Segundo examen parcial Simulación

David López

28/10/2020

Al inicio se presentan todas las preguntas que requirieron el uso de R para su interpretación, desarrollo y solución.

Pregunta 1

En esta pregunta modelo una simulación de variables aleatorias con función de densidad Rayleigh. La hago por medio de variables antiéticas para minimizar la varianza producida en la simulación. Utilizo el siguiente modelo de variables antiéticas: $\frac{X+Y}{2}$Entonces, para ver los resultados presento un histograma de los valores obtenidos a partir de la simulación. Para ello utilizo los siguientes valores: $n = 10000$ iteraciones y de parámetro de escala de la función de densidad de Rayleigh $\sigma = 10$.

```
# Pregunta 1 parte 1
library(VGAM)
```

```
## Warning: package 'VGAM' was built under R version 3.6.3
```

```
## Loading required package: stats4
```

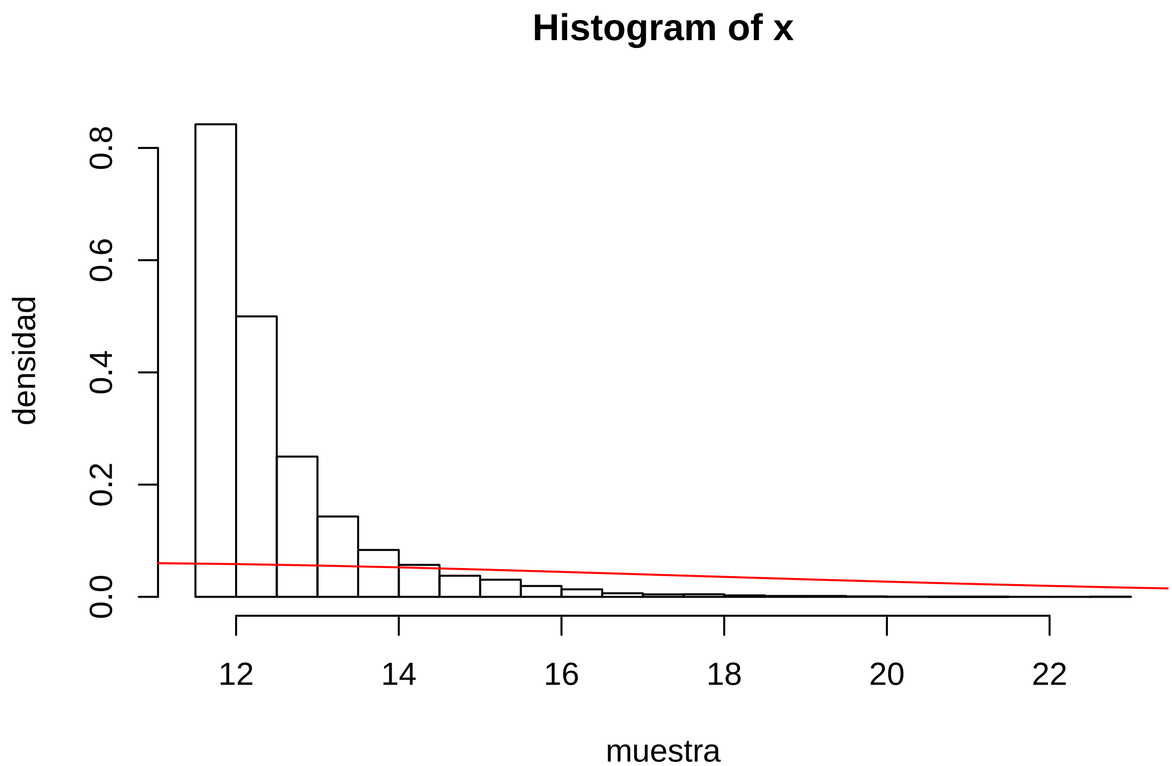
```
## Loading required package: splines
```

```
# Muestra aleatoria de una función de densidad Rayleigh por medio de variables antitéticas
X <- NULL; Y <- NULL; n <- 10000; sigma <- 10; x <- NULL;

for (i in 1:n){
  aux <- runif(1,0,1);
  X[i] <- sigma * sqrt(log(1/(1-aux)^2));
  Y[i] <- sigma * sqrt(log(1/(aux)^2));

  x[i] <- (X[i] + Y[i])/2;
}

hist(x, prob = T, breaks=20, xlab="muestra", ylab="densidad")
curve(drayleigh(x, scale = sigma, log = FALSE), from = 0, to = 100, add = T, col = "red")
```



Ahora, presento la varianza de la simulación de la muestra aleatoria por medio de variables antitéticas

```
varAnti <- var(x);
varAnti
```

```
## [1] 1.155486
```

