Tarea 10 de Modelos Probabilistas Aplicados

Ejercicios

5271

10 de noviembre de 2020

1. Introducción

En este documento se presentan los resultados de varios ejercicios del libro "*Introduction to Probability*"[1] encontrados de forma analítica, así como los resultados alcanzados numéricamente mediante simulación.

2. Valor esperado

Ahora para variables aleatorias discretas el valor esperado es:

$$E[X] = \sum_{x \in \Omega} x \times P(X = x). \tag{1}$$

3. Ejercicio 6 de la página 247

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?

3.1. $X = a + b$

Sea $X = a + b$ donde a es el valor del primer tiro y b el segundo, se tiene: $X = 2$–12. La frecuencia de los valores de X se muestran en el cuadro 1 de la página 2.

Cuadro 1: Frecuencia de valores de X

Posibles valores de x	**frecuencia (x)**	**$P(X = x)$**
2	1	0.028
3	2	0.056
4	3	0.083
5	4	0.111
6	5	0.139
7	6	0.167
8	5	0.139
9	4	0.111
10	3	0.083
11	2	0.056
12	1	0.028

Del cuadro 1 podemos obtener que:

$$P(X = x_i) = \frac{\text{frecuencia de } x_i}{36}. \tag{2}$$

Sustituyendo en la ecuación 1

$$\begin{aligned} E[X] &= \textstyle\sum_{i=1}^{11} x_i \times P(X = x_i) \\ E[X] &= 7. \end{aligned} \tag{3}$$

Para la comprobación y mejor entendimiento del resultado obtenido se realiza una simulación de la variable aleatoria $X = a + b$, la misma es realizada en R [2] como se muestra en el código 3.1 a continuación.

```r
restas = numeric()
sumas = numeric()
mult = numeric()
val = c(1:6)
a = 0
b = 0

for (i in c(1:1000)){
  a = sample(val,1)
  b = sample(val,1)
  r = a-b
  s = a+b
  restas= c(restas,r)
  sumas= c(sumas,s)
  mult= c(mult,r*s)
}
```

Tarea10.R

Los valores para X son almacenados en un *dataframe*, como se muestra en el cuadro 2 de la página 3, al aplicarle la función *summary* de R al *dataframe* se obtiene los siguientes valores.

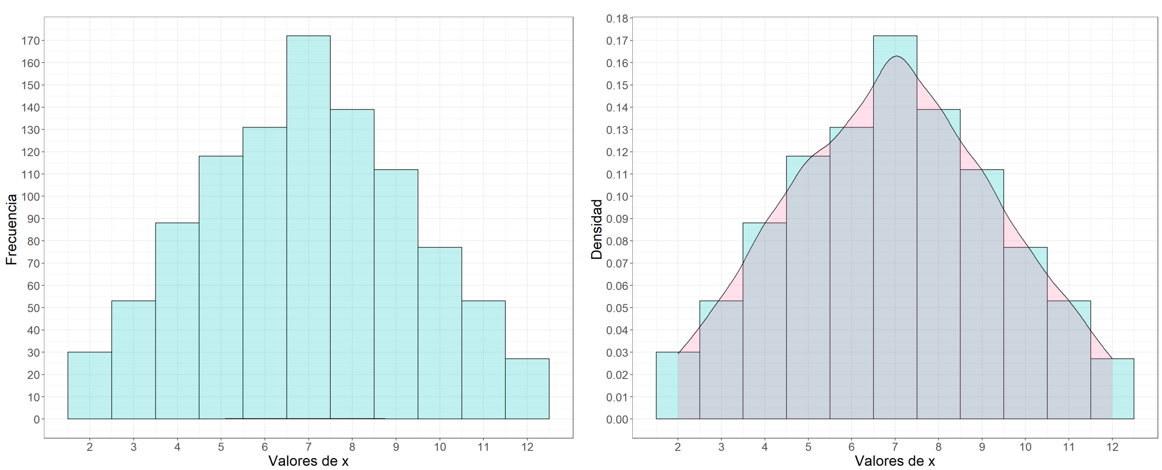
summary.txt

```
> summarydfs$x
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  2.000   5.000   7.000   6.948   9.000  12.000
```

En los valores obtenidos se puede observar que la media es igual a 6.848 por lo que se puede decir que el valor promedio obtenido es igual a la $E[X]$, además coincide con la mediana que es el valor con mayor frecuencia de ocurrencia de X como se muestra en la figura 4 de la página 9. Para mejor visualización del experimento se puede ver la figura $X = (a + b)$.gif, donde se observa el progreso de los lanzamientos.

