

Tarea 2 Simulación

Bernardo Mondragón Brozon, Diego García Santoyo, Rayan García Fabian, Karen Delgado Curiel

13 de septiembre de 2018

Problema 1

Probar por inducción matemática que para un generador lineal congruencial,

$$Z_i = a^i Z_0 + c \frac{a^i - 1}{a - 1} \pmod{m}$$

Demostración:

Para $i = 0$ se tiene que

$$\begin{aligned} Z_0 &= Z_0 \pmod{m} \\ &= a^0 Z_0 + c \frac{a^0 - 1}{a - 1} \pmod{m}, \end{aligned}$$

y para $i = 1$ se tiene que

$$\begin{aligned} Z_1 &= aZ_0 + c \pmod{m} \\ &= a^1 Z_0 + c \frac{a^1 - 1}{a - 1} \pmod{m}. \end{aligned}$$

Procediendo por inducción matemática, se supone válido para i de manera que

$$\begin{aligned} Z_{i+1} &= aZ_i + c \pmod{m} \\ &= a \left(a^i Z_0 + c \frac{a^i - 1}{a - 1} \right) + c \pmod{m} \\ &= a^{i+1} Z_0 + c \frac{a^{i+1} - 1}{a - 1} \pmod{m}. \end{aligned}$$

Por lo tanto es válido para toda $i \in \mathbb{N}$.

■

Problema 2

¿Qué se puede decir del periodo de $Z_i = 630,360,016Z_{i-1} \pmod{2^{31} - 1}$?

Solución:

Se puede decir que su periodo es menor o igual $2^{31} - 1$. Además no tiene periodo completo, recordemos que de acuerdo a la primera condición del teorema de periodo completo de Hull & Dobell, c y m deben ser primos relativos, en este caso $m = 2^{31} - 1$ y $c = 0$, de manera que no se cumple la condición.

Problema 3

Sin calcular ninguna Z_i , determinar cuál de los siguientes GLC's mixtos tienen periodo completo.

- a) $Z_i = 13Z_{i-1} + 13 \pmod{16}$.
- b) $Z_i = 12Z_{i-1} + 13 \pmod{16}$.
- c) $Z_i = 13Z_{i-1} + 12 \pmod{16}$.
- d) $Z_i = Z_{i-1} + 12 \pmod{13}$.
- e) El GLC con parámetros: $a = 2, 814, 749, 767, 109$, $c = 59, 482, 661, 568, 307$, $m = 248$.

Solución:

Considerando lo mencionado en la pregunta anterior, se puede determinar que (c) no tiene periodo completo. La segunda condición de Hull & Dobell indica que si q es un número primo que divide a m entonces q debe dividir a $a - 1$. Notemos que (b) no cumple (considerar $q = 2$), por lo que tampoco tiene periodo completo. Mencionando el último punto, si 4 divide a m , entonces 4 divide a $a - 1$. Vemos que esta condición no aplica (d) por lo tanto tiene periodo completo. Es fácil verificar que (a) y (d) cumplen las 3 condiciones.

- a) $Z_i = 13Z_{i-1} + 13 \pmod{16}$.
El generador lineal congruencial tiene periodo completo
- b) $Z_i = 12Z_{i-1} + 13 \pmod{16}$.
El generador lineal congruencial no tiene periodo completo
- c) $Z_i = 13Z_{i-1} + 12 \pmod{16}$.
El generador lineal congruencial no tiene periodo completo
- d) $Z_i = Z_{i-1} + 12 \pmod{13}$.
El generador lineal congruencial tiene periodo completo
- e) El GLC con parámetros: $a = 2, 814, 749, 767, 109$, $c = 59, 482, 661, 568, 307$, $m = 248$.
El generador lineal congruencial tiene periodo completo

Problema 4

Mostrar que el promedio de las U_i 's tomadas de un ciclo completo de un GLC de periodo completo es $\frac{1}{2} - \frac{1}{2m}$.

