Generating Random Variates

Max Turgeon

STAT 3150–Statistical Computing

Lecture Objectives

- Recognize when to use the inverse-transform method.
- Be able to generate random variates through transformations.
- Derive bounding densities for accept-reject sampling.

Motivation

- A staple of modern statistical research is the **simulation study**.
    - Finite sample properties can then be compared to theoretical expectations.
- More generally, by simulating data we can study the properties of a method or a model.
- **Bayesian statistics** strongly relies on generating data to estimate the posterior density of the parameters (cf. STAT 4150).

Inverse-Transform Method i

- *Recall*: Let X be a random variable with CDF $F(x)$. The **quantile** function is defined as

$$F^{-1}(p) = \inf\{x \in \mathbb{R} \mid F(x) \geq p\}.$$

- If X is continuous, this is simply the inverse function.

Inverse-Transform Method ii

Theorem
If U is uniform on $[0,1]$, then $F^{-1}(U)$ has the same distribution as X.

- In R, we can sample random variates from $U(0,1)$ by using the function `runif`:

```r
runif(5)
```

```
## [1] 0.2681359 0.3308333 0.4411671 0.8352923
0.9690489
```

Inverse-Transform Method iii

Algorithm
To generate random variates from F:

1. Generate random variates from $U(0,1)$.
2. Compute the quantile function F^{-1}.
3. Plug-in the uniform variates into F^{-1}.

Example i

- Let X follow an exponential distribution with parameter λ:

$$F(x) = 1 - \exp(-\lambda x).$$

- Since X is continuous, the quantile function is the inverse of F:

$$\begin{aligned} p = 1 - \exp(-\lambda x) &\Rightarrow \exp(-\lambda x) = 1 - p \\ &\Rightarrow -\lambda x = \log(1 - p) \\ &\Rightarrow x = \frac{-\log(1 - p)}{\lambda}. \end{aligned}$$

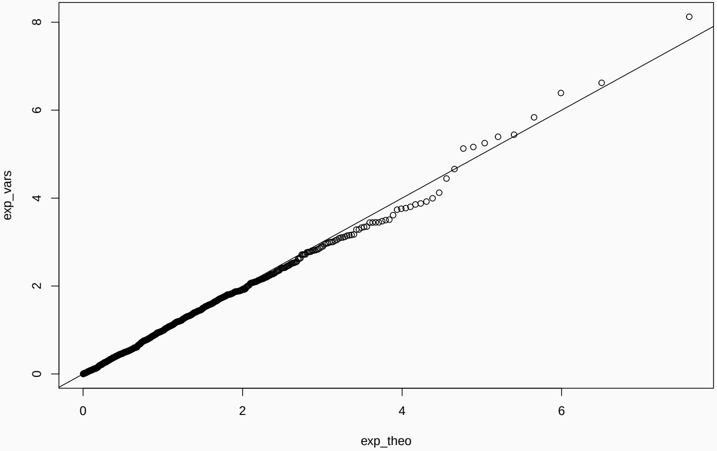
Example ii

```r
lambda <- 1
# We want 1000 samples
n <- 1000
unif_vars <- runif(1000)
exp_vars <- -log(1 - unif_vars)/lambda
```

Example iii

```r
# Compute theoretical quantiles
# using qexp
exp_theo <- qexp(ppoints(n))
qqplot(exp_theo, exp_vars)
# Add diagonal line
abline(a = 0, b = 1)
```

Example iv



Example v

Note: If U is uniform on $[0, 1]$, so is $1 - U$.

- Therefore $\frac{-\log(U)}{\lambda}$ also follows an $Exp(\lambda)$ distribution.

