Generating Random Numbers

**Aim:** produce random variables for given distribution

**Inverse Method**

Let $F$ be the distribution function of an univariate distribution and let $F^{-1}(y) = \inf\{x|F(x) \geq y\}$ (generalized inverse of $F$).

For a uniformly distributed random variable $U \sim U(0, 1)$, let

$$X = F^{-1}(U).$$

Then $X$ has distribution function $F$:

$$P(X \leq x) = P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x).$$

*Example:*

For the exponential distribution, we have $F(x) = 1 - \exp(-\theta x)$. Thus

$$X = F^{-1}(U) = -\frac{1}{\theta} \log(1 - U)$$

is exponentially distributed with parameter $\theta$.

*Remarks:*

The result shows that the generation of sequences of random variables (usually iid) for some given distribution is based on the production of uniform random variables.

Since the algorithms for generating uniform random variables are deterministic, we call them *pseudo random number generators*.

*References:*

Generating Random Numbers

Acceptance-Rejection Method
Let $f$ be the density of some univariate distribution we want to sample from. Suppose the $f(\cdot)$ is majorized by $c g(\cdot)$,

$$f(y) \leq c g(y),$$

for some simple probability density $g$ and some constant $c > 1$.
The idea of acceptance-rejection is to sample proposals $X$ from the simpler (simulationwise) density $g$ and then to reject some proposals which are likely to be overrepresented in the sample:

- Sample $X \sim g(x)$
- Sample $U \sim U(0, 1)$
- Accept $X$ if $U \leq \frac{f(X)}{c g(X)}$.

**Example:** Non-standard prior for binomial parameter
Suppose that $Y$ is binomially distributed

$$Y \sim \text{Bin}(n, \theta)$$

with non-conjugate prior distribution

$$\pi(\theta) = 4 \left( \frac{1}{2} - |\theta - \frac{1}{2}| \right), \quad \theta \in [0, 1].$$

The posterior distribution (conditional on data $Y = y$) satisfies

$$\pi(\theta|y) \sim \theta^y (1 - \theta)^{n-y} \left( \frac{1}{2} - |\theta - \frac{1}{2}| \right), \quad \theta \in [0, 1].$$
Generating Random Numbers

**Example:** Non-standard prior for binomial parameter (contd)

Since \( \frac{1}{2} - |\theta - \frac{1}{2}| \leq \frac{1}{2} - 2 (\theta - \frac{1}{2})^2 = 2 \theta (1 - \theta) \) we get for some \( c > 0 \)

\[
\pi(\theta | y) \leq c \theta^{y+1} (1 - \theta)^{n-y+1}
\]

The right side is proportional to the density of a beta distribution with parameters \( y + 2 \) and \( n - y + 2 \). This suggests the following sampling scheme:

- Define \( f(\theta) = \theta^y (1 - \theta)^{n-y} \left( \frac{1}{2} - |\theta - \frac{1}{2}| \right) \)
  \[ g(\theta) = 2 \theta^{y+1} (1 - \theta)^{n-y+1} \]

- Sample \( \theta \sim \text{Beta}(y + 2, n - y + 2) \)
  \[ U \sim \text{U}(0, 1) \]

- Accept \( \theta \) if \( U \leq \frac{f(\theta)}{g(\theta)} \).

- *Note:* We did not need to compute the proportionality factor
  \[
  \int_{\Theta} f(y | \theta) \pi(\theta) \, d\theta,
  \]
  since this cancels in the ratio \( \frac{f(\theta)}{g(\theta)} \).
Generating Random Numbers

For other distributions (e.g. the normal) exist special methods.

In this course, we will only use the built-in generators in $R$:

- $\mathcal{N}(\mu, \sigma^2)$: \texttt{rnorm(n,mean=0,sd=1)}
- $U[\text{min}, \text{max}]$: \texttt{runif(n,min=0,max=1)}
- Beta($a, b$): \texttt{rbeta(n,a,b)}
- Bin($s, p$): \texttt{rbinom(n,size,prob)}
- Cauchy($\alpha, \sigma$): \texttt{rnorm(n,loc=0,scale=1)}
- $\chi^2_n(\delta)$: \texttt{rchisq(n,df,ncp=0)}
- Exp($\text{rate}$): \texttt{rnorm(n,rate=1)}
- $F_{m,n}$: \texttt{rf(n,df1,df2)}
- $\Gamma(a, s)$: \texttt{rgamma(n,shape,rate=1,rate})
- Geom($p$): \texttt{rgeom(n,prob)}
- $H(m, n, k)$: \texttt{rhyper(nn,m,n,k)}
- log-normal($\mu, \sigma^2$): \texttt{rlnorm(n,mean=0,sd=1)}
- Logistic($\mu, \sigma^2$): \texttt{rlogis(n,loc=0,scale=1)}
- NegBinom($s, p$): \texttt{rnbinom(n,size,prob,mu)}
- Poisson($\lambda$): \texttt{rpois(n,lambda)}
- $t_n$: \texttt{rt(n,df)}
- Weibull($a, b$): \texttt{rweibull(n,shape,scale=1)}

$R$ also provides functions for calculating the density (dF), the distribution function (pF) and quantiles (qF), where F is the name of the distribution as in the above commands.
Monte Carlo Methods

**Aim:** Evaluate expectation

\[
E(h(Y)) = \int h(y) f(y) \, dy, \quad (1)
\]

where \( Y \) is some (possibly high-dimensional) random variable with distribution defined by \( f(y) \).