Solución:

El ciclo de un generador lineal congruencial de periodo completo dado por

$$Z_i = aZ_{i-1} + c \pmod{m}$$

tiene un total de m números. Además de probó que

$$Z_i = a^i Z_0 + c \frac{a^i - 1}{a - 1} \pmod{m},$$

de manera que el promedio de los m números generados está dado por

$$\frac{1}{m} \sum_{i=1}^m \left(a^i Z_0 + c \frac{a^i - 1}{a - 1} \pmod{m} \right) = \frac{1}{m} \sum_{i=1}^{m-1} i = \frac{1}{m} \left(\frac{m(m-1)}{2} \right) = \frac{1}{2} - \frac{1}{2m}$$

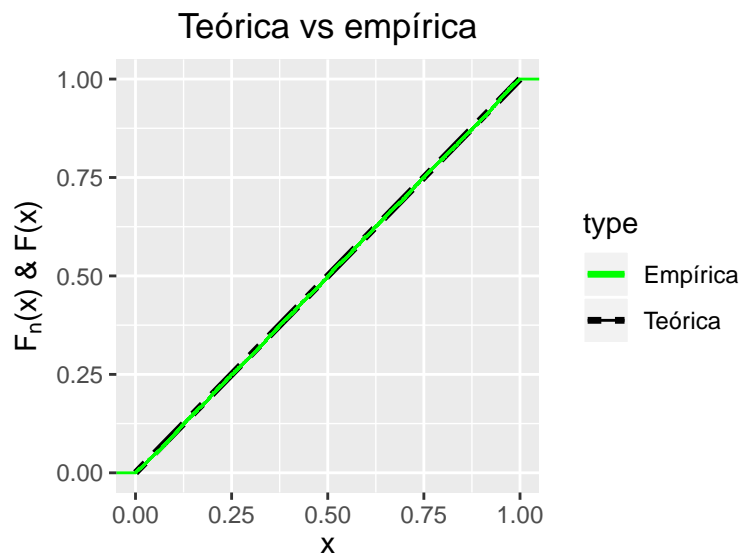
Problema 5

Generar 10,000 números uniformemente distribuidos entre 0 y 1 en Excel. Hacer un breve estudio para probar la calidad de los generadores; aplicar las pruebas de uniformidad e independencia a cada conjunto de datos. Resumir resultados en NO MAS de 2 cuartillas, incluyendo gráficas. De acuerdo a tus resultados, ¿cómo calificarías al generador de Excel?

Solución:

Pruebas de bondad de ajuste

A continuación se muestra la gráfica de la distribución empírica de los valores generados con Excel y la gráfica de la distribución uniforme en el intervalo $[0, 1]$:



Se realizará la prueba de Kolmogorov-Smirnov (prueba de bondad de ajuste) a los valores simulados para determinar si efectivamente provienen de la distribución uniforme en el intervalo $[0, 1]$. Para realizar la prueba se contrastarán las siguientes hipótesis:

$$H_0 : F(x) = x \quad \forall x \quad \text{vs.} \quad H_1 : F(x) \neq x \quad \text{para alguna } x$$

Entonces la estadística de prueba está dada por

$$D_n = \max_x |F_n(x) - x|.$$

Se rechazará la hipótesis H_0 si D_n es muy grande. En este caso, la estadística de prueba toma el siguiente valor

$$D_n = 0.0055188$$

El valor p correspondiente al valor de la estadística de prueba está dado por

$$\text{p-value} = 1.0876343$$

En este caso el valor p es mayor a 1 porque se calculó mediante una aproximación, sin embargo el verdadero valor es muy cercano a 1, de manera que no se rechaza la hipótesis de que los números aleatorios provienen de la distribución uniforme.

