{y}.$$

Example 9. Use a change of variables to compute

$$\int_0^1 \sqrt{1-x^2}\ dx.$$

Example 10. Use a change of variables to show that $f(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ is in fact a density function (i.e. $\int_{-\infty}^{\infty} f(x) = 1$). Hint: work with the joint density $f(x,y) = \frac{1}{2\pi}e^{-(x^2+y^2)/2}$.

Sampling a standard normal via transformation

Now let X and Y be independent standard normal variables and consider the vector $(X,Y) \in \mathcal{R}^2$. Introduce random variables $D \geq 0$ and $\Theta \in [-\frac{\pi}{2}\frac{\pi}{2}]$ that are defined by the transformation T whose equations are $D = X^2 + Y^2$ and $\Theta = \tan^{-1}(\frac{Y}{X})$. Now, since X and Y are independent, they have joint distribution

$$f(x,y) = \frac{1}{2\pi}e^{-(x^2+y^2)/2}.$$

Moreover, since $d = x^2 + y^2$, and $J_T(x,y) = 2$, it follows from the change-of-variables formula that D and Θ have the joint distribution

$$f(d,\theta) = \frac{1}{2\pi} \cdot \frac{1}{2}e^{-d/2},$$

which implies that D has the exponential distribution with $\lambda = 1/2$, and Θ has the uniform distribution over $[-\frac{\pi}{2}, \frac{\pi}{2}]$. Moreover, when sampling from these two distributions, one can recover standard normal Y using the equation

$$Y = \sqrt{d}\sin\Theta,$$

and hence Y may be sampled using $U_1, U_2 \sim \mathcal{U}(0,1)$ in the equation

$$Y = \sqrt{-2\ln U_1}\cos\pi U_2.$$

Miscellaneous Results

Theorem 5. For $\alpha > 1$ the following algorithm can be used to sample from $\text{Ga}(1,\alpha)$ with constant mean evaluation time.

Input α

$a = \alpha - 1$, $b = (\alpha - 1)/(6a\alpha)$, $m = 2/a$, $d = m + 2$, $c = \sqrt{\alpha}$

Repeat forever

- $X \leftarrow \infty$
- While $X \notin (0,1)$
  - Sample $X, Y \sim U(0,1)$
  - $X \leftarrow Y + (1 - 1.857764X)/c$
- $V \leftarrow bY/X$
- If $mX - d + V + V^{-1} \leq 0$, then return aV
- If $m\ln X - \ln V + V - 1 \leq 0$, then return aV

Theorem 6.

1. If $X \sim \text{Ga}(1, \alpha)$, then $\frac{X}{\lambda} \sim \text{Ga}(\lambda, \alpha)$.
2. If $X_1 \sim \text{Ga}(1, \alpha)$, $X_2 \sim \text{Ga}(1, \beta)$, and X_1, X_2 are independent, then $\frac{X_1}{X_1+X_2} \sim \text{Be}(\alpha, \beta)$.
3. If $X \sim N(0, 1)$, then $\mu + \sigma X \sim N(\mu, \sigma^2)$.
4. If $X \sim N(\mu, \sigma^2)$, then $e^X \sim \text{LN}(\mu, \sigma^2)$.

Exercises

1. If X is a random variable with $\text{dom}(X) = \{1, 2, 3, 4\}$ and $p(1) = 0.5$, $p(2) = 0.125 = p(4)$, and $p(3) = 0.25$, then use the discrete inverse transform method for providing a method of sampling X.

2. Use the discrete inverse transform method for providing a method of sampling $X \sim B(5, 0.25)$. Approximate all quantiles to three decimal places.

3. For the random variable X of Exercise 1, provide the Q sets if using the cutpoint method to sample X with $m = 4$.

4. For the random variable X of Exercise 2, provide the Q sets if using the cutpoint method to sample X with $m = 6$.

5. If $X \sim B(105, 0.1)$, then how many samples of a geometric random variable $Y \sim G(0.1)$ are expected to be taken in order to sample X? Explain.

6. Provide pseudocode for using a geometric random variable to sample $X \sim \text{NB}(r, p)$

7. Given a probability distribution $p_1, \ldots, p_n$, $n \geq 2$, prove that there is at least one i for which $p_i < 1/(n-1)$. For this i prove that there is at least one j for which $p_i + p_j \geq 1/(n-1)$. Hint: use proof by contradiction.

8. Let random variable X have domain $\{1, 2, \ldots\}$, and suppose $p_n = P(X = n)$, $n = 1, 2, \ldots$. Define the **hazard rate** λ_n as

$$\lambda_n = P(X = n | X > n - 1) = \frac{p_n}{1 - \sum_{i=1}^{n-1} p_i}.$$

For example, if X represents the month that a device will stop working, then λ_n gives the probability that the device will break during month n, on condition that it has been working for the first $n - 1$ months. Prove that $p_1 = \lambda_1$ and

$$p_n = (1 - \lambda_1) \cdots (1 - \lambda_{n-1}) \lambda_n,$$

for all $n \geq 2$.

