Unit 1: Generating Random Variables

Chapter 3 in "Statistical Computing with R"

Anna Ly
Department of Mathematical and Computational Sciences
University of Toronto Mississauga

January 5, 2026

Overview

1. The Inverse Transform Method
2. The Acceptance-Rejection Method
3. Transformation and Convolution Methods
4. Mixtures Methods

The Inverse Transform Method

The Inverse Transform Method

- Strictly speaking, it is impossible to get random numbers from a computer; but programs can produce pseudo-random numbers.
- In this course we assume that a suitable uniform pseudo-random number generator is available. (The methods for creating pseudo-randomness is not the focus of this course, and shall be omitted.)
- Refer to ‘help(.Random.seed)‘ for details about the default random number generator in R.

The Inverse Transform Method

Theorem 3.1 (Probability Integral Transform)

If X is a continuous random variable with cdf $F_X(x)$ then $U := F_X(X) \sim \mathit{Uniform}(0, 1)$. (Remark: U is defined as a composition of functions here!)

Proof. Define the inverse transform:

$$F_X^{-1}(u) = \inf\{x : F_X(x) = u\}, \quad 0 < u < 1.$$

Thus, for all $x \in \mathbb{R}$:

$$\begin{aligned}
\mathbb{P}(F_X^{-1}(U) \le x) &= \mathbb{P}(\inf\{x : F_X(x) = U\} \le x) \\
&= \mathbb{P}(F_X(\inf\{x : F_X(x) = U\}) \le F_X(x)) \\
&= \mathbb{P}(\inf\{F_X(x) : F_X(x) = U\} \le F_X(x)) \quad (\dagger) \\
&= \mathbb{P}(\inf\{U\} \le F_X(x)) \\
&= \mathbb{P}(U \le F_X(x)) \\
&= F_X(x) \quad (\text{CDF of } \mathit{Uniform}(0, 1)).
\end{aligned}$$

The Inverse Transform Method

Thus $F_X^{-1}(U)$ has the same distribution as X, and then U has the same distribution as $F_X(x)$. We will now show a proof for ($\dagger$).

- Suppose we define some cdf $F : A \to B$. First, we want to show that $F(\inf A)$ is a lower bound for $F(A)$.
- Pick $y \in A$. Then, since we have that F is non-decreasing, $\inf A < y \quad \Rightarrow \quad F(\inf A) < F(y)$ so $F(\inf A)$ is a lower bound.
- Now WTS that $F(\inf A)$ is the infimum. By right continuity of F, $\forall \epsilon > 0, \exists \delta > 0$ s.t. if $x \in [\inf A, \inf A + \delta)$, then $|F(\inf A) - F(x)| < \epsilon$.
- Since F is non-decreasing, we can drop absolute value and instead say $F(x) - F(\inf A) < \epsilon \quad \Rightarrow \quad F_X(x) < \epsilon + F(\inf A)$.
- Thus, $\epsilon + F(\inf A)$ is not a lower bound and thus $F(\inf A)$ must be the infimum of $F(A)$.

The Inverse Transform Method

Corollary

Let $U \sim Uniform(0, 1)$. Define $X = F^{-1}(U)$, where F is a cdf. Then, F is the cdf of X.

Proof.

$$\begin{aligned} F_X(x) &= \mathbb{P}(X \leq x) \\ &= \mathbb{P}(F^{-1}(U) \leq x) \\ &= \mathbb{P}(U \leq F(X)) \\ &= F(X). \end{aligned}$$

The Inverse Transform Method

The inverse transform method (continuous case) can be summarized as follows:

1. Find the cdf, $F_X(x)$.
2. Define the inverse function $F_X^{-1}(u)$.
3. For each random variate required:
    a Generate a random u from $Uniform(0, 1)$.
    b Deliver $x = F_X^{-1}(u)$.

Warning: we should only use this method if a closed form of $F_X^{-1}(x)$ exists!

The Inverse Transform Method

Example

Use the inverse transform method to simulate a random sample from the distribution with density:

$$f_X(x) = 3x^2, \quad 0 < x < 1.$$

Additionally, write the R code.

Solution. First, we should solve for the cdf:

$$F_X(x) = \int_0^x 3t^2 dt = t^3 + c\Big|_0^x = x^3, \;\; 0 < x < 1.$$

Now, if we generate u from $Uniform(0,1)$:

$$u = x^3 \;\;\Rightarrow\;\; x = u^{1/3} \overset{\text{set}}{=} F_X^{-1}(u).$$

For each random variate required,

1. Generate a random variable $u \sim Uniform(0,1)$.
2. Deliver $x = u^{1/3}$.

The Inverse Transform Method

Example

Use the inverse transform method to simulate a random sample from the exponential distribution, which has the density:

$$f_X(x) = \frac{1}{\theta} e^{-\frac{1}{\theta}x}, \quad x > 0,\, \theta > 0.$$

Additionally, write the R code.