Inverse Transform—Discrete Edition

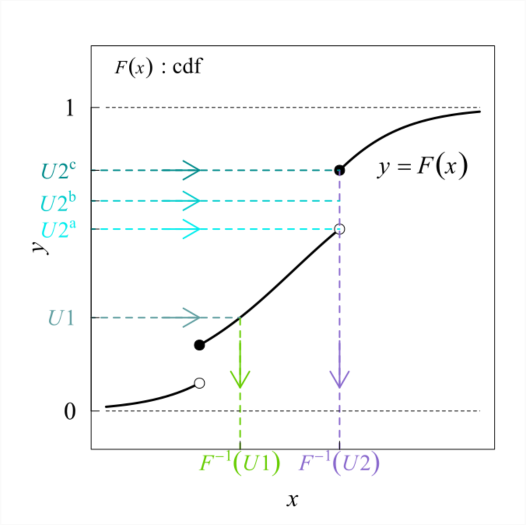


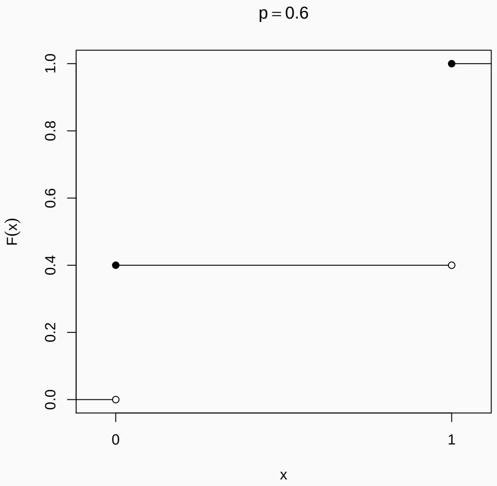
Figure 1: From Wikipedia

Example i

- Let X follow a Bernoulli distribution with parameter p:

$$F(x) = \begin{cases} 0 & x < 0, \\ 1 - p & x \in [0, 1), \\ 1 & x \geq 1. \end{cases}$$

Example ii



Example iii

- As we can see, we have

$$F^{-1}(u) = \begin{cases} 0 & u \leq 1 - p, \\ 1 & u > 1 - p. \end{cases}$$

- In other words, we sample U. If it is less than p, we set $X = 0$; else, we set $X = 1$.

Example iv

```r
p <- 0.6
n <- 1000
unif_vars <- runif(1000)
# as.numeric turns FALSE into 0
# and TRUE into 1
bern_vars <- as.numeric(unif_vars > 1 - p)
```

```r
c(mean(bern_vars), var(bern_vars))
```

```
## [1] 0.6400000 0.2306306
```

Example v

```r
# Compare with theory
c(p, p*(1 - p))
```

```
## [1] 0.60 0.24
```

More General Transformations

- Inverse transform is just one type of transformation!
- We can use relationships between distributions to generate random variates. For example:
    - If $Z \sim N(0,1)$, then $Z^2 \sim \chi^2(1)$.
    - If $V_1, \ldots, V_p \sim \chi^2(1)$, then $\sum_{i=1}^p \sim \chi^2(p)$.
    - If $U \sim \chi^2(p)$ and $V \sim \chi^2(q)$, then

$$\frac{U/p}{V/q} \sim F(p,q).$$

Example i

```r
# Choose degrees of freedom
p <- 2
q <- 4

# rnorm samples from a normal distribution
U <- sum(rnorm(p)^2)
V <- sum(rnorm(q)^2)

# Take ratio
(U/p)/(V/q)
```

Example ii

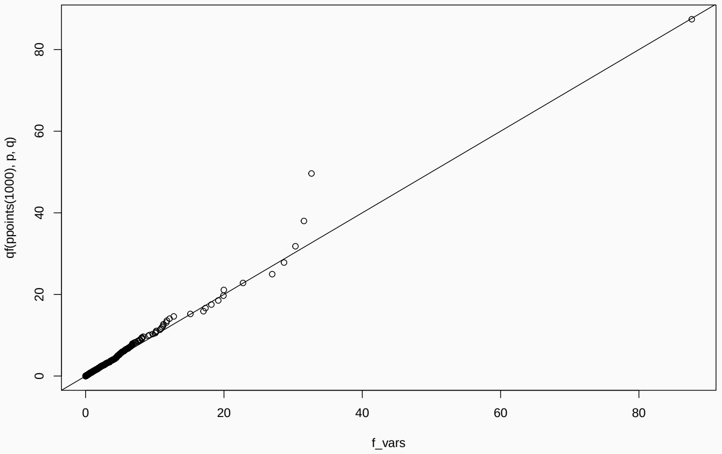
```
## [1] 3.389027
```

```
# What if we want 1000 replicates?
# Use the function replicate!
# First argument: number of replicates
# Second argument: expression to be run multiple times
f_vars <- replicate(1000, {
  U <- sum(rnorm(p)^2)
  V <- sum(rnorm(q)^2)
  (U/p)/(V/q)
})
```

Example iii

```r
qqplot(f_vars, qf(ppoints(1000), p, q))
# Add diagonal line
abline(a = 0, b = 1)
```

Example iv



Mixture of distributions i

- We say X with distribution function F follows a (discrete) **mixture of distributions** if we can write

$$F(x) = \sum_{k=1}^{K} \pi_k F_k(x),$$

  where F_k is a distribution function for all k and $\sum_{k=1}^{K} \pi_k = 1$.
    - For example, the distribution of adult heights can be well-approximated by a mixture of two normal distributions with equal weights.