**Examples:**

- Suppose \( \hat{\theta} = \hat{\theta}(Y) \) is an estimator for some parameter \( \theta \). Quantities of interest are the bias and the standard deviation of \( \hat{\theta} \),

\[
\text{bias}(\hat{\theta}) = E(\hat{\theta}) - \theta,
\]

\[
\sigma(\hat{\theta}) = \left( E(\hat{\theta} - E(\hat{\theta}))^2 \right)^{\frac{1}{2}}.
\]

- In cases, where the estimate \( \hat{\theta} \) is obtained by an iterative estimation procedures (e.g. by Newton-Raphson), the estimator \( \hat{\theta}(Y) \) cannot be written in closed form and the integral (1) cannot be computed by numerical integration.

- For two normally distributed samples \( Y_{11}, \ldots, Y_{n_11} \) and \( Y_{12}, \ldots, Y_{n_22} \), the hypothesis of equal means can be tested by the two-sample \( t \) test with test statistic

\[
T = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}
\]

The \( p \)-value of the test is defined as

\[
\mathbb{P}(T(Y) > t) = E((1_{(t, \infty)}(T(Y))).
\]

To see how the \( t \) test performs under departures from normality, we need to evaluate the \( p \)-value (or equivalently the significance level) of the test for nonnormal distributions \( f(y) \).
Monte Carlo Methods

Monte Carlo approach:

- Draw sample $Y^{(1)}, \ldots, Y^{(n)} \overset{\text{iid}}{\sim} f(y)$.
- Estimate expectation by

$$\frac{1}{n} \sum_{t=1}^{n} h(Y^{(t)})$$

(Monte Carlo integration)

- For independent samples, the Law of Large Numbers yields

$$\frac{1}{n} \sum_{t=1}^{n} h(Y^{(t)}) \to \mathbb{E}(h(Y)) \quad \text{as } n \to \infty.$$ 

Example: Two-sample $t$ test in $R$

```r
N<-10000 #number of MC repititions
n1<-8 #sample size
n2<-4
Y1<-rgamma(N*n1,1,1) #sample from gamma distribution
Y2<-rgamma(N*n2,2,2)
#Y1<-rnorm(N*n1,0,1) #alternatively:
#Y2<-rnorm(N*n2,0,2) #sample from normal distribution
Y<-c(Y1,Y2)
dim(Y)<-c(N,n1+n2)
tstat<-function(Y,n1,n2) {
  Y1<-Y[1:n1]
  Y2<-Y[(n1+1):(n1+n2)]
  #two-sample t test statistic
  T<-(mean(Y1)-mean(Y2))/sqrt(var(Y1)/n1+var(Y2)/n2)
  #Satterthwaite approximation of degrees of freedom
  df<-(var(Y1)/n1+var(Y2)/n2)^2/((var(Y1)/n1)^2/(n1-1)+(var(Y2)/n2)^2/(n2-1))
  return(c(T,df))
}
#calculate test statistic for N samples
R<-apply(Y,1,tstat,n1,n2)
#estimate significance level
a<-mean(ifelse(abs(R[1,])>qt(0.975,R[2,]),1,0))
a
```
Parametric Bootstrap

Suppose $Y_1, \ldots, Y_n \overset{iid}{\sim} P_\theta$ and $\hat{\theta} = S(Y)$ is an estimator for $\theta$.

**Aim:** Determine variance of $\hat{\theta}$.

**Problem:** Iterative algorithms such as the EM algorithm often do not give analytic expressions for the variance of the estimator.

