STATS 406 F15: Lab 04

1 Random number generation: beyond basic generators

1.1 Method 1: Inversion method

Gists:

1. Requires an explicit CDF expression.
2. More efficient than rejection sampling.

Formulation:

- R.v. $X \overset{\text{CDF}}{\sim} F(x)$, then $F(X) \sim \text{Unif}(0,1)$ and $F^{-1}(U) \overset{\text{CDF}}{\sim} F(x)$ for $U \sim \text{Unif}(0,1)$.

  Proof for continuous distributions. For any t in the sample space of X, we have

  $$\mathbb{P}(F^{-1}(U) < t) = \mathbb{P}(F(F^{-1}(U)) < F(t)) = \mathbb{P}(U < F(t)) = F(t)$$

  So $F^{-1}(U) \overset{\text{CDF}}{\sim} F(x)$. □

Examples:

(a) $\text{Exp}(\lambda)$ distribution, CDF:

$$F(x) = 1 - e^{-\lambda x}, \lambda > 0$$

```r
# Inversion method: exponential r.v.s
set.seed(2015);
n = 10000;
lambda = 3;
x = -(1/lambda)*log(runif(n));
hist(x, prob = TRUE, main="Inversion method: exponential r.v.s");
## (Recap last lab session: low-level plot)
x_dist = seq(0, 10, length.out = 1000);
lines(x=x_dist, y=dexp(x_dist, lambda), col='blue');
```

(b) Consequently, we know how to sample from a Gamma distribution with integer shape parameter(also called "Erlang distribution"): if $X_1, \ldots, X_k \overset{\text{iid}}{\sim} \text{Exp}(\lambda)$, then $X_1 + \ldots + X_k \overset{\text{iid}}{\sim} \text{Gamma}(k, \lambda)$.

(Code omitted)

(c) $\text{Pareto}(x_m, \alpha)$ distribution, CDF(following the notation in Wikipedia):

$$F(x; x_m, \alpha) = 1 - (x_m/x)^\alpha$$

where x_m is *scale* and α is *shape*. Sometimes there is a *location* that controlls where the domain starts(at 1 in Wikipedia and at 0 in R). Here we set domains to start at 0 following R.

We first preview the distribution to get an impression.

```
# Preview the distribtion
## (Study the code below as a good review of last lab session on plotting)
x_preview = seq(from=0.05, to=3, by=0.05);
matplot(x=x_preview,
    y=cbind(dpareto(x_preview, shape=1, scale=1),
        dpareto(x_preview, shape=3, scale=1),
        dpareto(x_preview, shape=5, scale=1)),
    col=1:3, lty=1, type='l',
    xlab='x', ylab='density', main='PDF for Pareto distribution, scale=1',
    cex.lab=1.5, cex.axis=1.5, cex.main=1.5);
legend('topright', legend=c('shape=1', 'shape=3', 'shape=5'), col=1:3, lty=1, cex=1.5);
```

Then we draw the sample and plot its histogram.

```
# Sample from a Pareto distribution
set.seed(2015);
n = 1000000;
x_m=1; alpha=3;
x = x_m*(1-runif(n))^(-1/alpha) - 1; # minus one to set location so that domain starts at 0
## (Recap the usage of paste() from last lab session)
hist(x, probability=TRUE, breaks=5000, xlim=c(0, 5),
    main=paste('Inversion method: Pareto r.v.s, shape=', alpha, ', scale=', x_m, sep=''));
x_dist = seq(0.01, 3, length.out = 1000);
lines(x=x_dist, y=dpareto(x_dist, shape=alpha, scale=x_m), col='blue');
```

(d) Multinomial distribution on $a_1, \ldots, a_m$, CDF:

$$\mathbb{P}(X \in \{a_1, \ldots, a_k\}) = p_1 + \ldots + p_k, \quad \text{for } 1 \leq k \leq m$$

(Illustration on the blackboard: understanding the inversion method here.) "Inverse function":

$$\text{“}F^{-1}(x)\text{”} = a_k, \quad \text{for } p_0 + \ldots + p_{k-1} \leq x < p_0 + \ldots + p_k, \text{ where } 1 \leq k \leq m \text{ and } p_0 := 0$$

(Code omitted)

1.2 Method 2: Rejection sampling

Know how to sample from distribution $\mathcal{F}_1$ with PDF $f_1(t)$, need to sample from $\mathcal{F}_2$ with PDF $f_2(t)$ dominated **pointwise** by $f_1(t)$ up to a constant, i.e.

$$f_2(t) \leq C f_1(t)$$

for some constant $C > 0$.