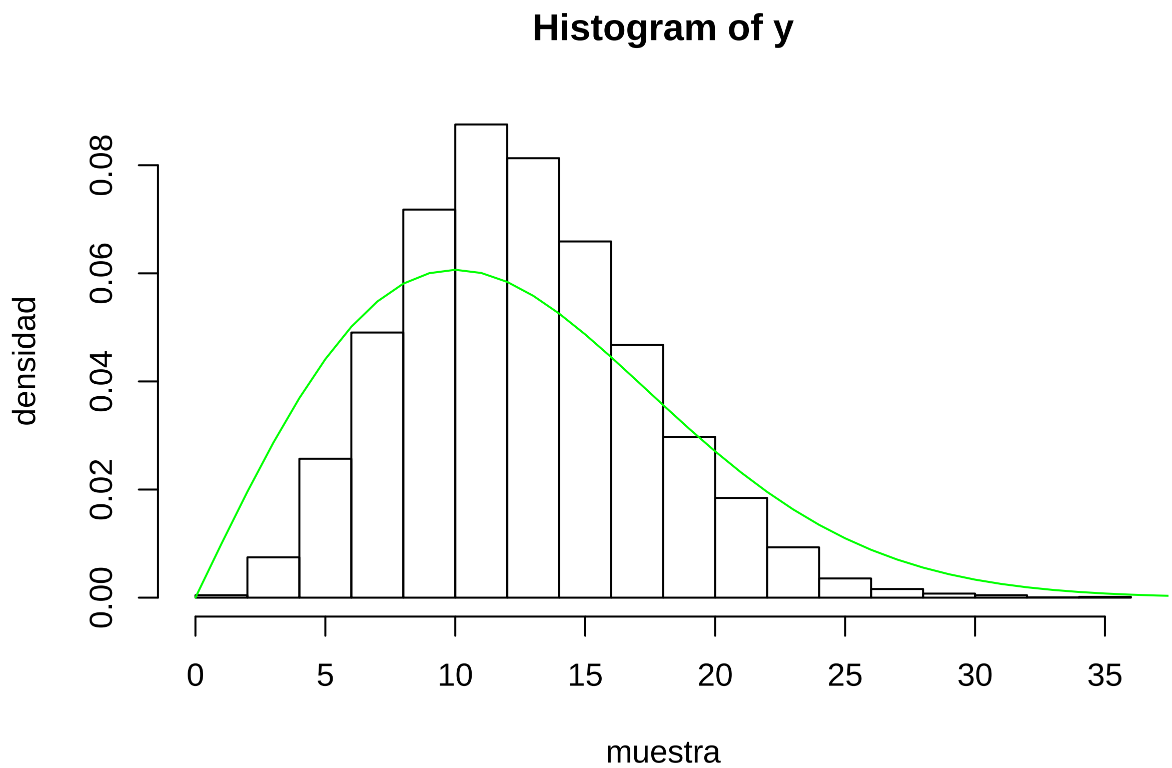
Por otro lado, simulo la otra muestra de variables aleatorias de la misma densidad de Rayleigh con exactamente los mismos valores $n = 10000$ y $\sigma = 10$, su parametro de escala. En esta ocasión, simulo dos variables aleatorias distintas e independientes entre sí y utilizo su combinación lineal de variables aleatorias dada por lo siguiente $\frac{U+V}{2}$. De esta manera, obtenemos el siguiente histograma:

```
#Pregunta 1 parte 2

library(VGAM)
# Muestra aleatoria de una función de densidad Rayleigh por medio de variables independientes
U <- NULL; V <- NULL; n <- 10000; sigma <- 10; x <- NULL;


for (i in 1:n){
  U[i] <- sigma * sqrt(-2 * log(runif(1,0,1)));
  V[i] <- sigma * sqrt(-2 * log(runif(1,0,1)));
}
y <- (U+V)/2;

hist(y, prob = T, breaks=20, xlab="muestra", ylab="densidad")
curve(drayleigh(x, scale = sigma, log = FALSE), from = 0, to = 100, add = T, col = "green")
```



El histograma anterior resulta ser diferente al primero que se produjo por medio de variables antitéticas. Sin embargo, la varianza producida en esta ocasión es mayor, ya que son dos variables distintas que se utilizan para obtener su información. Aunque pueda simular una aproximación cercana a la distribución, resulta que los valores son dispersos

```
varNormal <- var(y);
varNormal
```

```
## [1] 21.87448
```

Concluimos que el método de variables antitéticas reducen en gran porcentaje la varianza de la simulación de la función de densidad. La comparación de varianzas se hace con el supuesto de que la varianza por medio de variables antitéticas es el 100%.

```
porcentajeVarianza <- (varAnti * 100)/varNormal;
porcentajeVarianza
```

```
## [1] 5.282345
```

Con lo anterior, se traduce que la varianza antitética es cercana a la fracción $\frac{1}{20} = 0.05$ de la varianza obtenida por medio de variables independientes entre sí.

