Posgrado de Ingeniería de Sistemas

Ejercicios de valor esperado y varianza en R

Alberto Benavides

9 de noviembre de 2020

P. 247, 1

A card is drawn at random from a deck consisting of cards numbered 2 *through* 10*. A player wins* 1 *dollar if the number on the card is odd and loses* 1 *dollar if the number if even. What is the expected value of his winnings?*

En una variable `ganancias` se suman los dólares que se pueden ganar según las reglas descritas del problema que se traducen en la instrucción de R: `-1 + 2 * sample(2:10, 1) %% 2`. Al realizar este experimento $r = [1, 2, \ldots, 10\ 000]$ repeticiones, se tienen promedios cercanos a $E(X) = -1/9$ como se ve en la figura 1 (p. 2).

P. 247, 6

A die is rolled twice. Let X denote the sum of the two numbers that turn up, and Y the difference of the numbers (specifically, the number on the first roll minus the number on the second). Show that $E(XY) = E(X)E(Y)$. Are X and Y independent?

Se generan 100 000 tiradas de dos dados d_1 y d_2, las que se suman en X, se restan $Y = d_1 - d_2$ y se obtiene XY. Una muestra de estos resultados se halla en la tabla 1 (p. 2).

Además, se muestran los histogramas de los valores de X, Y así como XY en la figura 2 (p. 3), donde se puede constatar que los valores esperados son $E(X) = 7$, $E(Y) = 0$, $E(XY) = 0$ y $E(X) \cdot E(Y) = 0 = E(XY)$.