Solution. First, we should solve for the cdf:

$$F_X(x) = \int_0^x \frac{1}{\theta} e^{-t/\theta} dt = 1 - e^{-\frac{x}{\theta}}, \;\; x > 0,$$

Now, if we generate u from $Uniform(0, 1)$:

$$u = 1 - e^{-\frac{x}{\theta}} \;\; \Rightarrow \;\; x = -\theta \ln(1 - u) \overset{\text{set}}{=} F_X^{-1}(u).$$

Since u arises from $Uniform(0, 1)$, we have that $0 < u < 1$ and hence $\ln(1 - u) < 0$.

The Inverse Transform Method

For each random variate required,

1. Generate a random variable $u \sim Uniform(0, 1)$.
2. Deliver $x = u^{1/3}$.

The Inverse Transform Method

Example

Use the inverse transform method to simulate a random sample from the Weibull distribution, which has the density:

$$f_X(x) = \begin{cases} \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} e^{-\left(\frac{x}{\beta}\right)^{\alpha}} & x \geq 0, \alpha > 0, \beta > 0, \\ 0 & \text{otherwise.} \end{cases}$$

To make life easier, we'll consider the case where $\beta = 1$.

1. Additionally, write the R code.
2. Find $\mathbb{E}[X]$, the theoretical mean, and check that the simulated mean is close to the theoretical mean.

Solution.

$$F_X(x) = \int_0^x \alpha t^{\alpha-1} e^{-t^\alpha}\, dt$$

Let $w = t^\alpha$, then $dw = \alpha t^{\alpha-1} dt$. Furthermore, as $t \to 0, w \to 0$, and $t \to x, w \to x^\alpha$. Now,

$$F_X(x) = \int_0^{x^\alpha} e^{-w} dw = 1 - e^{-x^\alpha}, \;\; x \geq 0, \alpha > 0.$$

Furthermore, if we generate u from $Uniform(0,1)$:

$$x = (-\ln(1-u))^{1/\alpha} \overset{\text{set}}{=} F_X^{-1}(u).$$

Since u arises from $Uniform(0,1)$, we have that $0 < u < 1$ and hence $\ln(1-u) < 0$.

Solution. To find the expected value, we solve:

$$
\begin{aligned}
\mathbb{E}[X] &= \int_0^\infty \alpha x^\alpha e^{-x^\alpha} dx \\
&= \int_0^\infty w^{1/\alpha} e^{-w} dw, \quad \text{(Plug } w = x^\alpha\text{)} \\
&= \Gamma\left(\frac{1}{\alpha}+1\right) \underbrace{\int_0^\infty \frac{1}{\Gamma\left(\frac{1}{\alpha}+1\right)} w^{\left(\frac{1}{\alpha}+1\right)-1} e^{-w} dw}_{Gamma(\alpha'=1/\alpha+1, \beta'=1)} \quad \text{(Compare to Gamma density)} \\
&= \Gamma\left(\frac{1}{\alpha}+1\right).
\end{aligned}
$$

The Inverse Transform Method

If we just want to generate a random variable X with pmf:

$$\mathbb{P}(X = x_i) = p_i, \;\; i \in \mathbb{N}, \; \sum_i p_i = 1,$$

then inverse transform method (discrete case) can be summarized as follows:

1. Generate a random u from $Uniform(0, 1)$.
2. Transform u into X as follows:

$$X = x_j \text{ if } F_X(x_{j-1}) < u \leq F_X(x_j)$$

3. It follows that,

$$X = \begin{cases} x_1 & u \leq F_X(x_1), \\ x_2 & F_X(x_1) < u \leq F_X(x_2), \\ \dots & \\ x_j & F_X(x_{j-1}) < u \leq F_X(x_j), \\ \dots & \end{cases}$$

The Inverse Transform Method

In other words, discrete random variables can be generated by slicing up the interval $(0, 1)$ into subintervals which define a partition on $(0, 1)$:

$$(0, F_X(x_1)], (F_X(x_1), F_X(x_2)], (F_X(x_2), F_X(x_3)], \ldots, (F_X(x_{k-1}), 1].$$

We can also define:

$$p_1 = F_X(x_1), \quad p_2 = F_X(x_2) - p_1, \quad \ldots, \quad p_j = F_X(x_j) - \sum_{i=1}^{j-1} p_i.$$

Or equivalently,

$$F_X(x_1) = p_1, \quad F_X(x_2) = p_1 + p_2, \quad \ldots, \quad F_X(x_j) = \sum_{i=1}^{j} p_i.$$

The Inverse Transform Method

Proof of Previous Algorithm

Given X that was defined in the algorithm of the previous slides, Prove that $\mathbb{P}(X = x_j) = p_j$.