Mixture of distributions ii

Algorithm
To sample from $F = \sum_{k=1}^{K} \pi_k F_k$:

1. Select a component k by sampling from a multinomial distribution with probabilities $(\pi_1, \ldots, \pi_K)$.
2. Sample from F_k.

Example i

- Assume that adult heights for **male** follows a normal distribution with mean 175cm and standard deviation 7cm.
- Assume that adult heights for **female** follows a normal distribution with mean 160cm and standard deviation 7cm.
- Finally assume that we have an equal proportion of male and female.
- How can we sample from the distribution of adult heights?

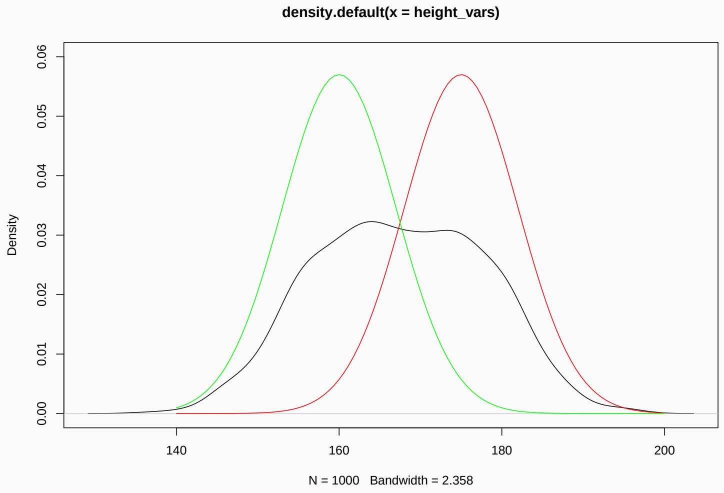
Example ii

```r
height_vars <- replicate(1000, {
  sex <- sample(c(”M”, ”F”), size = 1,
                prob = c(0.5, 0.5))
  if (sex == ”M”) {
    height <- rnorm(1, mean = 175, sd = 7)
  }
  if (sex == ”F”) {
    height <- rnorm(1, mean = 160, sd = 7)
  }
  return(height)
})
```

Example iii

```r
plot(density(height_vars), ylim = c(0, 0.06))
# Add component densities
x_seq <- seq(140, 200, length.out = 100)
lines(x_seq, dnorm(x_seq, 175, 7), col = ”red”)
lines(x_seq, dnorm(x_seq, 160, 7), col = ”green”)
```

Example iv


density.default(x = height_vars)
Density
0.00
0.01
0.02
0.03
0.04
0.05
0.06
140
160
180
200
N = 1000   Bandwidth = 2.358

Example v

```r
# Sample with less code
sex <- sample(c(1, 2), size = 1000,
              replace = TRUE,
              prob = c(0.5, 0.5))
height_vars <- rnorm(1000, c(175, 160)[sex], 7)
c(mean(height_vars), sd(height_vars))
```

```
## [1] 167.8058  10.3375
```

Summary

- When we can compute the quantile function, the inverse transform is simple to implement.
    - But it can be hard to compute!
- We can leverage relationships between distributions to transform one random variate into another.
- **Mixture distributions** are a flexible way to model heterogeneous data, and it's "easy" to sample from them.
    - **Note**: Do not confuse *convolutions* (i.e. adding random variables) with *mixtures* (i.e. convex combination of distribution functions).

Acceptance-Reject Method i

- Suppose you want to sample from a distribution X with density f, but you can only sample from a different distribution Y with density g.
- Further suppose that there exists a constant $c > 1$ such that

$$\frac{f(t)}{g(t)} \leq c$$

  for all t such that $f(t) > 0$.
- The *Acceptance-Reject method* is a way to transform random variates of Y into random variates of X.

Acceptance-Reject Method ii

Algorithm

1. Sample y from Y.
2. Sample a uniform variate u from $U(0,1)$.
3. Compute the ratio $r := \frac{f(y)}{cg(y)}$. If $u < r$, set $x = y$. Otherwise, reject y and repeat from Step 1.

Note: The number of iterations before we accept a draw from Y follows a geometric distribution with mean c. So we want the constant c to be as small as possible.

(If you want a proof of why this works, see UM Learn.)