Gists:

1. The domination must be **pointwise**, so it is required that the domain of $\mathcal{F}_1$ contains that of $\mathcal{F}_2$.
2. Although for one $f_2(t)$ there can be many useable $\{f_1(t), C\}$ combinations, there are good and bad ones among them. Find $f_1(t)$ that mimics the shape of $f_2(t)$ as much as possible, so that you generate more candidate random numbers around the high-probability regions under $f_2(t)$.
   A similar idea will be seen in *importance sampling*.

Illustration:

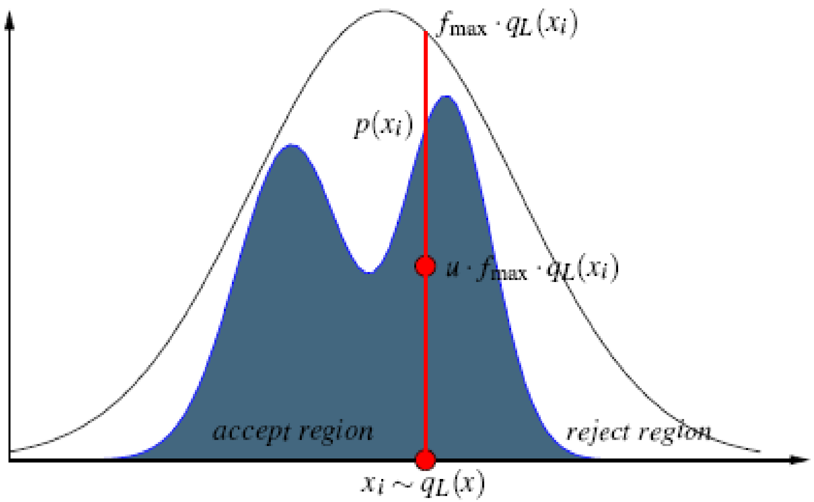
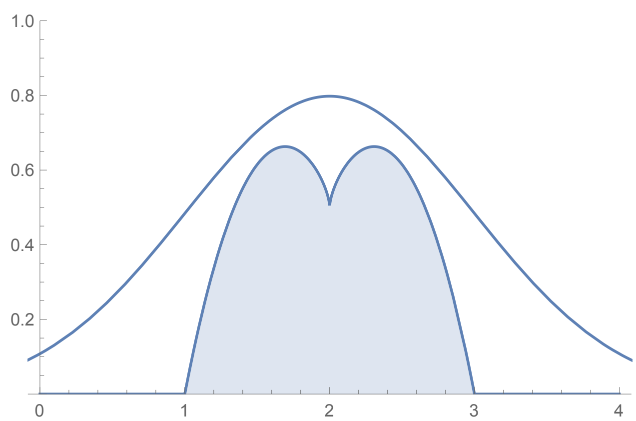


Figure 1: Importance sampling

Figure 1 source: `http://www.cs.ubc.ca/~heidrich/Projects/SamplingDirectIllumination/rj.png`

Example:

- Distribution to be sampled: $f_2(t) \propto -(x-1)(x-3)\left(|x-2|^{2/3}+1\right)$.
- Use $f_1(t) = 2 \times \text{PDF}\left(N(2,1);t\right)$ to envolope it.




```
# Example: see Lab_4.r
```

1.3 Method 3: Linear transformation(for multivariate normal distributions)

- Fact: if $\mathbf{X} \sim N(\mathbf{0}, \boldsymbol{I})$(standard multivariate normal distribution), then $\boldsymbol{\Sigma}^{1/2}\mathbf{X} + \boldsymbol{\mu} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ for semi-positive definite $\boldsymbol{\Sigma}$.
- The availability of other types of transformations of other distributions depends on the specific contexts. No universal guidance on "when it is possible" and "how". Analyze case-by-case.

```
# Example: see Lab_4.r
```

2 Briefly more on LLN: Empirical Distribution Functions

For a sample $X_1, \cdots, X_n \overset{\text{iid}}{\sim} N(0,1)$. The empirical CDF is:

$$\hat{F}(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(X_i \leq x)$$

Vary n and plot.

```
# Animated example: see Lab_4.r
```