Solution.

$$
\begin{aligned}
\mathbb{P}(X = x_j) &= \mathbb{P}(F_X(x_{j-1}) < U < F_X(x_j)) \\
&= F_X(x_j) - F_X(x_{j-1}) \\
&= \sum_{i=1}^{j} p_i - \sum_{i=1}^{j-1} p_i \\
&= p_j.
\end{aligned}
$$

The Inverse Transform Method

Example

Use the inverse transform method to simulate a random sample from the Bernoulli distribution with $p = 0.4$. Additionally, write the R code.

Solution. Since this is a Bernoulli RV, if we generate u from $Uniform(0,1)$:

$$\mathbb{P}(X=0)=0.6, \quad \mathbb{P}(X=1)=0.4, \quad F_X(0)=0.6, \quad F_X(1)=1$$

Deliver the following:

$$X = \begin{cases} 0 & u \leq 0.6, \\ 1 & 0.6 < u < 1. \end{cases}$$

Example

Let X be a discrete random variable with the following pmf:

x	1	2	3	4
$p_X(x)$	0.2	0.5	0.2	0.1

Find $F_X(x)$. Additionally, write the R code.

Solution. Generate u from $Uniform(0, 1)$. Deliver the following:

$$X = \begin{cases} 1 & u \leq 0.2, \\ 2 & 0.2 \leq u < 0.7, \\ 3 & 0.7 \leq u < 0.9, \\ 4 & u \geq 0.9. \end{cases}$$

The Inverse Transform Method

Generalized Inverse Function

For a discrete random variable X with cdf $F_X(x)$, the inverse CDF $F_X^{-1}(p)$, also called the quantile function, is defined as:

$$F_X^{-1}(x) := \inf\{x : F_X(x) \geq p\}, \quad 0 < p < 1.$$

Note that there isn't a unique x where $F(x) = p$ for every $p \in (0,1)$.

Example

From the previous example,

- Find $F_X^{-1}(0.1)$. *Solution.* 1
- Find $F_X^{-1}(0.5)$. *Solution.* 2
- Find $F_X^{-1}(0.85)$. *Solution.* 3

The Inverse Transform Method

Example

Use the inverse transform method to simulate a random sample from the Geometric distribution with pmf:

$$\mathbb{P}(X = i) = pq^i, \;\; i \in \mathbb{N}, \;\; q = 1 - p.$$

Additionally, write the *R* code.

Solution.

$$F_X(x) = \mathbb{P}(X \leq x) = 1 - \mathbb{P}(X > x) = 1 - \mathbb{P}(X \geq x + 1) = 1 - \sum_{i=x+1}^{\infty} pq^i$$

Remember that $\sum_{i=a}^{\infty} r^i = \frac{r^a}{1-r}$. Thus,

$$1 - \sum_{i=x+1}^{\infty} pq^i = 1 - p\frac{q^{x+1}}{(1-q)} = 1 - q^{x+1}$$

The Inverse Transform Method

Now, if we generate u from $Uniform(0,1)$:

$$F_X(x-1) < u \leq F_X(x) \quad \Rightarrow \quad x \ln q > \ln(1-u) \geq (x+1)\ln q$$

Since $0 < q < 1$, $\ln(q) < 0$ and so:

$$x < \frac{\ln(1-u)}{\ln(q)} \leq x+1$$

Deliver:

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

The Inverse Transform Method

Remember that now we are slicing up the interval $(0, 1)$ into subintervals which define a partition on $(0, 1)$:

$$(0, F_X(x_1)], (F_X(x_1), F_X(x_2)], (F_X(x_2), F_X(x_3)], \ldots, (F_X(x_{k-1}), 1].$$

Which basically means we are trying to evaluate $F_X(x_i)$ for $i \in \mathbb{N}$.

- However, sometimes it's hard to obtain a closed form for $F_X(x_i)$.
- Thus, it is more useful to find a recursive formula.