Pregunta 5

En este problema simulo una cadena de Markov que simula los ataques a los puertos de unas computadoras. Simulo de Markov dada una distribución condicional y una matriz permutación P definidas como siguen:

$\pi = (0, 0, 0, 0, 1)$

$$P = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 8/13 & 3/13 & 1/13 & 1/13 \\ 1/16 & 3/16 & 3/8 & 1/4 & 1/8 \\ 0 & 1/11 & 4/11 & 5/11 & 1/11 \\ 0 & 1/8 & 1/2 & 1/8 & 1/4 \end{bmatrix}$$

```r
# Pregunta 5 (Cadenas de Markov)

library("markovchain")
```

```
## Warning: package 'markovchain' was built under R version 3.6.3
```

```
## Package:  markovchain
## Version:  0.8.5-2
## Date:     2020-09-07
## BugReport: https://github.com/spedygiorgio/markovchain/issues
```

```r
n <- 100;
estadoInicial <- c(0,0,0,0,1);
P <- matrix(c(0, 0, 0, 0, 1,
              0, 8/13, 3/13, 1/13, 1/13,
              1/16, 3/16, 3/8, 1/4, 1/8,
              0, 1/11, 4/11, 5/11, 1/11,
              0, 1/8, 1/2, 1/8, 1/4),
            byrow=T, nrow=5)


# Se denota -1 a que no hay ataques
cadenaAtaques <- new("markovchain", states = c("80","134","139","445","-1"),
                    transitionMatrix  = P, name = "Ataques a Puertos de computadora");

estadoNveces <- function(estadoIni, numVeces, matriz){
  estadoN <- estadoInicial* (cadenaAtaques)^numVeces;
  estadoN;
}

estado100veces <- estadoNveces(estadoInicial, n, P);
estado100veces;
```

```
##              80       134       139       445     -1
## [1,] 0.02146667 0.2669333 0.3434667 0.2273333 0.1408
```

Después de hacer la simulación, también busco cuál es la distribución estacionaria, a partir del espacio de estados de la cadena de Markov.

```r
distribucionEstacionaria <- steadyStates(cadenaAtaques);
distribucionEstacionaria;
```

```
##              80       134       139       445     -1
## [1,] 0.02146667 0.2669333 0.3434667 0.2273333 0.1408
```

Pregunta 3

La siguiente pregunta se escribió a mano y se adjunta en la siguiente página.

(Parcial 2) Simulación David Isaac López Romero

CU: 173993

3) Si $\hat{\theta}_1$ y $\hat{\theta}_2$ son cualesquiera dos estimadores insesgados de θ, encontrar el valor c^* que minimiza la varianza del estimador $\hat{\theta}_c = c\hat{\theta}_1 + (1-c)\hat{\theta}_2$

Sol:

Vemos que $\hat{\theta}_1$ y $\hat{\theta}_2$ son independientes pues son dos estimadores distintos. (Usan los mismos valores, pero no dependen uno del otro)

$$Var(\hat{\theta}_c) = Var(c\hat{\theta}_1 + (1-c)\hat{\theta}_2) = c^2 Var(\hat{\theta}_1) + (1-c)^2 Var(\hat{\theta}_2)$$

Como $\hat{\theta}_1$ y $\hat{\theta}_2$ tienen varianza positiva, $Var(\hat{\theta}_1) = k\,Var(\hat{\theta}_2)$, $k>0$

(Si $Var(\hat{\theta}_1), Var(\hat{\theta}_2) = 0$ ya conozco a θ y no sirve tener un estimador)

$$\therefore Var(\hat{\theta}_c) = c^2 Var(\hat{\theta}_1) + (1-c)^2 Var(\hat{\theta}_2) = c^2 k\,Var(\hat{\theta}_2) + (1-c)^2\,Var(\hat{\theta}_2)$$

$$= \{c^2 k + (1-c)^2\}\,Var(\hat{\theta}_2)$$

Busco minimizar la varianza; entonces, $\min_{c\in\mathbb{R}} \{c^2 k + (1-c)^2\}$

Sea $F(c) = c^2 k + (1-c)^2$ entonces,

$$F'(c) = 2ck + 2(1-c)(-1) = 2ck + 2(c-1) = 2ck + 2c - 2$$

Ahora,

$$F'(c) = 0 \iff 2ck + 2c - 2 = 0 \iff ck + c - 1 = 0 \iff c(k+1) = 1$$

$$\iff c = \frac{1}{k+1}$$

Además,

$F''(c) = 2k + 2 > 0$, pues $k > 0$ $\therefore F$ es cóncava. $\forall c \in \mathbb{R}$

$\therefore c^* = \frac{1}{k+1}$ es mínimo

Así pues $\hat{\theta}_c$ minimiza la varianza si $\hat{\theta}_c = c\hat{\theta}_1 + (1-c)\hat{\theta}_2$, $c = \frac{1}{k+1}$

i.e. $\hat{\theta}_{c^*} = \frac{1}{k+1}\hat{\theta}_1 + \frac{k}{k+1}\hat{\theta}_2$

//