Aplicando la prueba de Crámer-von Misses se obtiene el siguiente valor de la estadística de prueba:

$$W_n = n \int_{-\infty}^{\infty} (F_n(x) - F(x))^2 dF(x) = 10000 \int_{-\infty}^{\infty} (F_n(x) - x)^2 dx = 0.0459533.$$

De manera que el valor p está dado por

$$p\text{-value} = 0.9003661$$

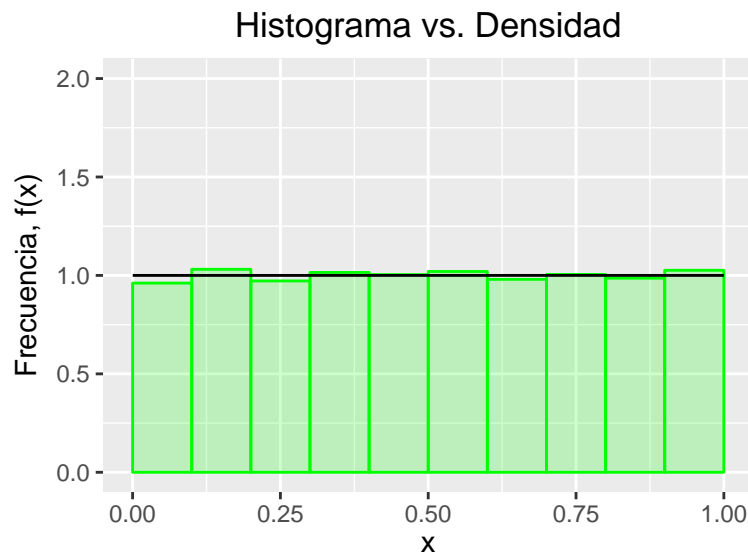
Aplicando la prueba de Anderson-Darling se obtiene el siguiente valor de la estadística de prueba:

$$A_n^2 = n \int_{-\infty}^{\infty} \frac{(F_n(x) - F(x))^2}{F(x)(1 - F(x))} dF(x) = 10000 \int_{-\infty}^{\infty} \frac{(F_n(x) - x)^2}{x(1 - x)} dx = \infty$$

De manera que el valor p está dado por

$$p\text{-value} = 6 \times 10^{-8}$$

A continuación se muestra el histograma de los datos generados en Excel y la función de densidad uniforme en el intervalo $[0, 1]$:



Para la prueba de bondad de ajuste χ^2 de Pearson, las hipótesis a contrastar son las mismas que las de la prueba de Kolmogorov-Smirnov. En este caso la estadística de prueba toma el siguiente valor:

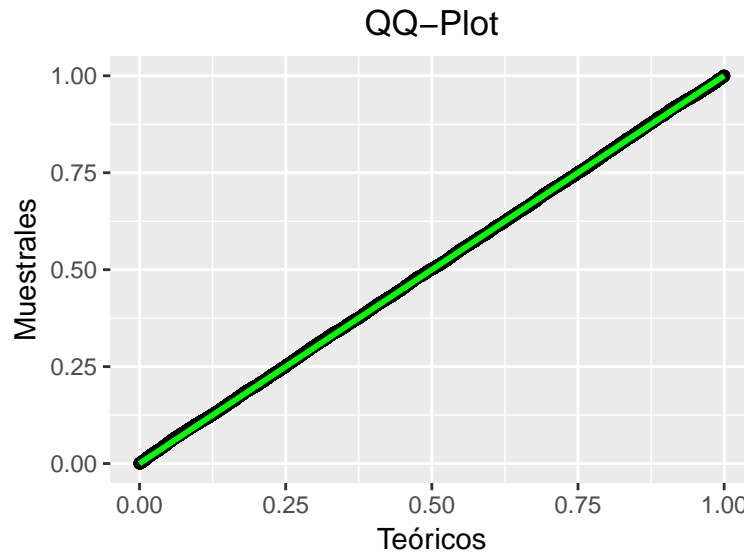
$$\chi^2 = \sum_{i=1}^k \frac{(N_j - np_j)^2}{np_j} = \sum_{i=1}^1 \frac{(N_j - 10000p_j)^2}{10000p_j} = 2026.8$$

De manera que el valor p está dado por

$$p\text{-value} = 0.3270129$$

Por lo tanto no se rechaza la hipótesis nula de que los datos generados por Excel provienen de la distribución uniforme en el intervalo $[0, 1]$.

A continuación se presenta un qq-plot de los valores generados por Excel. Si el qq-plot sigue la recta identidad cuando se grafica contra la distribución teórica, entonces se puede decir que los datos siguen adecuadamente la distribución uniforme en el intervalo $[0, 1]$.