The Inverse Transform Method

Example

Use the inverse transform method to simulate a random sample from the Binomial distribution with pmf:

$$p_i := \mathbb{P}(X = i) = \binom{n}{i} p^i (1-p)^{n-i}, \quad i = 0, 1, \ldots, n.$$

1. First, derive the following recursive formula:

$$p_{i+1} = \left(\frac{n-i}{i+1}\right) \left(\frac{p}{1-p}\right) p_i, \;\; i = 0, 1, \ldots, n.$$

2. Write the *R* code to simulate the algorithm.

Solution.

$$
\begin{aligned}
p_{i+1} &= \binom{n}{i+1} p^{i+1}(1-p)^{n-(i+1)} \\
&= \frac{n!}{(i+1)!}\frac{1}{(n-(i+1))!} p p^i \frac{(1-p)^{n-i}}{1-p} \\
&= \frac{n!}{(i+1)\cdot i!}\frac{(n-i)}{(n-i)!}\frac{p}{1-p} p^i (1-p)^{n-i} \\
&= \left(\frac{n-i}{i+1}\right)\left(\frac{p}{1-p}\right)\binom{n}{i} p^i (1-p)^{n-i} \\
&= \left(\frac{n-i}{i+1}\right)\left(\frac{p}{1-p}\right) p_i
\end{aligned}
$$

The Inverse Transform Method

Example

Use the inverse transform method to simulate a random sample from the logarithmic distribution with pmf:

$$p_i := \mathbb{P}(X = i) = \frac{a\theta^i}{i}, \quad i \in \mathbb{N}.$$

Where $0 < \theta < 1$ and $a = (-\ln(1-\theta))^{-1}$.

1. First, derive the following recursive formula:

$$p_{i+1} = \left(\frac{\theta i}{i+1}\right) p_i, \quad i \in \mathbb{N}.$$

2. Write the R code to simulate the algorithm.

The Inverse Transform Method

Solution.

$$p_{i+1} = \frac{a\theta^{i+1}}{i+1} = \frac{a\theta\theta^{i}}{(i+1)}\frac{i}{i} = \frac{\theta i}{(i+1)}\left(\frac{a\theta^{i}}{i}\right) = \frac{\theta i}{(i+1)}p_i$$

The Inverse Transform Method

In general:

$$X \quad \Rightarrow \quad F_X^{-1}(x) \quad \Rightarrow \quad F_X^{-1}(u)$$

However, this only works if we can define the inverse. We can think of many different functions where the inverse would be hard to find: the Gaussian distribution, beta, etc...

The Acceptance-Rejection Method

Acceptance-Rejection

Suppose we want to generate the random variable X with target density f using the acceptance-rejection method.

1. Find another random variable, Y with trial/candidate/envelope density g where there exists $c \in \mathbb{R}$ such that:
$$\frac{f(t)}{g(t)} < c.$$
2. For each random variate required:
   2.1 Generate y from the distribution with density g.
   2.2 Generate u from the $Uniform(0,1)$ distribution.
   2.3 If
   $$u < \frac{f(y)}{cg(y)}$$
   accept y and deliver $x = y$. Otherwise, reject y and generate a random variate again.

Review

Consider the probability triple $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\{B_1, B_2, \dots\}$ be a partition of Ω. Then, for any $A \in \mathcal{F}$, we have that:

Total Law of Probability

$$\mathbb{P}(A) = \sum_i \mathbb{P}(A|B_i)\mathbb{P}(B_i), \quad i \in \mathbb{N}$$

Bayes Rule

$$\mathbb{P}(B_j|A) = \frac{\mathbb{P}(B_j \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B_j)\mathbb{P}(B_j)}{\sum_{i=1}^{n} \mathbb{P}(B_i)\mathbb{P}(A|B_i)}, \quad i, j \in \mathbb{N}$$

Acceptance-Rejection

Probability of Accepting

Given the acceptance-rejection algorithm, evaluate the probability of acceptance for any iteration (see part 2.3 in the previous slide).

Proof. First we'll show for the discrete case. Note that in step 2.3,

$$\mathbb{P}(accept|Y=y) = \mathbb{P}\left(U < \frac{f(y)}{cg(y)} \middle| Y=y\right) = \frac{f(y)}{cg(y)}$$

Using the law of total probability,

$$\begin{aligned}\mathbb{P}(accept) &= \sum_y \mathbb{P}(accept|Y=y)\mathbb{P}(Y=y)\\ &= \sum_y \frac{f(y)}{cg(y)}g(y)\\ &= \frac{1}{c}\end{aligned}$$

Acceptance-Rejection

Now for the continuous case, let $Y \in A \subseteq \mathbb{R}$ and $\text{dom}(f) \subseteq A$. Then, using the law of total probability:

$$\begin{aligned}
\mathbb{P}(accept) &= \int_A \mathbb{P}(accept | Y = y) g(y) dy \\
&= \int_A \frac{f(y)}{cg(y)} g(y) dy \\
&= \frac{1}{c} \int_A f(y) dy \\
&= \frac{1}{c}
\end{aligned}$$

Aside: Y is a random variable, which is neither random nor a variable. It's a function $Y : \Omega \to \mathbb{R}$. The notation $Y \in A$ means $\{\omega \in \Omega : Y(\omega) \in A\}$.

