Εξέταση Προσομοίωση Ιούλιος 2019

Πρόβλημα 1

Η συνάρτηση πυκνότητας πιθανότητας είναι

$$f(x) = \begin{cases} 0, & x < 0 \\ \dfrac{e^x}{e-1}, & 0 \le x \le 1 \\ 0, & x > 1 \end{cases}$$

(α)

Η πρώτη γεννήτρια βασίζεται στη μέθοδο αντιστροφής. Η συνάρτηση κατανομής είναι

$$F(x) = \begin{cases} 0, & x \le 0 \\ \dfrac{e^x - 1}{e-1}, & 0 \le x \le 1 \\ 1, & x \ge 1 \end{cases}$$

Επομένως αν $U \sim U(0,1)$, ισχύει $F(X) = U$ και λύνοντας ως προς X παίρνουμε

$$X = \log(1 + (e-1)U).$$

(β)

Η δεύτερη γεννήτρια βασίζεται στη μέθοδο απόρριψης. Ως βοηθητική κατανομή χρησιμοποιούμε την $U(0,1)$, επομένως $g(x) = 1, 0 \le x \le 1$.

Βρίσκουμε πρώτα τη μέγιστη τιμή του λόγου $f(x)/g(x)$:

$$c = \max_{0\le x\le 1} \frac{f(x)}{g(x)} = \max_{0\le x\le 1} \frac{e^x}{e-1}.$$

Επειδή η e^x είναι αύξουσα, μεγιστοποιείται για $x = 1$, επομένως

$$c = \frac{e}{e-1}$$

και ο λόγος

$$\frac{f(x)}{cg(x)} = e^{x-1}.$$

Επομένως ο αλγόριθμος απόρριψης διαρθρώνεται ως εξής:

1. Δημιουργία $Y \sim U(0,1)$.

2. Δημιουργία $U \sim U(0,1)$.

3. Αν $U \leq e^{Y-1}$, δεχόμαστε την Y και θέτουμε $X = Y$, διαφορετικά επιστροφή στο βήμα 1.

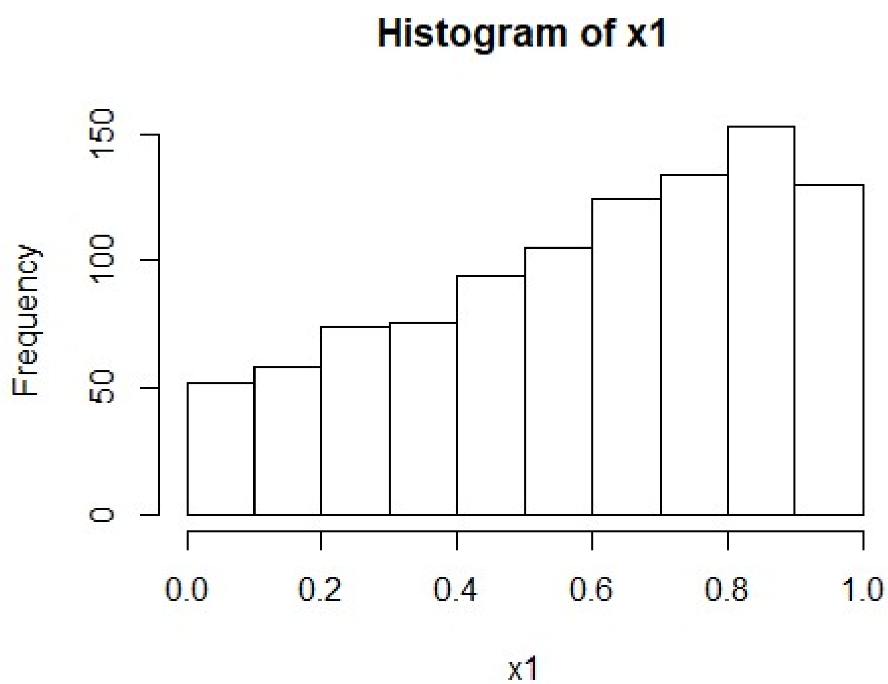
(γ)