Al parecer se tiene un buen ajuste, sin embargo, esto no es una prueba en sentido estricto; es simplemente una guía visual para determinar si se tiene un buen ajuste.

Pruebas de independencia

Aplicando la prueba de independencia que utiliza a las rachas en la sucesión de números aleatorios se obtiene el siguiente número de rachas:

$$R = 6619$$

Mediante la aproximación normal se tiene el siguiente valor del estadístico de prueba que sigue la distribución normal estándar:

$$Z = \frac{R - (2n - 1)/3}{\sqrt{(16R - 29)/90}} = -1.3800399.$$

El correspondiente valor p para el valor dado de la estadística de prueba está dado por

$$\text{p-value} = 0.0837872$$

Aplicando la prueba de independencia de Levene-Wolfowitz se obtiene el siguiente valor para la estadística de prueba:

$$R_n = 9.8723467.$$

El correspondiente valor p para el valor dado de la estadística de prueba está dado por

$$\text{p-value} = 0.1301317$$

Aplicando la prueba de gaps se obtiene el siguiente valor para la estadística de prueba:

$$R_n = 8.99.$$

El correspondiente valor p para el valor dado de la estadística de prueba está dado por

$$\text{p-value} = 0.77$$

Aplicando la prueba del poker se obtiene el siguiente valor para la estadística de prueba:

$$R_n = 7.07.$$

El correspondiente valor p para el valor dado de la estadística de prueba está dado por

$$\text{p-value} = 0.13$$

Aplicando pruebas de autocorrelacion, se obitnene los siguiente:

```
## [1] 0.77
##
## Box-Pierce test
##
## data:  unifs
## X-squared = 1, df = 3, p-value = 0.8
##
## Box-Ljung test
##
## data:  unifs
## X-squared = 1, df = 3, p-value = 0.8
```

De acuerdo a las pruebas anteriores, se puede concluir que el generador de números aleatorios de excel es un buen generador de números aleatorios. De manera que no se rechaza la hipótesis de que los números son independientes.

Problema 6

Probar que la parte fraccional de la suma de uniformes en $[0, 1]$: $U_1 + \dots + U_k$ es también uniforme en el intervalo $[0, 1]$.

Solución:

Sea $\{x\} = x - \lfloor x \rfloor$ la parte fraccional del número x , y $S_k = \sum_{i=1}^k U_i$ la suma de k variables aleatorias independientes e uniformemente distribuidas en el intervalo $[0, 1]$. Para $k = 1$ se tiene que $\{S_1\} = \{U_1\} = U_1 \sim U[0, 1]$. Para $k = 2$ se tiene que $\{S_2\} = \{U_1 + U_2\}$ con

$$f_{S_2}(s) = \begin{cases} s & 0 \leq s \leq 1 \\ 2 - s & 1 < s \leq 2 \end{cases}.$$

de manera que

$$\begin{aligned} F_{\{S_2\}}(x) &= Pr[\{S_2\} \leq x] \\ &= \int_{u=0}^x f_{S_2}(u) du + \int_{u=1}^{1+s} f_{S_2}(u) du = s. \end{aligned}$$

Entonces $\{S_2\}$ sigue la distribución uniforme en el intervalo $[0, 1]$. Procediendo mediante inducción se supone válido para k , de manera que

$$\{S_k\} \sim U[0, 1].$$

Por lo tanto, para $k + 1$ se tiene que

$$\{S_{k+1}\} = \{U_1 + \cdots + U_{k+1}\} = \{\{U_1 + \cdots + U_k\} + U_{k+1}\} = \{\{S_k\} + U_{k+1}\}.$$

Pero $\{S_k\}$ y U_{k+1} tienen la distribución uniforme y son independientes, entonces

$$\{S_{k+1}\} \sim U[0, 1].$$

■

Problema 7

Un generador Fibonacci obtiene X_{n+1} a partir de X_n y X_{n-1} de la siguiente forma:

$$X_{i+1} = X_i + X_{i-1} \mod m$$

donde X_0 y X_1 están especificados. Suponga que $m = 5$. Sólo dos ciclos son posibles. Encontrarlos, así como su respectivo periodo.