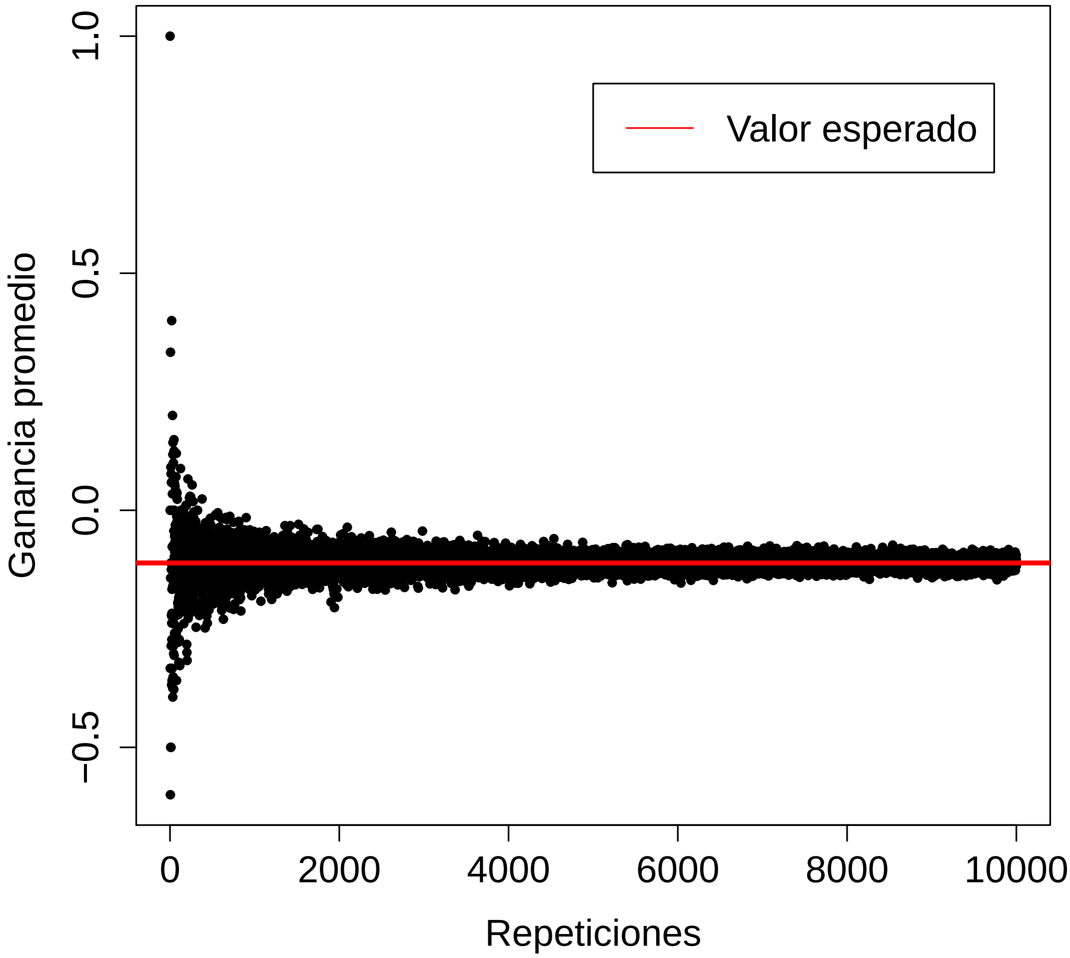

Figura 1: P. 274, 1.

Tabla 1: Muestra de resultados de p. 247, 6.