Acceptance-Rejection

Probability of Accepting

Given the acceptance-rejection algorithm, prove that the accepted sample has the same distribution as X.

Proof. First we will prove the discrete case. Suppose we accept $Y = k$, then:

$$\mathbb{P}(K|accepted) = \frac{\mathbb{P}(accepted|k)g(k)}{\mathbb{P}(accepted)} = \frac{\frac{f(k)}{cg(k)}g(k)}{1/c} = f(k) = \mathbb{P}(X = k)$$

For the continuous case, let $A \subseteq \mathbb{R}$. Then,

$$\begin{aligned}
\mathbb{P}(Y \in A|accepted) &= \frac{\mathbb{P}(accepted \cap \{Y \in A\})}{\mathbb{P}(accepted)} \\
&= \frac{\int_{\mathbb{R}} \mathbb{P}(accepted \cap \{Y \in A\}|Y = y)\mathbb{P}(Y \in dy)}{\mathbb{P}(accepted)}
\end{aligned}$$

Where the previous line is from the law of total probability.

Acceptance-Rejection

$\mathbb{P}(Y \in dy)$ is notation commonly used in probability measure theory, where the probability is computed on a "small" neighborhood of y (layman explanation).

$$= \frac{\int_{\mathbb{R}} \mathbb{P}(accepted \cap \{Y \in A\} | Y = y) g(y) dy}{\mathbb{P}(accepted)}$$

$$= \frac{\int_{\mathbb{R}} \mathbf{1}_{\{y \in A\}} \mathbb{P}(accepted \,|\, Y = y) g(y) dy}{\mathbb{P}(accepted)}$$

$$= \frac{\int_{A} \mathbb{P}(accepted \,|\, Y = y) g(y) dy}{\mathbb{P}(accepted)}$$

$$= \frac{\int_{A} \frac{f(y)}{cg(y)} g(y) dy}{\frac{1}{c}}$$

$$= \int_{A} f(y) dy = \mathbb{P}(X \in A)$$

Acceptance-Rejection

The Choice of c

Given the acceptance-rejection algorithm, show that that $c \geq 1$.

Solution. Recall:

$$0 \leq \mathbb{P}(\cdot) \leq 1, \quad \Rightarrow \quad \mathbb{P}(accept) = \frac{1}{c} \leq 1 \quad \Rightarrow \quad c \geq 1.$$

Acceptance-Rejection

- Ideally, you want g to be easy to sample from, consistent with the support of f.
- Theoretically, as long as you know that f indeed has longer tails than g, you can choose c to be ridiculously large and this will still yield a valid algorithm.
- Since the acceptance rate is equal to $1/c$, you will have to, on average, generate $c \times n$ draws from the trial distribution and from the uniform distribution just to get n draws from the target distribution.
- So choosing c to be too large will yield an inefficient algorithm.
- Ideally: choose c close to 1; so f and g are similar.

The Distribution of N

Let N represent the number of iterations that the acceptance-rejection algorithm needs to successfully generate one value of X. What is the distribution of N?

Solution. In this scenario, waiting until we accept a value of X is analogous to waiting for the first success from a sample of multiple Bernoulli trials; this sounds exactly like the Geometric distribution:

$$\mathbb{P}(X = x) = (1 - p)^{x-1}p, \quad \text{where } p \text{ represents the probability of "success"}.$$

Thus, we say that $N \sim \textit{Geometric}\left(\frac{1}{c}\right)$ since we showed earlier that $\mathbb{P}(\textit{accept}) = \frac{1}{c}$.

Acceptance-Rejection

Example

Use the acceptance-rejection method to simulate a random sample from the distribution with density:

$$f_X(x) = 3x^2, \quad 0 < x < 1.$$

Let the trial distribution be $Uniform(0,1)$. Write the R code to simulate the algorithm. Compare the number of iterations to the value of $\frac{1}{c}$.

Solution. To find c:

$$c = \max_{x \in (0,1)} \left(\frac{f(x)}{g(x)} \right) = \max_{x \in (0,1)} (3x^2)$$

Let $h(x) = f(x)/g(x)$. Note that,

$$h'(x) = 6x \overset{\text{set}}{=} 0 \quad \Rightarrow \quad x = 0$$

Test some points (you do not need to show it is a maximum unless I ask):

$$h(0) = 0, \;\; h(1) = 3 \quad \Rightarrow \quad c = 3$$

Acceptance-Rejection

For each random variate required,

1. Generate $Y \sim Uniform(0, 1)$.
2. Generate $U \sim Uniform(0, 1)$.
3. If
$$u < \frac{f(y)}{cg(y)} = \frac{1}{3}3y^2 = y^2$$
accept y and set $x = y$. Otherwise, reject y and generate a random variate again.