Solución:

Problema 8

Genera 10,000 números con una semilla de $Z_0 = 1$ usando el generador $Z_n = 7^5 Z_{n-1} \mod 2^{31} - 1$ Clasifica los números en 10 celdas de igual tamaño y prueba uniformidad usando la prueba χ^2 con un nivel de confianza del 90%. Aplicar también la prueba de rachas.

Solución:

En la prueba χ^2 el valor del estadístico de prueba esta dado por

$$\chi^2 = 950 < \chi_{9,0.1}^2 = 15.$$

De manera que se obtiene un valor p dado por

$$p\text{-value} = 0.86.$$

Por lo tanto no rechaza la hipótesis de que los números generados por el generador lineal congruencial siguen la distribución uniforme en el intervalo $[0, 1]$.

Aplicando la prueba de rachas de Wald-Wolfowitz, se tiene el siguiente valor del estadístico de prueba:

$$R = -1.18.$$

De manera que se obtiene un valor p dado por

$$\text{p-value} = 0.24.$$

El valor p es muy pequeño, entonces se rechaza la hipótesis de que los números generados por el generador lineal congruencial provienen de la distribución uniforme en el intervalo $[0, 1]$.

Problema 9

Aplicar a los datos del ejercicio las pruebas de correlación, gaps y poker.

Solución:

Aplicando la prueba de autocorrelación para varios rezagos se obtienen los siguientes valores de los estadísticos de prueba:



Aplicando la prueba de Gaps se obtiene el siguiente valor para la estadística de prueba:

$$G = 7.65.$$

El valor p correspondiente esta dado por

$$\text{p-value} = 0.87.$$

Aplicando la prueba del Poker se obtiene el siguiente valor para la estadística de prueba:

$$P = 14.3.$$

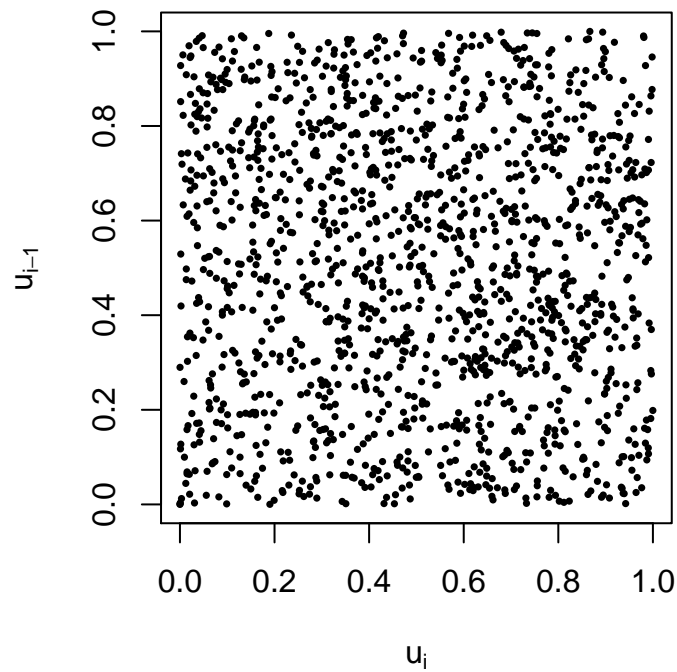
El valor p correspondiente esta dado por

$$\text{p-value} = 0.01.$$

Problema 10

Generar 1500 números del generador RANDU. Hacer una prueba de Kolmogorov-Smirnov al 95 % de confianza

```
## [1] 0.000031 0.000183 0.000824 0.003296 0.012360 0.044495 0.155732
## [8] 0.533939 0.802042 0.006802 0.822440 0.873416 0.838541 0.170501
## [15] 0.476134 0.322291 0.648545 0.990648 0.106986 0.726078
```