9. Let random variable X have domain $\{1, 2, \ldots\}$, and hazard rates (see previous exercise) $\lambda_1, \lambda_2, \ldots$. Moreover, suppose $\lambda_n \leq \lambda$, for all $n \geq 1$. Consider the following algorithm for sampling X.

   Step 1: $S = 0$.

   Step 2: sample $Y \sim G(\lambda)$.

   Step 3: $S = S + Y$.

   Step 4: sample $U \sim \mathcal{U}(0, 1)$

   Step 5: if $U \leq \lambda_S/\lambda$, then return $X = S$. Otherwise go to Step 2.

   Prove that this algorithm is correct. In other words, prove that the probability of sampling $X = n$ is equal to

$$p_n = (1 - \lambda_1) \cdots (1 - \lambda_{n-1})\lambda_n.$$

10. Suppose $p_0, p_1, \ldots$ and $r_0, r_1, \ldots$ are probability distributions for which no probability from either distribution is equal to zero. Moreover, suppose that $p_i/p_j = r_i/r_j$ for all $i, j \geq 0$. Prove that the distributions are identical, i.e. $p_i = r_i$, for all $i \geq 0$. Note: an analagous result holds for continuous density functions.

11. Suppose X, Y, and W are discrete random variables with the property that, for some fixed j,

$$P(W = i) = P(X = i|Y = j),$$

   for all $i = 1, 2, \ldots$. Assume an algorithm exists for sampling X. Prove that the following algorithm may be used to sample W.

   Step 1: sample X to obtain value i.

   Step 2: sample $U \sim \mathcal{U}(0, 1)$.

   Step 3: if $U \leq P(Y = j|X = i)$, return i.

   Step 4: go to Step 1.

12. Provide a method for sampling random variable X with density function $f(x) = e^x/(e - 1)$, for $0 \leq x \leq 1$.

13. Provide a method for sampling random variable X with density function

$$f(x) = \begin{cases} \frac{x-2}{2} & \text{if } 2 \leq x \leq 3 \\ \frac{2-x/3}{2} & \text{if } 3 \leq x \leq 6 \end{cases}$$

14. Use the inverse transform method for providing a method for sampling random variable X with CDF $F(x) = \frac{x^2+x}{2}$, $0 \leq x \leq 1$.

15. The following data is to be used for the creation of an empirical CDF $F(x)$ with linear interpolation:

$$1.58, 1.83, 0.71, 0.10, 0.88, 0.70, 1.36, 0.65, 3.37, 0.42.$$

   Assuming $a = 0$, compute $F(1.58)$, $F(0.5)$, and $F(0.025)$.

16. For the emprical CDF $F(x)$ from the previous exercise, compute $F^{-1}(0.75)$, $F^{-1}(0.34)$, and $F^{-1}(0.01)$.

17. Suppose X has CDF

$$F(x) = \sum_{i=1}^{n} p_i F_i(x)$$

where $p_1 + \cdots + p_n = 1$ and F_i is a CDF with a well-defined inverse F_i^{-1}, for all $i = 1, \ldots, n$. Consider the following method for sampling X. First sample finite random variable I, where $\text{dom}(I) = \{1, \ldots, n\}$, and $p(i) = p_i$, for all $i = 1, \ldots, n$. Let i be the sampled value. Next, sample $U \sim U(0,1)$, and return $Y = F_i^{-1}(U)$. Prove that Y has a CDF equal to $F(x)$.

18. Use the result from the previous exercise to provide a method for sampling random variable X with CDF

$$F(x) = \begin{cases} \frac{1-e^{-2x}+2x}{3} & \text{if } 0 < x < 1 \\ \frac{3-e^{-2x}}{3} & \text{if } x \geq 1 \end{cases}.$$

19. If $F_1(x), \ldots, F_n(x)$ are CDFs, prove that

$$F(x) = \prod_{i=1}^{n} F_i(x)$$

is a CDF. Provide an algorithm for sampling from $F(x)$, assuming algorithms for sampling each of the $F_i(x)$.

20. Given random variable X having density function $f(x) = 1/4 + 2x^3 + 5/4x^4$, $0 < x < 1$, find an appropriate $\eta(x)$ so that X can be sampled using acceptance-rejection method. Determine $\kappa(f, \eta)$.