Cuadro 2: fragmento de *dataframe* de la variable aleatoria $X = a + b$

Lanzamientos	x
1	6
2	11
3	6
4	9
5	4
161	7
162	3
163	6
164	7
165	8
996	6
997	9
998	11
999	11
1000	5



(a) Histograma de frecuencia de la variable aleatoria X

(b) Histograma de densidad de la variable aleatoria X

Figura 1: Histogramas de la variable aleatoria $X = a + b$

3.2. $Y = a - b$

Análogamente, sea $Y = a - b$. De Y se obtiene los valores que muestra el cuadro 3 de la página 4.

Cuadro 3: Frecuencia de valores de X

Posibles valores de *x*	**frecuencia *(x)***	***P(X = x)***
-5	1	0.028
-4	2	0.056
-3	3	0.083
-2	4	0.111
-1	5	0.139
0	6	0.167
1	5	0.139
2	4	0.111
3	3	0.083
4	2	0.056
5	1	0.028

Sustituyendo en la ecuación 1

$$\begin{aligned} E[X] &= \textstyle\sum_{i=1}^{11} x_i \times P(X = x_i) \\ E[X] &= 0. \end{aligned} \tag{4}$$

Al comprobar con la simulación de la variable aleatoria $X = a - b$, los valores para X son almacenados en un *dataframe*, como se muestra en el cuadro 4 de la página 5, al aplicarle la función *summary* de R al *dataframe* se obtiene los siguientes valores.

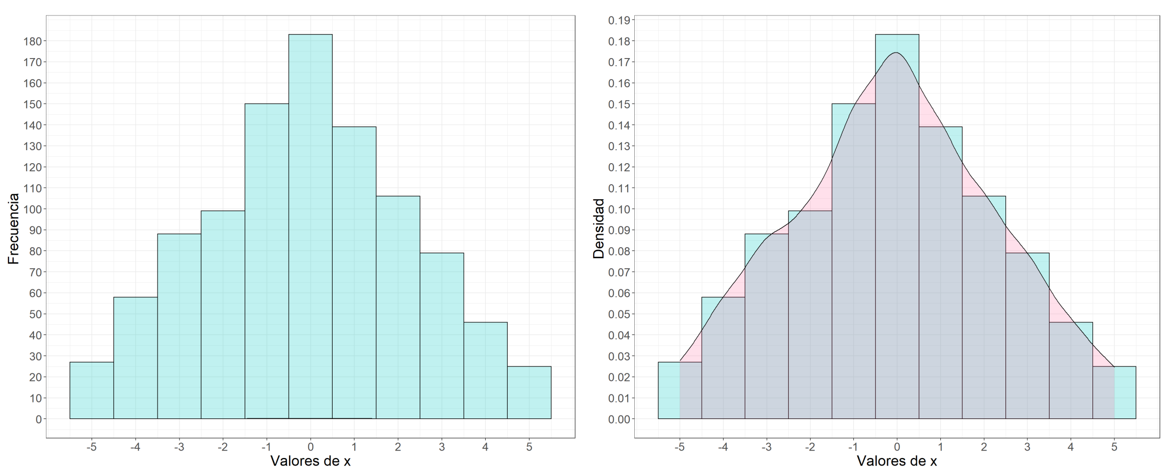
summary.txt

```
> summarydfr$x
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 -5.000  -2.000   0.000  -0.082   2.000   5.000
```

Cuadro 4: Fragmento de *dataframe* de la variable aleatoria Y = a - b

Lanzamientos	y
1	-4
2	-1
3	-2
4	-3
5	2
201	-3
202	0
203	1
204	0
205	5
996	2
997	-1
998	1
999	-1
1000	-1

En los valores obtenidos se puede observar que la media es igual a -0.082 por lo que se puede decir que el valor promedio obtenido es igual a la $E[Y]$, además coincide con la mediana que es el valor con mayor frecuencia de ocurrencia de Y como se muestra en la figura 3 de la página 7. Para mejor visualización del experimento se puede ver la figura $Y = (a - b)$.gif, donde se observa el progreso de los lanzamientos.



(a) Histograma de frecuencia de la variable aleatoria Y (b) Histograma de densidad de la variable aleatoria Y

Figura 2: Histogramas de la variable aleatoria $Y = a - b$

3.3. $E(XY) = E((a + b)(a - b))$

Dado que $E(XY) = E((a + b)(a - b))$, tenemos para XY los valores que se muestran en el cuadro 5 de la página 6.

