STAT 330
Tutorial 7

Mengqi (Molly) Cen
Department of Statistics & Actuarial Science

Nov 2nd/ Nov 4th

Outline

1. **The Method of Monte Carlo**
    - Transformations of commonly used rvs
    - Accept-Reject Algorithm

2. **Bootstrap Procedures**

The Method of Monte Carlo

4.8.5. Determine a method to generate random observations for the logistic pdf .

Review:

Monte Carlo methods are a class of computational algorithms that obtain numerical results by repeated random sampling.

How to generate random variables ?

Use transformations of commonly used rvs: for example, If given a cdf $F(\cdot)$, $Y = F^{-1}(U)$ with $U \sim U(0,1)$ has $Y \sim F(\cdot)$.

Answer:

The cdf of the logistic distribution is

$$F(x) = \frac{1}{1 + e^{-x}}, \quad -\infty < x < \infty.$$

To determine the inverse of this function, set $u = \frac{1}{1+e^{-x}}$ and then solve for x.

Take inverse of both sides, we have

$$u^{-1} = 1 + e^{-x}$$

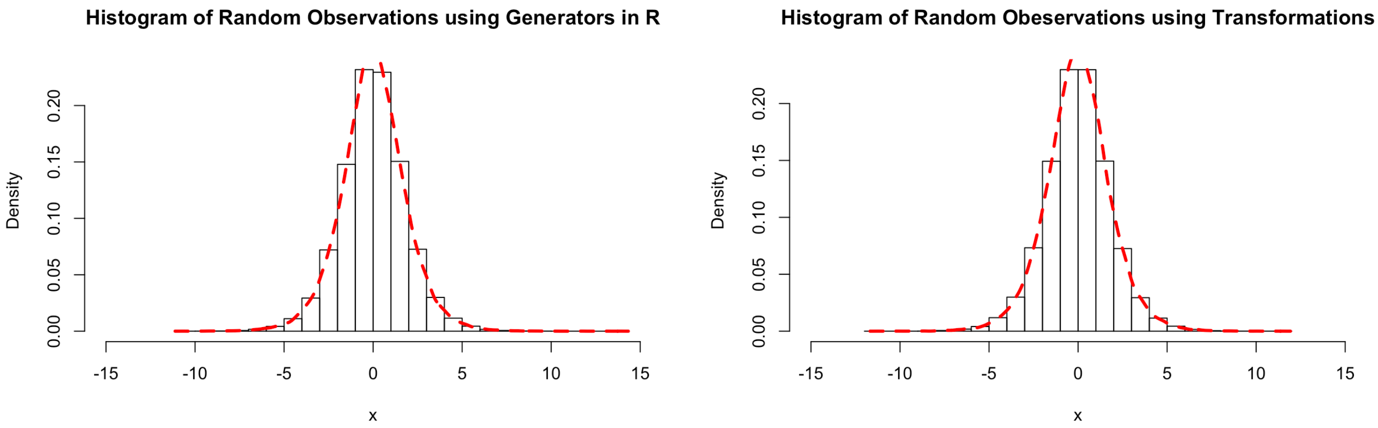
Then we take logrithm of both sides and get

$$-x = \log(u^{-1} - 1).$$

Thus, $F^{-1}(u) = x = \log\left(\frac{u}{1-u}\right)$.

Hence, if U is uniform $(0,1)$ then $\log \frac{U}{1-U}$ has a logistic distribution with cdf $F(x)$.

The Method of Monte Carlo




```r
Fx = function(n){
  U = runif(n)
  Fx = log(U/(1-U))
  Fx
}

par(mfrow=c(1,2))

logistic<-rlogis(100000, location = 0, scale = 1)
hist(logistic,freq=FALSE,main="Histogram of Random Observations using Generators in R",xlab="x",xlim=c(-15,15))
lines(density(logistic),col=2,lty=2,lwd=3)

mclogistic<-Fx(100000)
hist(mclogistic,freq=FALSE,main="Histogram of Random Obeservations using Transformations",xlab="x",xlim=c(-15,15))
lines(density(mclogistic),col=2,lty=2,lwd=3)
```

The Method of Monte Carlo

4.8.20. For $\alpha > 0$ and $\beta > 0$, consider the following accept–reject algorithm:

1. Generate U_1 and U_2 iid uniform$(0,1)$ random variables. Set $V_1 = U_1^{1/\alpha}$ and $V_2 = U_2^{1/\beta}$.

2. Set $W = V_1 + V_2$. If $W \le 1$, set $X = V_1/W$; else go to step 1.

3. Deliver X.

Show that X has a beta distribution with parameters α and β.

Review: **Accept - Reject Algorithm**

If $f(\cdot)$ is a pdf and $f(x) \le Mg(x)$ with M a constant and $g(\cdot)$ the instrumental pdf.

Step 1. Generate $Y \sim g(\cdot)$ and $U \sim U(0,1)$ independently.

Step 2. If $U \le \frac{f(Y)}{[Mg(Y)]}$, take $X = Y$ and go to Step 3;

otherwise, return to Step 1.

Step 3. Obtain X, which follows $f(\cdot)$.

The Method of Monte Carlo

Answer :

We note that $F_X(x) = P(X \le x) = P(\frac{V_1}{W} \le x \mid W \le 1)$

$$= \frac{P(V_1 \le xW, W \le 1)}{P(W \le 1)}$$

$$\propto c\, P(V_1 \le xW, W \le 1)$$

First, we want to find the cdf of V_1 and V_2.