21. For transformation $T(x,y) = (xy - 3x^2, 2xy - y^2)$, compute $J_T(x,y)$.

22. Let C denote the circular region defined by the equation $x^2 + y^2 = 16$. Use a change of variables to evaluate

$$\int_C 100 - x^2 - y^2 dxdy.$$

Exercise Solutions

1. For $U \sim U(0,1)$,

$$X = \begin{cases} 1 & \text{if } U < 0.5 \\ 2 & \text{if } 0.5 \leq U < 0.625 \\ 3 & \text{if } 0.625 \leq U < 0.875 \\ 4 & \text{otherwise} \end{cases}$$

2. For $U \sim U(0,1)$,

$$X = \begin{cases} 0 & \text{if } U < 0.237 \\ 1 & \text{if } 0.237 \leq U < 0.633 \\ 2 & \text{if } 0.633 \leq U < 0.896 \\ 3 & \text{if } 0.896 \leq U < 0.984 \\ 4 & \text{if } 0.984 \leq U < 0.999 \\ 5 & \text{otherwise} \end{cases}$$

3. $Q_1 = \{q_1\}$, $Q_2 = \{q_1\}$, $Q_3 = \{q_2, q_3\}$, $Q_4 = \{q_3, q_4\}$.

4. $Q_1 = \{q_0\}$, $Q_2 = \{q_0, q_1\}$, $Q_3 = \{q_1\}$, $Q_4 = \{q_1, q_2\}$., $Q_5 = \{q_2\}$, $Q_6 = \{q_2, q_3, q_4, q_5\}$

5. The expected number of Y samples is $\lceil 105/10 \rceil = 11$ since the expected value of each Y sample is 10.

6. The code assumes the existence of function `sample_geo` which returns a geometric sample on input probability p.

```
int sample_negative_binomial(int r, double p)
{
     int count = 0; //count additional trials beyond r
     int i;

     for(i = 0; i < r; i++)
          //subtract one (i.e. the success) from the sample
          //and add to count
          count += sample_geo(p)-1;

     return count;
}
```

7. Assume $p_i \geq 1/(n-1)$ for all $i = 1, \ldots, n$. Then

$$\sum_{i=1}^{n} p_i = 1 \geq n/(n-1) > 1,$$

is a contradiction.

Now suppose $p_i < 1/(n-1)$, and, for $j \neq i$, $p_i + p_j < 1/(n-1)$. Then

$$\sum_{j \neq i} (p_i + p_j) < (n-1)/(n-1) = 1.$$

But on the other hand,

$$\sum_{j \neq i} (p_i + p_j) = (n-1)p_i + (1 - p_i) = 1 + (n-2)p_i \geq 1,$$

a contradiction.

8. By definition,

$$\lambda_1 = P(X = 1 | X > 0) = \frac{p_1}{1 - \sum_{i=1}^{0} p_i} = p_1.$$

Hence, $p_1 = \lambda_1$ and $P(X > 1) = 1 - \lambda_1$. Now assume that

$$P(X > n) = (1 - \lambda_1) \cdots (1 - \lambda_n),$$

is true for some $n \geq 1$ (it is certainly true for $n = 1$). Then

$$P(X > n+1) = P(X > n+1|X > n)P(X > n) + P(X > n+1|X \leq n)P(X \leq n) =$$

$$(1-\lambda_{n+1})(1-\lambda_1)\cdots(1-\lambda_n) + 0 = (1-\lambda_1)\cdots(1-\lambda_n)(1-\lambda_{n+1}).$$

Hence, by induction,

$$P(X > n) = (1-\lambda_1)\cdots(1-\lambda_n)$$

for all $n \geq 1$. Therefore,

$$p_n = P(X = n|X > n-1)P(X > n-1) = \lambda_n(1-\lambda_1)\cdots(1-\lambda_{n-1}).$$

9. Let $p_n = P(X = n)$. The key idea is that the geometric random variable Y may be replaced by a sequence of independent Bernoulli random variables $B_1, B_2, \ldots$, where $P(B_n = 1) = \lambda$, for all $n \geq 1$. Now suppose stage $n \geq 1$ has been reached, if $B_n = 0$, then proceed to stage $n+1$. Otherwise, sample U and return n if $U \leq \lambda_n/\lambda$. Otherwise, proceed to stage $n+1$.

   Notice how the above algorithm is identical to the one described in the exercise, since sampling a geometric with success probability λ is equivalent to continually sampling independent Bernoulli random variables until the value 1 has been observed. Notice also that $p_1 = (\lambda)(\lambda_1/\lambda) = \lambda_1$, and thus $P(X > 1) = 1 - \lambda_1$.