Πρώτα δημιουργούμε 1000 παρατηρήσεις από τη γεννήτρια αντίστροφου μετασχηματισμού:

```
n=1000
u=runif(n,0,1)
x1=log(1+(exp(1)-1)*u)
```

Βλέπουμε το ιστόγραμμα των παρατηρήσεων από την πρώτη γεννήτρια.

```
hist(x1)
```



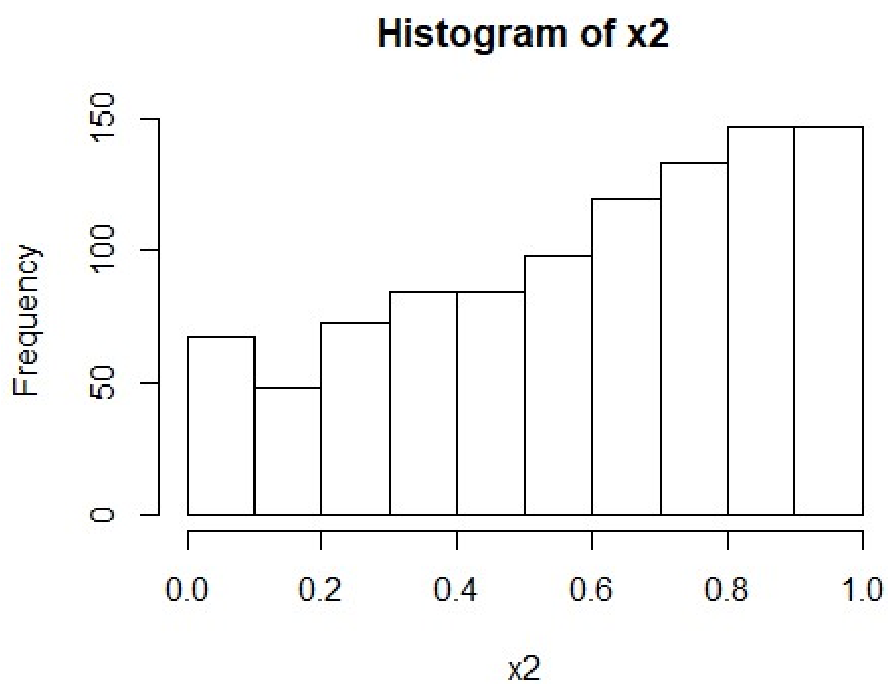
Στη συνέχεια, για την υλοποίηση του αλγορίθμου απόρριψης ορίζουμε την παρακάτω συνάρτηση R που δημιουργεί μια παρατήρηση από την f:

```
accrej=function ()
{
  acc=F
  while (acc==F)
  {
        Y=runif(1)
```

```
        U=runif(1)
        if (U<=exp(Y-1))
           acc=T
  }
  return(Y)
}
```

Χρησιμοποιώντας τη συνάρτηση accrej δημιουργούμε 1000 παρατηρήσεις από την κατανομή, με το αντίστοιχο ιστόγραμμα:

```
x2=c()
for (i in 1:n)
{
  x2[i]=accrej();
}
hist(x2)
```



Τέλος εφαρμόζουμε τον έλεγχο Kolmogorov-Smirnoff για να ελέγξουμε την υπόθεση ότι τα δύο παραπάνω δείγματα προέρχονται από την ίδια κατανομή:

```
ks.test(x1,x2)

##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  x1 and x2
## D = 0.033, p-value = 0.6476
## alternative hypothesis: two-sided
```

Με βάση την τιμή $p = 0.5$, δεν προκύπτει στατιστικά σημαντική ένδειξη ότι τα δείγματα προέρχονται από διαφορετικές κατανομές.