Acceptance-Rejection

Example

Use the acceptance-rejection method to simulate a random sample from the beta distribution with density:

$$f_X(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad 0 < x < 1.$$

Assume $\alpha = 2, \beta = 4$. Let the trial density be *Uniform*$(0,1)$. Write the *R* code to simulate the algorithm. Compare the number of iterations to the value of $\frac{1}{c}$.

Solution. Technically we have that:

$$f_X(x) = \frac{\Gamma(6)}{\Gamma(2)\Gamma(4)} x^1(1-x)^3 = \frac{5!}{1!3!} x(1-x)^3 = 20x(1-x)^3$$

Acceptance-Rejection

To find c:

$$c = \max_{x \in (0,1)} \left(\frac{f(x)}{g(x)} \right) = \max_{x \in (0,1)} (20x(1-x)^3)$$

Let $h(x) = f(x)/g(x)$. Then,

$$h'(x) = 20(1-x)^3 - 60x(1-x)^2 = (1-x)^2(20-80x) \overset{\text{set}}{=} 0 \quad \Rightarrow \quad x = 1, 1/4$$

Test values:

$$h(0) = 0, \;\; h(1) = 0, \;\; h(1/4) = 135/64 = 2.109375$$

Hence, $c = 2.109375$.

Acceptance-Rejection

For each random variate required,

1. Generate $Y \sim \mathit{Uniform}(0,1)$.
2. Generate $U \sim \mathit{Uniform}(0,1)$.
3. If

$$u < \frac{f(y)}{cg(y)} \approx \frac{20y(1-y)^3}{2.11}$$

accept y and set $x = y$. Otherwise, reject y and generate a random variate again.

Acceptance-Rejection

Example

Use the acceptance-rejection method to simulate a random sample from the distribution with density:

$$f_X(x) = \begin{cases} \frac{2}{\pi}\sqrt{1-x^2} & |x| < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Let the trial density be *Uniform*$(-1, 1)$. Write the *R* code to simulate the algorithm. Compare the number of iterations to the value of $\frac{1}{c}$.

Solution.

$$g(x) = \frac{1}{1-(-1)} = \frac{1}{2}, \quad -1 < x < 1.$$

To find c:

$$c = \max_{x \in (0,1)} \left(\frac{f(x)}{g(x)} \right) = \max_{x \in (0,1)} \left(\frac{4}{\pi}\sqrt{1-x^2} \right)$$

Acceptance-Rejection

Let $h(x) = f(x)/g(x)$. Then,

$$h'(x) = \frac{-4}{\pi} \frac{x}{\sqrt{1-x^2}} \overset{\text{set}}{=} 0 \quad \Rightarrow \quad x = 0$$

Test values:

$$h(1) = 0, \;\; h(-1) = 0, \;\; h(0) = \frac{4}{\pi}$$

Hence $c = \frac{4}{\pi}$. For each random variate required,

1. Generate $Y \sim Uniform(-1, 1)$.
2. Generate $U \sim Uniform(0, 1)$.
3. If

$$u < \frac{f(y)}{cg(y)} = \frac{\pi}{4}\left(\frac{4}{\pi}\sqrt{1-y^2}\right) = \sqrt{1-y^2}$$

   accept y and set $x = y$. Otherwise, reject y and generate a random variate again.

Transformation and Convolution Methods

Transformation and Convolution Methods

You've done transformations or convolution methods before (or at least I hope!)

- If $X_i \sim Exponential(a)$ with $i \in \{1, 2, ..., n\}$ then $\sum_i X_i \sim Gamma(n, a)$.
- If $U \sim Gamma(r, \lambda)$ and $V \sim Gamma(s, \lambda)$ and $U \perp\!\!\!\perp V$ then

$$X := \frac{U}{U+V} \sim Beta(r, s).$$

- If $Z \sim N(0, 1)$ then $V := Z^2 \sim \chi^2(1)$.
- If $U \sim \chi^2(m)$ and $V \sim \chi^2(n)$ and $U \perp\!\!\!\perp V$ then

$$\tilde{F} := \frac{U/m}{V/n} \sim F(m, n).$$

- If $Z \sim N(0, 1)$ and $V \sim \chi^2(n)$ and $Z \perp\!\!\!\perp V$ then:

$$\tilde{T} := \frac{Z}{\sqrt{V/n}} \sim t(n).$$

Example

Use `rexp()` to generate the $Gamma(\alpha = 10, \beta = 1/2)$ distribution.