   Now assume that

$$P(X > n) = (1-\lambda_1)\cdots(1-\lambda_n),$$

   is true for some $n \geq 1$ (it is certainly true for $n = 1$). Then, as in the previous exercise,

$$P(X > n+1) = P(X > n+1|X > n)P(X > n) = P(X > n+1|X > n)(1-\lambda_1)\cdots(1-\lambda_n).$$

   Now, the probability of moving past stage $n+1$ given that stage $n+1$ was reached, is equal to

$$(1-\lambda) + (\lambda)(1 - \frac{\lambda_{n+1}}{\lambda}) = 1 - \lambda_{n+1}.$$

   Hence,

$$P(X > n+1) = (1-\lambda_1)\cdots(1-\lambda_n)(1-\lambda_{n+1}).$$

   and so, by induction,

$$P(X > n) = (1-\lambda_1)\cdots(1-\lambda_n)$$

   for all $n \geq 1$. And a consequence of this is that

$$p_n = P(X = n) = P(X = n|X > n-1)P(X > n-1) = (1-\lambda_1)\cdots(1-\lambda_{n-1})(\lambda)(\lambda_n/\lambda) =$$

$$(1-\lambda_1)\cdots(1-\lambda_{n-1})\lambda_n.$$

10. For fixed and arbitrary j, it follows that the sequence of numbers $p_0/p_j, p_1/p_j, \ldots$ and $r_0/r_j, r_1/r_j, \ldots$ are identical. Hence,

$$\sum_{i=0}^{\infty} p_i/p_j = \sum_{i=0}^{\infty} r_i/r_j.$$

    But

$$\sum_{i=0}^{\infty} p_i/p_j = \frac{1}{p_j}\sum_{i=1}^{\infty} p_i = (\frac{1}{p_j})(1) = \frac{1}{p_j}.$$

Similarly,

$$\sum_{i=0}^{\infty} r_i/r_j = \frac{1}{r_j}.$$

Thus, $1/p_j = 1/r_j$, i.e. $p_j = r_j$ and, since j was arbitrary, the two distributions are equal.

11. Notice that

$$P(W = i) = P(X = i|Y = j) = \frac{P(Y = j|X = i)P(X = i)}{P(Y = j)}.$$

Moreover, in a single pass through Steps 1-3 of the algorithm, i will be sampled/returned with probability $p_i = P(Y = j|X = i)P(X = i)$. Hence, the probability that the algorithm returns i equals p_i/c, where

$$c = \sum_{r=1}^{\infty} p_r.$$

Hence, for arbitrary i and k,

$$p_i/p_k = P(Y = j|X = i)P(X = i)/(P(Y = j|X = k)P(X = k)) = P(W = i)/P(W = k).$$

Therefore, by the previous exercise, the algorithm samples a random variable that has the same probability distribution as W.

12. Using the inverse transform method, $X = \ln(U(e-1) + 1)$ has the desired distribution.

13. Using the inverse transform method,

$$X = \begin{cases} 2 + 2\sqrt{U} & \text{if } 0 \le U \le 1/4 \\ 6 - 2\sqrt{3 - 3U} & \text{if } 1/4 \le U \le 1 \end{cases}$$

14. Using the inverse transform method, $X = \frac{-1+\sqrt{1+8U}}{2}$ has the desired distribution.

15. $F(1.62) = 0.8 + \frac{(0.1)(0.04)}{1.83-1.58} = 0.816$, $F(0.5) = 0.71$, and $F(0.025) = 0.1(0.025)/(0.1) = 0.025$.

16. $F^{-1}(0.75) = (1.36 + 1.58)/2 = 1.47$, $F^{-1}(0.34) = 0.65 + (0.4)(0.05) = 0.67$, and $F^{-1}(0.01) = 0.01$.

17. Let $I = i$ denote the event that i was sampled in the first step of the algorithm, and let U denote the uniform random variable that is sampled in the algorithm.

$$P(Y \le x) = \sum_{i=1}^{n} P(Y \le x|I = i)P(I = i) = \sum_{i=1}^{n} P(F_i^{-1}(U) \le x)p_i =$$

$$\sum_{i=1}^{n} P(F_i(F_i^{-1}(U)) \le F_i(x))p_i = \sum_{i=1}^{n} P(U \le F_i(x))p_i = \sum_{i=1}^{n} F_i(x)p_i = F(x).$$

Therefore, Y has CDF equal to F and has the same distribution as X.

18. $F(x)$ can be written as

$$F(x) = 1/3(1 - e^{-2x}) + 2/3F_2(x),$$

where

$$F_2(x) = \begin{cases} x & \text{if } 0 < x < 1 \\ 1 & \text{if } x \geq 1 \end{cases} .$$

Thus, $F_2(x)$ is the CDF for $U(0,1)$. Therefore, the algorithm is to first sample $U \sim U(0,1)$. If $U \leq 1/3$, then return a sample with exponential distribution $E(2)$. Otherwise, return a sample from $U(0,1)$.