d_1	d_2	X	Y	XY
4	4	8	0	0
3	6	9	-3	-27
2	1	3	1	3
4	1	5	3	15
5	1	6	4	24

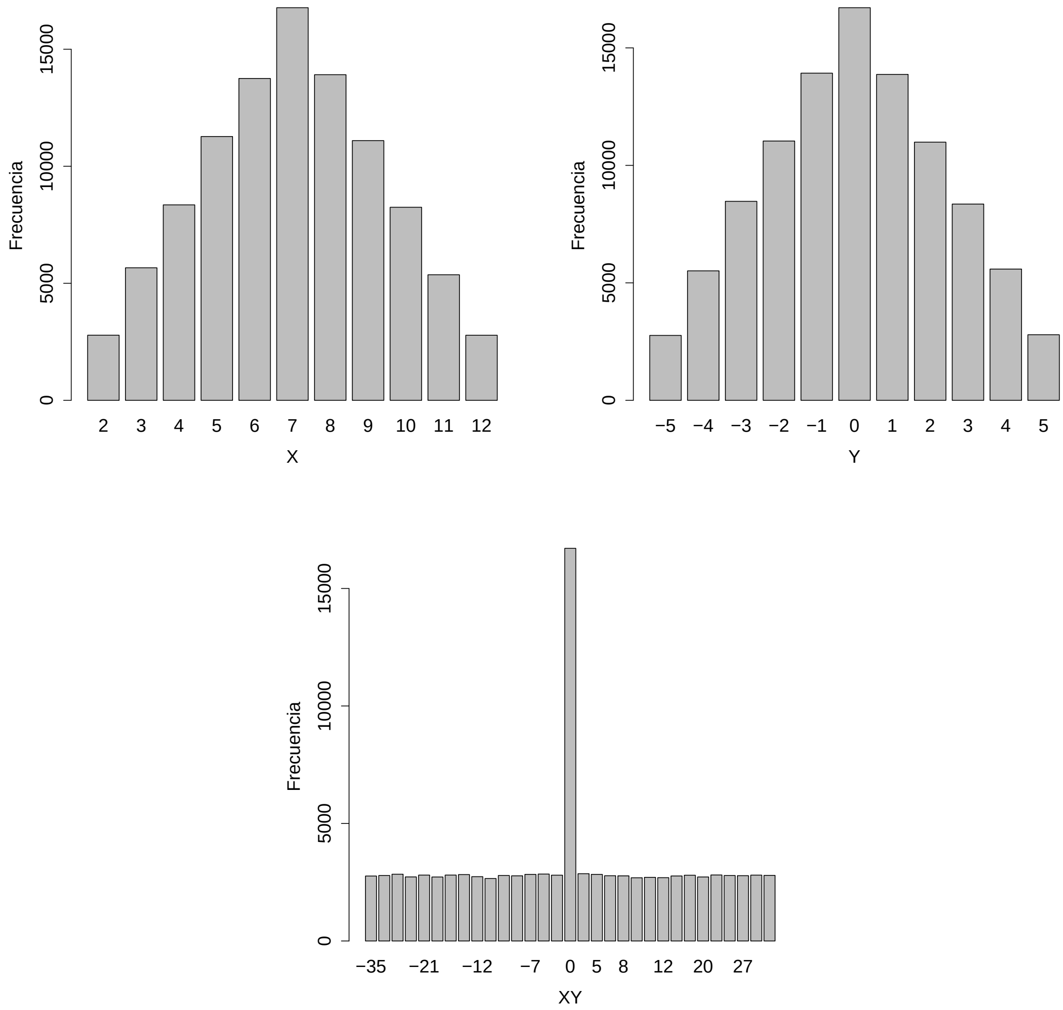

Figura 2: Histogramas de p. 247, 6.

P. 249, 15

A box contains two gold balls and three silver balls. You are allowed to choose successively balls from the box at random. You win 1 dollar each time you draw a gold ball and lose 1 dollar each time you draw a silver ball. After a draw, the ball is not replaced. Show that, if you draw until you are ahead by 1 dollar or until there are no more gold balls, this is a favorable game.

En este caso, se realiza un diseño de experimentos en el que se obtienen cincuenta promedios de 10 000 repeticiones del experimento mostrado en el código 1 entre las líneas 6 a la 24. En la línea 1 se crea la variable `caja` que contiene un vector $[-1, -1, -1, 1, 1]$ análogo a los dólares que se obtienen por extraer una bola plateada, -1, y dorada 1. Esta `caja` se desordena en la línea 7 y se simula la extracción de una en una de las bolas en la línea 10. La variable `ganancias` declarada en la línea 3 se utiliza para almacenar las ganancias (calculadas en la línea 11) del experimento descrito, de modo que cuando se va ganando por un dólar (línea 12) o cuando han salido las dos bolas de oro (contadas en la línea 15 y revisadas en la 16), se termina el juego y se almacenan las ganancias en la línea 22.

Código 1: P. 249, 15

```
1 caja = c(rep(-1, 3), rep(1, 2))
2 orden = c()
3 ganancias = c()
4 promedios = c()
5 for (k in 1:50){
6   for (i in 1:10000){
7     muestra = sample(caja)
8     ganancia = 0
9     oro = 0
10    for (j in 1:length(caja)){
11      ganancia = ganancia + muestra[j]
12      if (ganancia == 1){
13        break
14      } else if (muestra[j] == 1){
15        oro = oro + 1
16        if (oro == 2){
17          break
18        }
19      }
20    }
21    orden = c(orden, toString(caja))
22    ganancias = c(ganancias, ganancia)
23  }
24  promedios = c(promedios, mean(ganancias))
25 }
```

La variable `orden` almacena el acomodo en que se desordenan al azar las cajas de bolas. Cinco muestras de estas combinaciones y sus ganancias dadas las reglas pueden verse en la tabla 2 (p. 5). Adicionalmente, se muestra un diagrama de cajas y bigotes con los resultados en la figura 3 (p. 5).

Tabla 2: Muestra de resultados de p. 249, 15.

Orden	Ganancias (USD)
-1, -1, -1, 1, 1	1
-1, -1, -1, 1, 1	-1
-1, -1, -1, 1, 1	1
-1, -1, -1, 1, 1	1
-1, -1, -1, 1, 1	1