We have $V_1 = U_1^{1/\alpha}$ and $V_2 = U_2^{1/\beta}$ where U_1 and U_2 are iid uniform $(0,1)$ random variables.

$$F_{V_1}(x) = P(V_1 \le x) = P(U_1^{1/\alpha} \le x) = P(U_1 \le x^\alpha)$$
$$= \int_0^{x^\alpha} 1\, dx = x^\alpha .$$

$$F_{V_2}(x) = P(V_2 \le x) = P(U_2^{1/\beta} \le x) = P(U_2 \le x^\beta)$$
$$= \int_0^{x^\beta} 1\, dx = x^\beta .$$

Take the derivative of $F_{V_1}(x)$ and $F_{V_2}(x)$ and we get the pdf of V_1 and V_2 :

$$f_{V_1}(x) = \alpha x^{\alpha-1}, \quad 0 < x < 1.$$

$$f_{V_2}(x) = \beta x^{\beta-1}, \quad 0 < x < 1.$$

The Jacobian determinant of transformation for $V = V_1$, $W = V + V_2$ is,

$$|J| = \begin{vmatrix} \frac{\partial v_1}{\partial v_1} & \frac{\partial v_1}{\partial v_2} \\ \frac{\partial w}{\partial v_1} & \frac{\partial w}{\partial v_2} \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ 1 & 1 \end{vmatrix} = 1$$

Thus, the joint distribution is $f_{V,W}(v,w) = \alpha v^{\alpha-1} \beta (w-v)^{\beta-1}$, where $0 < v < 1$ and $v < w < v+1$.

We have that $F_X(x) \propto \int_0^1 \int_0^{wx} \alpha v^{\alpha-1} \beta (w-v)^{\beta-1}\, dv\, dw$

$$= \int_0^1 \alpha (wx)^{\alpha-1} \beta (w - wx)^{\beta-1}\, dw$$

$$= x^{\alpha-1} (1-x)^{\beta-1} \int_0^1 \alpha\beta w^{\alpha+\beta-2}\, dw$$

Thus, X has a beta distribution with parameters α and β.

Bootstrap Procedures

Review:

Consider rv $X \sim F(\cdot)$: iid observations $X_1, \ldots, X_n$.

Bootstrap variance estimation:

Step 1. Generate $X_{1b}^*, \ldots, X_{nb}^*$ iid from the empirical function $F_n(\cdot)$.

(Resample with size n from $X_1, \ldots, X_n$ with replacement)

Step 2. Calculate $\hat{\theta}(X_{1b}^*, \ldots, X_{nb}^*)$, denoted by $\hat{\theta}_b^*$.

Repeat step 1 and 2 B times and obtain

$$\{\hat{\theta}_b^* : b = 1, \ldots, B\}$$

Calculate $S_{\hat{\theta}}^2 = \sum_{b=1}^{B} (\hat{\theta}_b^* - \bar{\hat{\theta}}^*)^2 / B-1$

with $\bar{\hat{\theta}}^* = \sum_{b=1}^{B} \hat{\theta}_b^* / B$.

Use $S_{\hat{\theta}^*}^2$ to estimate $Var(\hat{\theta})$.

Rcodes:

```
B<-1000
n<-100

X<-rnorm(n)
thetabstarhat<-rep(0,B)

for(i in 1:B){
  Xstar<-sample(X,n,replace=TRUE)
  thetabstarhat[i]<-mean(Xstar)
}

thetastarhatbar<-sum(thetabstarhat)/B
ssquared<-0
for(j in 1:B){
  ssquared<-ssquared+(thetabstarhat[j]-thetastarhatbar)^2/(B-1)
}

ssquared
var(X)/n
```

Results:

```
> ssquared
[1] 0.01128843
> var(X)/n
[1] 0.01149071
```

Bootstrap Procedures

4.9.9. Let $\mathbf{x}' = (x_1, x_2, \ldots, x_{n_1})$ and $\mathbf{y}' = (y_1, y_2, \ldots, y_{n_2})$ be the realizations of the two random samples. The test statistic is the difference in sample means $\overline{y} - \overline{x}$. The estimated p-value of the test is calculated as follows:

1. Combine the data into one sample $\mathbf{z}' = (\mathbf{x}', \mathbf{y}')$.
2. Obtain all possible samples of size n_1 drawn without replacement from $\mathbf{z}$. Each such sample automatically gives another sample of size n_2, i.e., all elements of $\mathbf{z}$ not in the sample of size n_1. There are $M = \binom{n_1+n_2}{n_1}$ such samples.
3. For each such sample j:
   - **(a)** Label the sample of size n_1 by $\mathbf{x}^*$ and label the sample of size n_2 by $\mathbf{y}^*$.
   - **(b)** Calculate $v_j^* = \overline{y}^* - \overline{x}^*$.
4. The estimated p-value is $\widehat{p}^* = \#\{v_j^* \geq \overline{y} - \overline{x}\}/M$.

Bootstrap Procedures

Suppose we have two samples each of size 3 which result in the realizations: $\mathbf{x}' = (10, 15, 21)$ and $\mathbf{y}' = (20, 25, 30)$. Determine the test statistic and the permutation test described above along with the p-value.

```
library(gtools)

x<-c(10,15,21)
y<-c(20,25,30)
z<-c(x,y)

xstar<-combinations(n=6,r=3,v=z)
xstarsum<-apply(xstar[1:20,], 1, sum)
xstarbar<-apply(xstar[1:20,], 1, mean)
zsum<-sum(z)
ystarsum<-zsum-xstarsum
ystarbar<-ystarsum/3
vstar<-ystarbar-xstarbar
ybar<-mean(y)
xbar<-mean(x)
pstarhat<-sum((ystarbar-xstarbar)>=(ybar-xbar))/20
```

```
[1] 0.1
```

Questions