Πρόβλημα 2

Έστω N η τυχαία μεταβλητή που δείχνει την κατάσταση ενός ατόμου με βάση τις δύο μετρήσεις του και παίρνει τις τιμές 1,2,3 αν το άτομο βρίσκεται στην κατηγορία υψηλού κινδύνου, χαμηλού κινδύνου, ή επιτήρησης, αντίστοιχα. Επομένως η N ορίζεται ως εξής:

$$N = \begin{cases} 1, & \text{αν } \max(X_1, X_2) > 17 \\ 2, & \text{αν } \max(X_1, X_2) < 13 \\ 1, & \text{διαφορετικά} \end{cases}$$

Έστω $p_j = P(N = j), j = 1,2,3$. Για την εκτίμηση των p_j χρησιμοποιούμε 1000 σενάρια προσομοίωσης. Σε κάθε σενάριο δημιουργούμε μια τιμή για την X_1 από την $N(14.2, 2.4^2)$ και δεδομένης της τιμής $X_1 = x_1$, μια τιμή για τη X_2 από την $N(x_1, 1.2^2)$. Στον πίνακα δεδομένων που σχηματίζουμε προσθέτουμε μια τρίτη στήλη που περιέχει την τιμή της κατηγορίας N για κάθε σενάριο με βάση τον παραπάνω ορισμό.

```
n=1000
dat=rep(0,3*n)
dat=matrix(dat,n,3)
for (i in 1:n)
{
  x1=rnorm(1,14.2,2.4)
  x2=rnorm(1,x1,1.2)
  if (max(x1, x2)>17)
    N=1
  else if (max(x1,x2)<13)
    N=2
  else
    N=3
c(x1,x2,N)
dat[i,]=c(x1,x2,N)
}
```

Με την εντολή table μπορούμε να δούμε τους αριθμούς των παρατηρήσεων για κάθε κατηγορία και διαιρώντας με το μέγεθος του δείγματος παίρνουμε τις αντίστοιχες εκτιμήσεις των πιθανοτήτων:

```
T=table(dat[,3])
T

##
##   1   2   3
## 158 263 579
```

```
pest=T/n
pest

##
##     1     2     3
## 0.158 0.263 0.579
```

Πρόβλημα 3

Σκοπός της προσομοίωσης είναι να εκτιμηθεί η μεροληψία της εκτιμήτριας $\hat{\theta} = 1/\overline{X}_n$ της παραμέτρου θ της εκθετικής κατανομής, όπου $\overline{X}_n$ ο δειγματικός μέσος ενός τυχαίου δείγματος μεγέθους n από την κατανομή, όταν $n = 30, \theta = 2$.

Επειδή η μεροληψία είναι $B = E(\hat{\theta} - \theta)$, πρέπει να εκτιμηθεί η μέση τιμή της εκτιμήτριας μέσω προσομοίωσης.

Σε κάθε σενάριο προσομοίωσης δημιουργούμε ένα δείγμα μεγέθους 30 από την κατανομή $Exp(2)$ και παίρνουμε τον αντίστροφο του δειγματικού μέσου. Αυτή είναι η τιμή που παίρνει η εκτιμήτρια για αυτό το σενάριο. Επαναλαμβάνοντας για 1000 σενάρια, παίρνουμε μια εκτίμηση του $E(\hat{\theta})$ και επομένως και του B.

Η παρακάτω συνάρτηση δημιουργεί ένα σενάριο. Έχει σχεδιαστεί πιο γενικά με ορίσματα n, θ, όπου n το μέγεθος του δείγματος σε κάθε σενάριο και θ η πραγματική τιμή της παραμέτρου.

```
thetaest=function(n, theta)
{
  x=rexp(n,theta)
  thest=1/mean(x)
  return(thest)
}
```

Για τη δημιουργία των σεναρίων καλούμε αυτή τη συνάρτηση 1000 φορές με ($n = 30, theta = 2$, καταγράφουμε τις τιμές της εκτιμήτριας σε κάθε σενάριο και βρίσκουμε την εκτίμηση της μεροληψίας.

```
N=1000
Y=rep(0,N)
for (i in 1:N)
{
  Y[i]=thetaest(30,2)
}
Best=mean(Y)-2
Best

## [1] 0.06250767
```

Επομένως η δοσμένη εκτιμήτρια υπερεκτιμά την αληθή τιμή $\theta = 2$ κατά 3.13%.