Solution. Let $X_i \sim Exp(1/2)$. Then, $\sum_{i=1}^{10} X_i \sim Gamma(10, 1/2)$.
For each random variate required,

1. Generate $X_i \sim Exp(1/2)$ for $i = 1, 2, \ldots, 10$.
2. Deliver the sum of what was generated, i.e., $\sum_{i=1}^{10} X_i$.

Example

Use `rexp()` to generate the $Beta(\alpha = 2, \beta = 3)$ distribution.

Solution. Consider $X_i \sim Exp(1)$. Then, $U := \sum_{i=1}^{2} X_i \sim Gamma(2, 1)$ and $V := \sum_{i=3}^{5} X_i \sim Gamma(3, 1)$. Furthermore, we have:

$$\frac{U}{U + V} \sim Beta(2, 3)$$

Note: in this case, the choice of scale for the exponential was arbitrary; they just needed to be consistent.

For each random variate required, and let the choice of λ will be arbitrary.

1. Generate $X_i \sim Exp(\lambda)$ for $i = 1, 2, \ldots, 5$.
2. Let $U = \sum_{i=1}^{2} X_i$ and $V = \sum_{i=3}^{5} X_i$.
3. Deliver $\frac{U}{U+V}$.

Mixture Methods

Finite Mixture Model

A finite mixture model is a statistical model that represents a probability distribution as a mixture of several component distributions. Mathematically, given k component distributions $f_1(x), ..., f_k(x)$, each with associated mixing probabilities (also known mixing weights) $\pi_1, \dots, \pi_k$, a finite mixture model $f(x)$ is defined as:

$$f(x) := \sum_{i=1}^{k} \pi_i f_i(x),$$

where $0 \leq \pi_i \leq 1$ and $\sum_{i=1}^{k} \pi_i = 1$.

Mixture Model Properties

We want to show that $f(x)$ satisfies the properties of a density function. That is, show that $f(x)$ is non-negative and $\int_{-\infty}^{\infty} f(x)dx = 1$.

Solution. By assumption, since we have that:

$$0 \leq \pi_i, \ \ 0 \leq f_i(x), \ \ \forall i \in \{1, 2, \ldots, n\},$$
$$\Rightarrow \ \ 0 \leq \pi_i f_i(x) \ \ \forall i \in \{1, 2, \ldots, n\},$$
$$\Rightarrow 0 \leq \pi_1 f_1(x) + \pi_2 f_2(x) + \cdots + \pi_n f_n(x) = \sum_{i=1}^{n} \pi_i f_i(x)$$

Now, to show that this integrates to 1:

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{\infty} \sum_{i=1}^{n} \pi_i f_i(x)dx = \sum_{i=1}^{n} \int_{-\infty}^{\infty} \pi_i f_i(x)dx$$

Where the previous line is us applying the linearity property of the integral...

Mixture Models

$$
\begin{aligned}
\sum_{i=1}^{n} \int_{-\infty}^{\infty} \pi_i f_i(x) dx &= \sum_{i=1}^{n} \pi_i \underbrace{\int_{-\infty}^{\infty} f_i(x) dx}_{=1} \\
&= \sum_{i=1}^{n} \pi_i \\
&= 1 \quad \text{(By assumption.)}
\end{aligned}
$$

Mixture Model CDF

Compute $F(x)$.

Solution.

$$
\begin{aligned}
F(x) &= \int_{-\infty}^{x} \sum_{i=1}^{n} \pi_i f_i(t) dt \\
&= \sum_{i=1}^{n} \pi_i \int_{-\infty}^{x} f_i(t) dt \\
&= \sum_{i=1}^{n} \pi_i F_i(x)
\end{aligned}
$$

Mixture Model Expected Value

Let X_i have density f_i. Denote $\mu_i := \mathbb{E}[X_i] = \int_{-\infty}^{\infty} x f_i(x) dx$; this represents the mean for the i-th component distribution. Compute $\mu := \mathbb{E}[X]$. What does it represent?

Solution.

$$\begin{aligned}
\mathbb{E}[X] &= \int_{-\infty}^{\infty} x \sum_{i=1}^{k} \pi_i f_i(x) dx \\
&= \sum_{i=1}^{k} \pi_i \int_{-\infty}^{\infty} x f_i(x) dx \\
&= \sum_{i=1}^{k} \pi_i \mu_i
\end{aligned}$$

Hence, this represents the weighted average of the means of the individual components.