**Idea:** Simulate from the distribution $P_\theta$ to obtain samples

$$Y^{(b)} = (Y_1^{(b)}, \ldots, Y_n^{(b)}), \quad b = 1, \ldots, B$$

and for each sample an estimate

$$\hat{\theta}^{(b)} = S(Y^{(b)}).$$

Then the standard error of $\hat{\theta} = S(Y)$ can be estimated by

$$\hat{\sigma}^2(\hat{\theta}) = \frac{1}{B} \sum_{b=1}^{B} (\hat{\theta}^{(b)} - \bar{\theta})^2.$$  

**Problem:** We cannot sample from $P_\theta$ since $\theta$ is unknown.

**Idea:** Approximate $P_\theta$ by $P_{\hat{\theta}}$

**Parametric bootstrap:**

- Estimate $\theta$ by $\hat{\theta} = S(Y)$
- For $b = 1, \ldots, B$, simulate bootstrap replications
  $$\overset{*}{Y}_j^{(b)} \overset{iid}{\sim} P_{\hat{\theta}}.$$
- Compute bootstrap estimate $\hat{\theta}^*(b)$ from the bootstrap sample $Y^*(b)$.
- Estimate the variance of $\hat{\theta} = S(Y)$ by
  $$\hat{\sigma}^2(\hat{\theta}) = \frac{1}{B} \sum_{b=1}^{B} (\hat{\theta}^*(b) - \bar{\theta}^*)^2,$$
  where
  $$\bar{\theta}^* = \frac{1}{B} \sum_{b=1}^{B} \hat{\theta}^*(b).$$
Importance Sampling

**Aim:** Reduction of variance

Rewrite $\mathbb{E}(h(Y))$ with $Y \sim f$ as

$$\int h(y) f(y) dy = \int \frac{h(y) f(y)}{g(y)} g(y) dy.$$ 

Thus if $Z^{(1)}, \ldots, Z^{(n)} \sim g$ then

$$\frac{1}{n} \sum_{t=1}^{n} \frac{h(Z^{(t)}) f(Z^{(t)})}{g(Z^{(t)})}$$

offers an alternative approximation for $\mathbb{E}(h(Y))$.

The variance of the Monte Carlo estimate for $\mathbb{E}(h(Y))$ is minimized if

$$\frac{|h(y)| f(y)}{g(y)} \approx \text{constant}$$

**Example:**

Suppose we want to approximate $p = \mathbb{P}(Y \geq c)$ for

(i) $Y \sim \mathcal{N}(0, 1)$

(ii) $Y \sim \text{Cauchy}(0, 1)$

An intuitive Monte Carlo approximation is

$$p_n = \frac{1}{n} \sum_{t=1}^{n} 1_{(c, \infty)}(Y^{(t)})$$

where $Y^{(t)} \overset{iid}{\sim} \mathcal{N}(0, 1)$ resp. $Y^{(t)} \overset{iid}{\sim} \text{Cauchy}(0, 1)$.

Since $p_n$ is the mean of $n$ Bernoulli random variables, its relative error is

$$\frac{\sqrt{\text{var}(p_n)}}{p} = \sqrt{\frac{1 - p}{pn}}.$$ 

**Problem:** For small $p$ the relative error is large.
Importance Sampling

**Alternative approach:** Let $Z^{(t)}$ be exponentially distributed on $(c, \infty)$,

$$Z^{(t)} \sim g(z) = \exp(c - z) \mathbf{1}_{(c, \infty)}(z).$$

Then $p$ can be approximated by

$$p'_n = \frac{1}{n} \sum_{t=1}^{n} \frac{f(Z^{(t)})}{g(Z^{(t)})},$$

where $f$ is the density of a (i) normal or (ii) Cauchy distribution.

**Remarks:**

- If the ratio $f/g$ is unbounded, the weights $f(Z^{(t)})/g(Z^{(t)})$ vary widely and the approximation is large influence by only a few values.
- The distribution $g$ should have heavier tails than $f$. 
Markov Chain Monte Carlo

Suppose that \( Y = (Y_1, \ldots, Y_d)^T \) has density \( f(y) \).

**Aim:** Approximation

\[ \mathbb{E}(h(Y)) \approx \frac{1}{n} \sum_{t=1}^{n} h(Y^{(t)}) \quad \text{with } Y^{(1)}, \ldots, Y^{(n)} \overset{\text{iid}}{\sim} f(y). \]

**Problem:** Independent sampling from a multivariate distribution is often difficult.

**Solution:**

- The Law of Large Numbers

\[ \frac{1}{n} \sum_{t=1}^{n} h(Y^{(t)}) \to \mathbb{E}(h(Y)) \quad \text{as } n \to \infty. \]

still applies if \( Y^{(t)} \) are (not too) dependent observations with \( Y^{(t)} \sim f(y) \).

- Idea: Generate a sequence of random numbers \( Y^{(k)} \) which converges to a dependent sample from the joint distribution \( f(y) \)

\[ Y^{(t)} \sim f(y) \quad \text{but NOT } \quad Y^{(t)} \overset{\text{iid}}{\sim} f(y). \]

- We will show that this can be accomplished by sampling \( Y_i^{(t)} \) from the conditional distribution

\[ f(y_i|Y_1^{(t)}, \ldots, Y_i^{(t)}, Y_{i+1}^{(t-1)}, \ldots, Y_d^{(t-1)}) \]

for \( i = 1, \ldots, d \)