Cuadro 5: Fragmento de valores de frecuencia de XY

Posibles valores de ***xy***	**frecuencia** ***(xy)***	$\boldsymbol{P(XY = xy)}$
-35	1	0.028
-32	1	0.028
-9	1	0.028
-8	1	0.028
-7	1	0.028
-5	1	0.028
-3	1	0.028
0	6	0.167
3	1	0.028
5	1	0.028
7	1	0.028
8	1	0.028
9	1	0.028
32	1	0.028
35	1	0.028

Sustituyendo en la ecuación 1

$$\begin{aligned} E[XY] &= \sum_{i=1}^{31} xy_i \times P(XY = xy_i) \\ E[XY] &= 0 \end{aligned} \tag{5}$$

Ahora para comprobar si X e Y son variables aleatorias independientes, hay que probar que se cumple para todos los casos se cumpla que $P(X,Y) = P(X)P(Y)$. Por lo anteriormente mencionado, con encontrar un caso donde no se cumpla la propiedad para decir que las variables aleatorias no son independientes. Como es el caso:

$$P(X = 12, Y = 0) = P(a = 6, b = 6) = \frac{1}{36}. \tag{6}$$

$$P(X = 12)P(Y = 0) = \frac{1}{36} \times \frac{1}{6} = \frac{1}{216}. \tag{7}$$

Por tanto $P(X = 12, Y = 0) \neq P(X = 12)P(Y = 0)$, lo que implica que X e Y no son variables aleatorias independientes.

Al realizar la simulación, los valores para la variable aleatoria $XY = (a + b)(a - b)$, los valores para XY son almacenados en un *dataframe*, como se muestra en el cuadro 6 de la página 7, al aplicarle la función *summary* de R al *dataframe* se obtiene los siguientes valores.

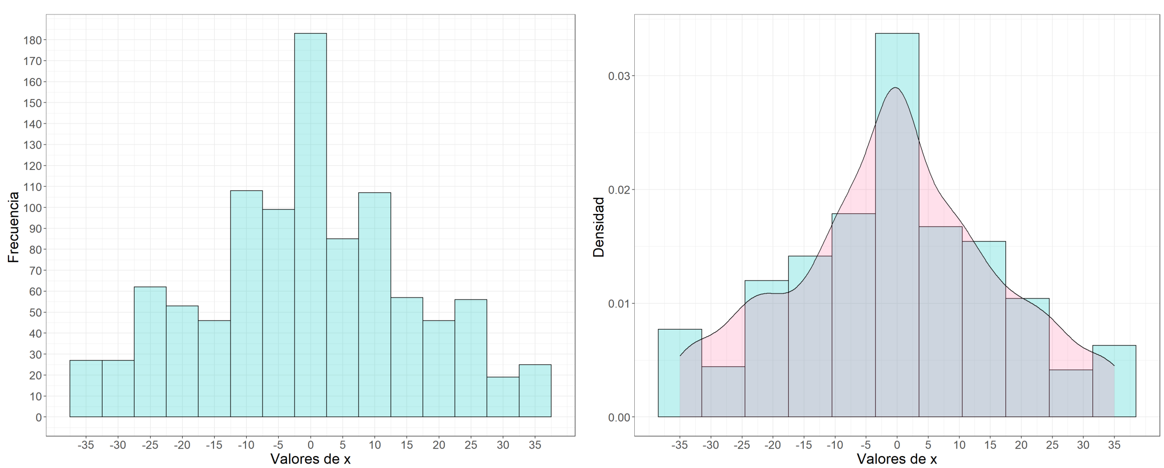
summary.txt

```
summarydfm$x
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-35.000 -11.000   0.000  -0.504  11.000  35.000
```

Cuadro 6: Fragmento de *dataframe* de la variable aleatoria $XY = (a + b)(a - b)$

Lanzamientos	xy
1	-24
2	-11
3	-12
4	-27
5	8
614	0
615	-15
616	0
617	-24
618	9
996	12
997	-9
998	11
999	-11
1000	-5

En los valores obtenidos se puede observar que la media es igual a -0.504 por lo que se puede decir que el valor promedio obtenido es similar a la $E[XY]$, además la $E[XY]$ coincide con la mediana que es el valor con mayor frecuencia de ocurrencia de XY como se muestra en la figura 3 de la página 7. Para mejor visualización del experimento se puede ver la figura $XY = (a + b)(a - b)$.gif, donde se observa el progreso de los lanzamientos.



(a) Histograma de frecuencia de la variable aleatoria XY (b) Histograma de densidad de la variable aleatoria XY

Figura 3: Histogramas de la variable aleatoria $XY = (a + b)(a - b)$

Mediante la simulación se pudo constatar que los los resultados para la esperanza

4. Ejercicio 18 de la página 249

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?

Respuesta: La variable aleatoria X viene dada por el número de intentos fallidos antes de un éxito. Por lo que se tienen los valores en el cuadro 7 de la página 8.