Mixture Model Variance

Let $\sigma_i^2 := \mathbb{V}[X_i]$ represent the standard deviation for the i-th component distribution. Show that:

$$\sigma^2 := \mathbb{V}[X] = \sum_{i=1}^{k} \pi_i \sigma_i^2 + \sum_{i=1}^{k} \pi_i (\mu_i - \mu)^2$$

What does it represent?

Solution.

$$\begin{aligned}
\mathbb{V}[X] &= \mathbb{E}[(X-\mu)^2] \\
&= \int_{-\infty}^{\infty} (x-\mu)^2 \sum_{i=1}^{k} \pi_i f_i(x) dx \\
&= \sum_{i=1}^{k} \pi_i \int_{-\infty}^{\infty} (x^2 - 2x\mu + \mu^2) f_i(x) dx
\end{aligned}$$

Mixture Models

$$
\begin{aligned}
\mathbb{V}[X] &= \sum_{i=1}^{k} \pi_i \left[\int_{-\infty}^{\infty} x^2 f_i(x)dx - 2\mu \int_{-\infty}^{\infty} x f_i(x)dx + \mu^2 \int_{-\infty}^{\infty} f_i(x)dx \right] \\
&= \sum_{i=1}^{k} \pi_i \left[\mathbb{E}[X_i^2] - 2\mu\mu_i + \mu^2 \right] = \sum_{i=1}^{k} \pi_i \left[\mathbb{V}[X_i] + (\mathbb{E}[X_i])^2 - 2\mu\mu_i + \mu^2 \right] \\
&= \sum_{i=1}^{k} \pi_i \left[\sigma_i^2 + \mu_i^2 - 2\mu\mu_i + \mu^2 \right] \\
&= \sum_{i=1}^{k} \pi_i \left[\sigma_i^2 + (\mu_i - \mu)^2 \right] = \sum_{i=1}^{k} \pi_i \sigma_i^2 + \sum_{i=1}^{k} \pi_i (\mu_i - \mu)^2
\end{aligned}
$$

This represents a weighted average of the variances of the individual components, along with the weighted variance of the component means.

Mixture Models

Convolutions and mixtures look similar but the represented distributions differ!

Example

Suppose $X_1 \sim N(0,1)$, $X_2 \sim N(3,1)$, and $X_1 \perp\!\!\!\perp X_2$.

- Convolution representation:
$$S := 0.4X_1 + 0.6X_2$$
- Mixture representation:
$$F_X(x) := 0.4F_{X_1}(x_1) + 0.6F_{X_2}(x_2)$$

Note: for the convolution, the coefficients in front of the random variables do not necessarily have to add to one; we could have done $S := aX_1 + bX_2$ for any $a, b \in \mathbb{R}$. However, for finite mixture models, these coefficients must add to 1.

Mixture Models

Simulating a variable from a finite k-mixture distribution is typically carried out by the composition method: Consider $F(x) = \sum_{i=1}^{k} \pi_i F_{X_i}(x)$. Then,

Composition Method

- Generate an integer $I \in \{1, \ldots, k\}$ such that:

$$\mathbb{P}(I = i) = \pi_i, \quad i \in \{1, 2, \ldots, k\}.$$

- Deliver X with cdf $F_{X_I}(x)$.

Mixture Models

Example

Suppose $X_1 \sim N(0,1)$, $X_2 \sim N(3,1)$, and $X_1 \perp\!\!\!\perp X_2$. Simulating the following using R:

$$F_X(x) := 0.4F_{X_1}(x_1) + 0.6F_{X_2}(x_2)$$

Then, compare the above to the following convolution counterpart:

$$S := 0.4X_1 + 0.6X_2$$

Solution. Consult the R code.

Mixture Models

The first two methods we discussed, inverse-transform and acceptance-rejection method, depend on the uniform distribution. Similarly, we can do the same for generating finite mixture models. Consider $F(x) = \sum_{i=1}^{k} \pi_i F_{X_i}(x)$. Then:

Modified Composition Method

- Generate u from the *Uniform*$(0, 1)$ distribution.
- If:

$$\sum_{i=1}^{I-1} \pi_i \leq u < \sum_{i=1}^{I} \pi_i$$

  then generate a random x from $F_{X_I}(x)$ where $I = 1, \ldots, k$ with the convention $\sum_{i=1}^{0} \pi_i = 0$.

Mixture Models

Example

Using the alternative method... Suppose $X_1 \sim N(0,1)$, $X_2 \sim N(3,1), X_3 \sim N(5,1)$, and assume they're all independent of each other. Code the following algorithm using R:

$$F_X(x) := 0.4F_{X_1}(x_1) + 0.3F_{X_2}(x_2) + 0.3F_{X_3}(x_3)$$

Solution. Consult the R code.