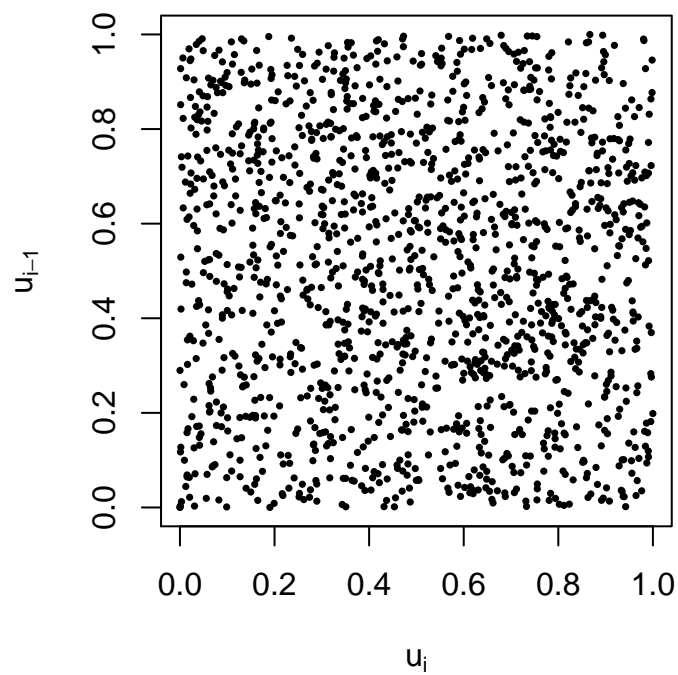
```
##
## One-sample Kolmogorov-Smirnov test
##
## data: randu0
## D = 0.03, p-value = 0.3
## alternative hypothesis: two-sided
```

Problema 10

Generar 1500 números del generador RANDU. Hacer una prueba de Kolmogorov-Smirnov al 95 % de confianza

```
-----
x
-----
0.00
0.00
0.00
0.00
0.01
```

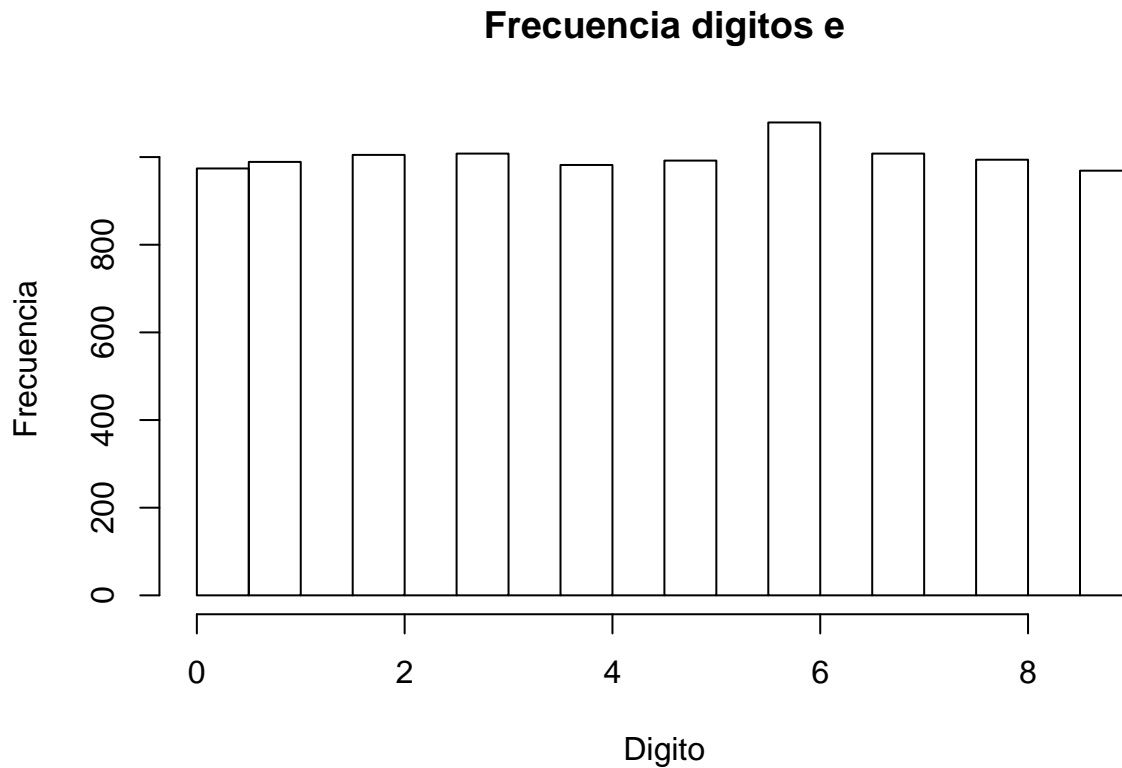
<u>x</u>
0.04
0.16
0.53
0.80
0.01
0.82
0.87
0.84
0.17
0.48
0.32
0.65
0.99
0.11
0.73



```
##
## One-sample Kolmogorov-Smirnov test
##
## data:  rand5
## D = 0.03, p-value = 0.3
## alternative hypothesis: two-sided
```