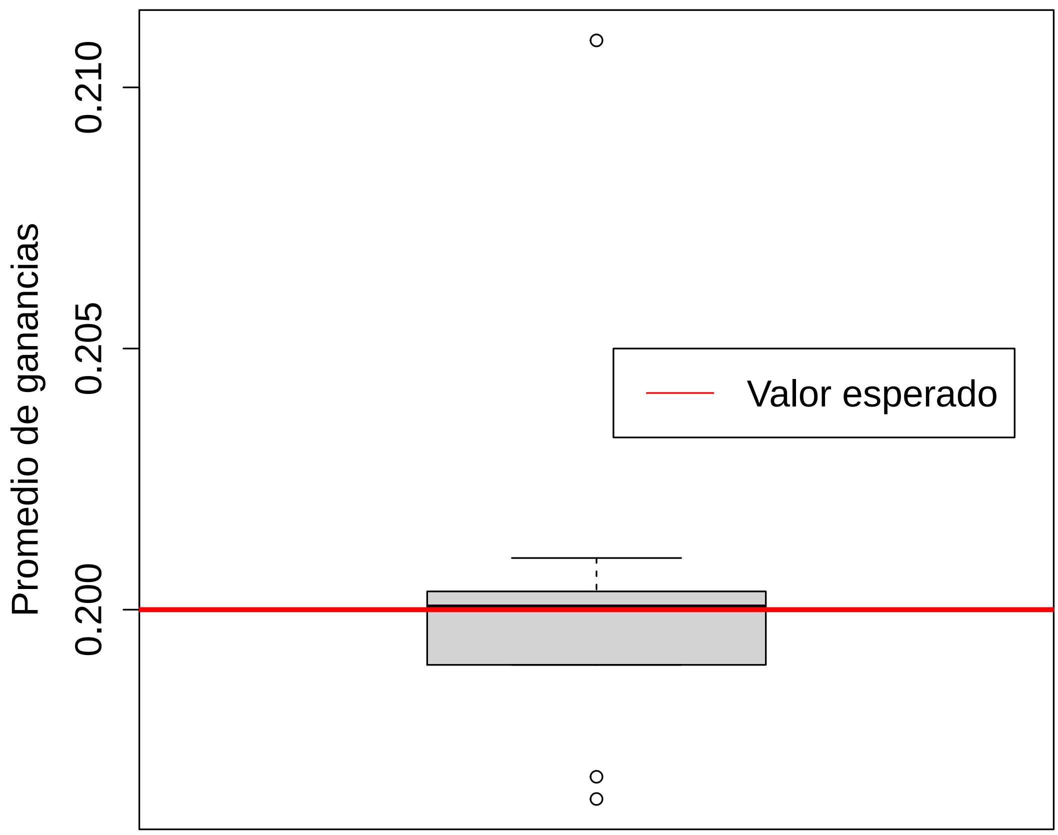

Figura 3: Diagrama de cajas y bigotes de 50 promedios de 10 000 repeticiones para el problema P. 279, 15.

P. 249, 18

Exactly one of six similar keys opens a certain door. If you try the keys, one after another, what is the expected number of keys that you will have to try before success?

Se generan 100 000 números aleatorios entre 0 y 5 que corresponden a los intentos que se harían antes de abrir la puerta mediante `sample(0:5, 100000, replace = TRUE)`. La media de esos intentos es aproximadamente 2.5, lo que coincide con su valor esperado calculado analíticamente.

P. 249, 19

A multiple choice exam is given. A problem has four possible answers, and exactly one answer is correct. The student is allowed to choose a subset of the four possible answers as his answer. If his chosen subset contains the correct answer, the student receives three points, but he loses one point for each wrong answer in his chosen subset. Show that if he just guesses a subset uniformly and randomly his expected score is zero.

Esto se puede lograr a partir del procedimiento mostrado en el código 2, mediante el cálculo del promedio (línea 1) de 100 000 repeticiones (línea 3) de la suma (línea 4) de los valores de respuestas (línea 6) de un subconjunto elegido al azar de esas respuestas (línea 7). La función `unlist` permite convertir la lista resultante de la función `lapply` en un vector al que se le puede aplicar la función `mean`.

Código 2: P. 249, 19

```
1  mean(
2    unlist(
3      lapply(1:100000,
4        function(x) {
5          sum(
6            sample(c(-1, -1, -1, 3),
7              sample(1:4)
8            )
9          )
10       }
11     )
12   )
13 )
```

P. 263, 1

A number is chosen at random from the set $S = \{-1, 0, 1\}$. Let X be the number chosen. Find the expected value, variance, and standard deviation of X.