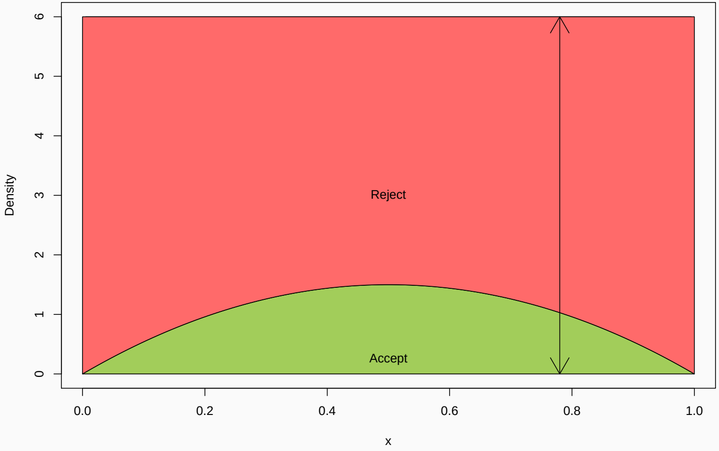
Example i

- We want to sample from $X \sim Beta(2,2)$ whose density is $f(x) = 6x(1-x)$.
    - The *proposal* distribution will be $Y \sim Beta(1,1)$ (i.e. a uniform distribution).
- Let $t \in (0,1)$. We have

$$\frac{f(t)}{g(t)} = \frac{6t(1-t)}{1} \leq 6,$$

since the maximum t and $1-t$ can take is 1. So we can set $c = 6$.

Example ii



Example iii

```r
# Set parameters----
C <- 6 # Constant
n <- 1000 # Number of variates
k <- 0 # counter for accepted
j <- 0 # iterations
y <- numeric(n) # Allocate memory
```

Example iv

```r
# A while loop runs until condition no longer holds
while (k < n) {
  u <- runif(1)
  j <- j + 1
  x <- runif(1) # random variate from g
  if (u < 6*x*(1-x)/C) {
    k <- k + 1
    y[k] <- x
    }
}
```

Example v

```r
# How many iterations did we need?
j
```

```
## [1] 6144
```

```r
# Compare theoretical and empirical quantiles
p <- seq(0.1, 0.9, by = 0.1)
Qhat <- quantile(y, p) # empirical
Q <- qbeta(p, 2, 2) # theoretical
```

Example vi

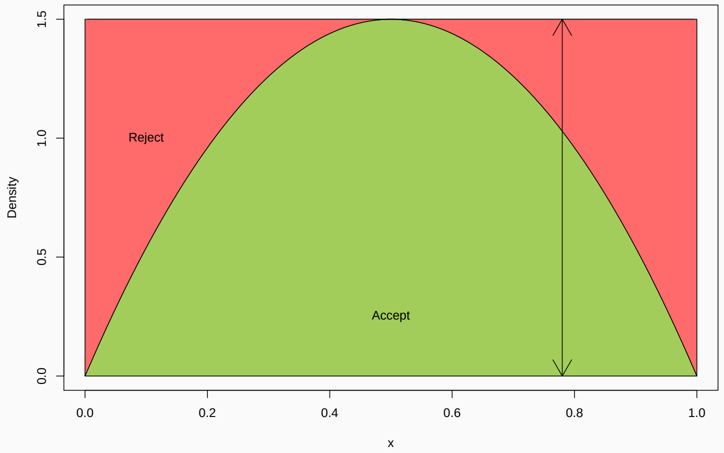
```r
round(cbind(Qhat, Q, diff = abs(Qhat - Q)), 3)
```

```
##      Qhat     Q  diff
## 10% 0.181 0.196 0.014
## 20% 0.262 0.287 0.025
## 30% 0.341 0.363 0.022
## 40% 0.409 0.433 0.024
## 50% 0.477 0.500 0.023
## 60% 0.558 0.567 0.009
## 70% 0.637 0.637 0.000
## 80% 0.715 0.713 0.002
## 90% 0.793 0.804 0.011
```

Example vii

- As the graph showed, the “Rejection” region is very large.
    - In fact, it is unnecessarily large.
- With a little bit of calculus, we can show that the maximum value of $6x(1-x)$ is 1.5.
    - In other words, we can set the constant $c = 1.5$.
    - This means that we can sample from X while rejecting 4 times *less* often.

Example viii



Example ix

```r
C <- 1.5; k <- j <- 0 # Reset counters
while (k < n) {
  u <- runif(1)
  j <- j + 1
  x <- runif(1)
  if (u < 6*x*(1-x)/C) {
    k <- k + 1
    y[k] <- x
    }
}
```

Example x

```
# How many iterations did we need this time?
j
```

```
## [1] 1517
```