Problema 11

La página The number e to one million digits (<https://apod.nasa.gov/htmltest/gifcity/e.1mil>) contiene el primer millón de dígitos de e (pueden usar cualquier otra página). Considerando estos dígitos. Realizar un histograma y verificar la hipótesis de que los dígitos corresponden a una distribución uniforme discreta. Verificar independencia de los dígitos, considerando las pruebas de gaps, de poker y de rachas.



```
##
##          Gap test
##
## chisq stat = 365, df = 15, p-value = 1.8e-68
##
##      (sample size : 10000)
##
## length  observed freq    theoretical freq
## 1         1002         1250
## 2          628          625
## 3          325          312
## 4          221          156
## 5          102           78
## 6           68           39
## 7           55           20
## 8           24           9.8
## 9            8           4.9
## 10           9           2.4
## 11           6           1.2
```

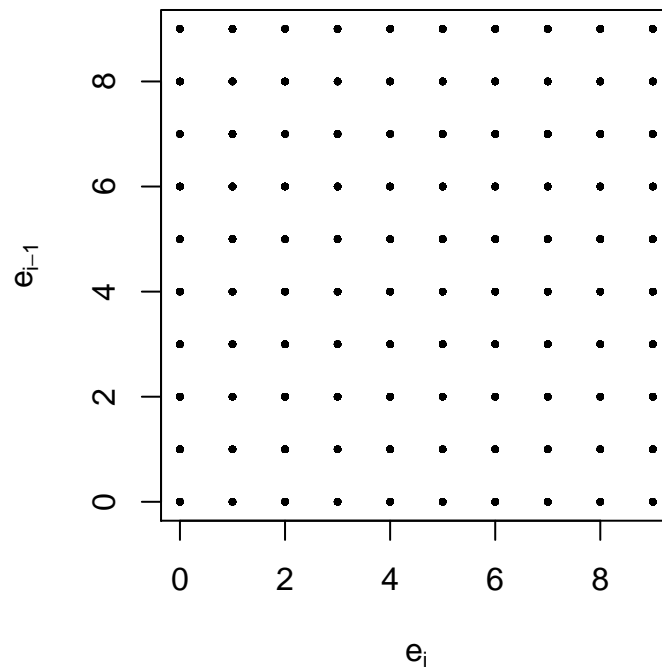
```

## 12          2          0.61
## 13          2          0.31
## 14          2          0.15
## 15          0          0.076
## 16          2          0.038

##
##          Poker test
##
## chisq stat = 3.5, df = 4, p-value = 0.48
##
##          (sample size : 10000)
##
## observed number  2 211 935 769 83
## expected number  3.2 192 960 768 77

##
## Runs Test
##
## data:  pi by 2
## statistic = 2, runs = 5000, n1 = 5000, n2 = 5000, n = 10000,
## p-value = 0.03
## alternative hypothesis: nonrandomness

```



Nota: Lo anterior se realizó considerando valores entre 0 y 0.9, con diferencias de 0.1

Problema 12

Escriban un programa que utilice el método de la transformación inversa para generar números de la densidad siguiente: $f(x) = \frac{1}{x^2}I(1 \leq x)$. Para probar el programa, hagan un histograma de 10,000 números junto con la densidad f . Verificar la hipótesis de que la muestra sigue la distribución teórica dada y hacer un qq-plot e interpretar.