Es posible usar las funciones `mean`, `var` y `sd` de R en una muestra de un millón de valores del arreglo $S = \{-1, 0, 1\}$, dado por la instrucción `sample(c(-1, 0, 1), 1000000, replace=TRUE)`, para calcular aproximadamente la esperanza $E^*(X)$, varianza

$V^*(X)$ y desviación estándar $D^*(X)$, en ese orden. Estos valores son $E^*(X) = -0.0007 \approx E(X) = 0$; $V^*(X) = 0.6669 \approx V(X) = 2/3$; y $D^*(X) = 0.8164 \approx D(X) = \sqrt{\frac{2}{3}}$.

P. 264, 9

A die is loaded so that the probability of a face coming up is proportional to the number on that face. The die is rolled with outcome X. Find $V(X)$ and $D(X)$.

La función `sample` de R permite asignar probabilidades a los elementos de un vector sobre el que se desean extraer muestras con el uso del atributo `prob`, de modo que `sample(1:6, prob = (1/21 * 1:6), 100000, replace = T)` devuelve 100 000 valores entre $[1, 2, 3, 4, 5, 6]$ con probabilidades $[\frac{1}{21}, \frac{2}{21}, \frac{3}{21}, \frac{4}{21}, \frac{5}{21}, \frac{6}{21}]$ respectivamente, a los cuales se les puede calcular el promedio, varianza y desviación estándar con `mean`, `var` y `sd`. Los resultados, en ese orden, son 4.3388, 2.2129, 1.4876, muy cercanos a los calculados analíticamente.

P. 278, 3

The lifetime, measure in hours, of the ACME super light bulb is a random variable T with density function $f_T(t) = \lambda^2 te^{-\lambda t}$, where $\lambda = 0.05$. What is the expected lifetime of this light bulb? What is its variance?

La probabilidad de una variable aleatoria continua x en un intervalo $[a, b]$ se calcula mediante

$$P(a \leq x \leq b) = \int_a^b f_x(X)\mathrm{d}X,$$

mientras que su valor esperado es

$$E(X) = \int x f_X(x)\mathrm{d}x$$

y la varianza

$$\mathrm{E}(X^2) = \int (x - \mu_X)^2 f_X(x)\mathrm{d}x.$$

En este problema,

$$f_T(t) = 0.05^2 te^{-0.05t}. \tag{1}$$

Esto se puede calcular en R con el uso de la función `integrate`, lo cual se realiza en el código 3, donde se definen `f` como $f_T(t)$, `g` como $E(T)$ y `h - g ^ 2` como $E(T^2)$. Posteriormente se usa `integrate` para calcular las integrales de las variables `g` y `h`, de donde se obtienen el valor esperado y la varianza en las líneas 5 y 6 respectivamente.

Código 3: P. 278, 3; solución analítica

```
1 f <- function(t) 0.05^2 * t * exp(-0.05 * t)
2 g <- function(t) t * f(t)
3 h <- function(t) t^2 * f(t)
4
5 E <- integrate(g, lower = 0, upper = Inf)$value
6 V <- integrate(h, lower = 0, upper = Inf)$value - E^2
```

Como la ecuación 1 es una función de distribución, es posible calcular las probabilidades de cualquier valor t que reciba. Así, se generan un millón de valores aleatorios con distribución $\mathcal{U}(0, 150)$ usando la instrucción `a = runif(100000, 0, 150)`, de las que se calculan sus probabilidades almacenadas en `p = f(a)`. Un histograma de la función de distribución de esta variable se encuentra en la figura 3 (p. 7). Posteriormente, se obtienen un millón de valores aleatorios del vector `a` con las probabilidades almacenadas en el vector `p` mediante la variable `s` que recibe el vector resultante de `sample(a, 100000, prob=p, replace=TRUE)`, con lo que se puede calcular tanto el valor esperado como el promedio de esos valores, obtenidos por la función `mean`, y la varianza con `var`, como se observa en el código 4.

Código 4: P. 278, 3; aproximación experimental

```
a = runif(100000, 0, 1000)
p = f(a)
s = sample(a, 100000, prob=p, replace=TRUE)
mean(s)
var(s)
```

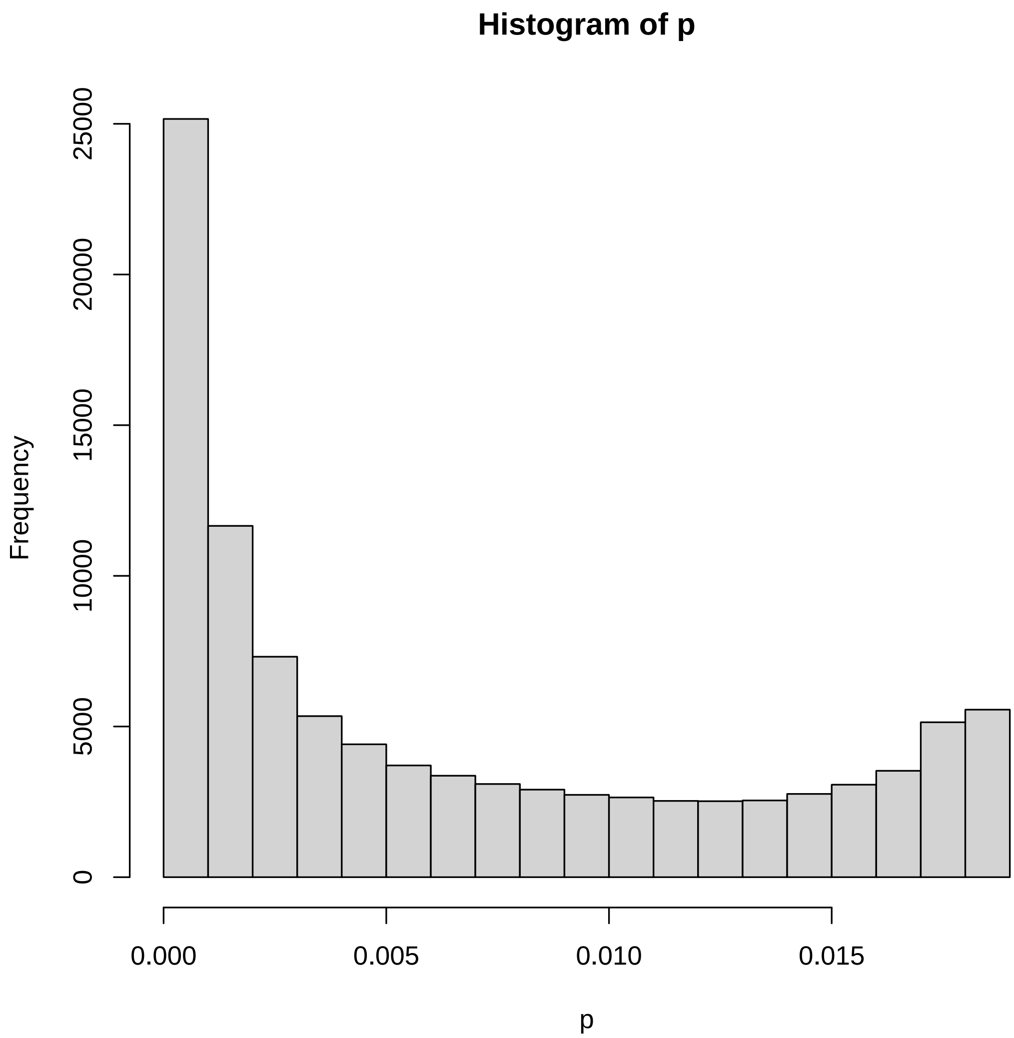

Figura 4: Histograma de la función de distribución de probabilidades dada por la ecuación 1 para un millón de variables aleatorias con distribución $\mathcal{U}(0, 150)$.