Cuadro 7: Probabilidades de ocurrencia de x

# de intentos (x)	$P(X = x)$
0	$1/6$
1	$(5/6) \times (1/5) = 1/6$
2	$(5/6) \times (4/5) \times (1/4) = 1/6$
3	$(5/6) \times (4/5) \times (3/4) \times (1/3) = 1/6$
4	$(5/6) \times (4/5) \times (3/4) \times (2/3) \times (1/2) = 1/6$
5	$(5/6) \times (4/5) \times (3/4) \times (2/3) \times (1/2) \times 1 = 1/6$

Entonces se sustituye en la ecuación 1:

$$\begin{aligned} E[X] &= 0 \times \left(\tfrac{1}{6}\right) + 1 \times \left(\tfrac{1}{6}\right) + 2 \times \left(\tfrac{1}{6}\right) + 3 \times \left(\tfrac{1}{6}\right) + 4 \times \left(\tfrac{1}{6}\right) + 5 \times \left(\tfrac{1}{6}\right) \\ E[X] &= \left(\tfrac{5}{2}\right). \end{aligned} \tag{8}$$

Por lo que el número esperado de intentos antes de abrir la puerta es 2.5.

Para la comprobación y mejor entendimiento del resultado obtenido se realiza una simulación de la variable aleatoria X, la misma es realizada por el código 5 a continuación.

```
llaves = c(1:6)
correcta= 2
inten = 0
intentos= numeric()
for (k in c(1:1000)){
  for (j in c(1:6)){
    int = sample(llaves ,1)
    if(int ==correcta){
    inten= j-1
      }
  }
  intentos= c(intentos ,inten)
}
```

Tarea10.R

Los valores para X son almacenados en un *dataframe*, como se muestra en el cuadro **??** de la página **??**, al aplicarle la función *summary* de R al *dataframe* se obtiene los siguientes valores.

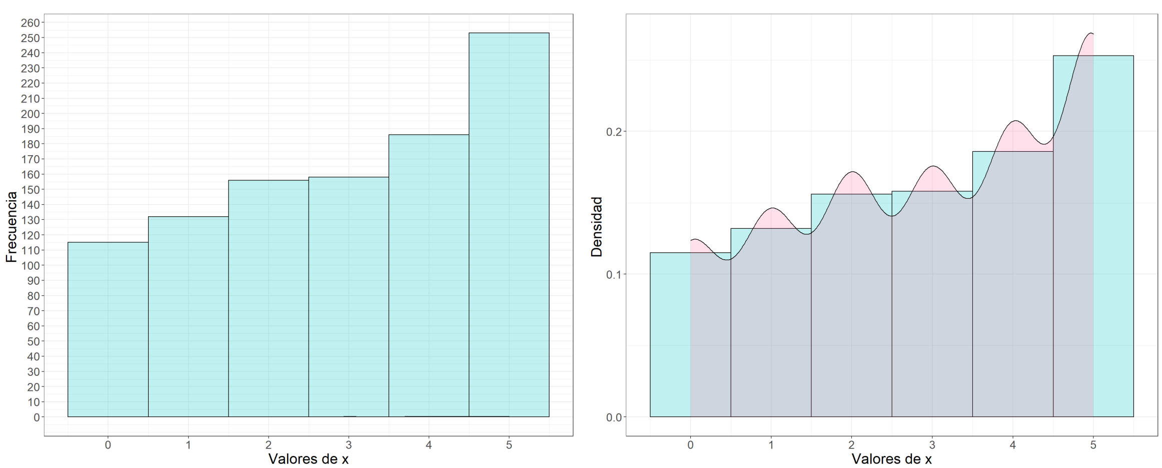
summary.txt

```
> summaryintentos
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.000   2.000   3.000   2.927   5.000   5.000
```

Cuadro 8: Fragmento de *dataframe* con valores de X obtenido en los experimentos

Experimentos	X
1	3
2	5
3	3
4	3
5	5
177	5
178	0
179	0
180	1
181	1
996	3
997	3
998	2
999	4
1000	5

En los valores obtenidos se puede observar que la media es igual a 2.927 por lo que se puede decir que el valor promedio obtenido no es igual a la $E[X]$, además no coincide con la mediana ni con el valor de valor con mayor frecuencia de ocurrencia de X que es cinco intentos. Lo anterior se muestra en la figura 4 de la página 9.



(a) Histograma de frecuencia de la variable aleatoria X

(b) Histograma de densidad de la variable aleatoria X

Figura 4: Histogramas de la variable aleatoria X

Con los valores numéricos obtenidos en la simulación se llega a una contradicción a los resultados obtenidos de forma analítica.

5. Ejercicio 19 de la página 249

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.

Respuesta: Sea la puntuación obtenida en la elección del subconjunto de posibles respuestas una variable aleatoria X, se tiene que el espacio muestral está conformado por dieciséis formas de escoger un subconjunto. En el cuadro 9 de la página 10 se muestra la cardinalidad de los posibles subconjuntos, así como los posibles valores de x y las probabilidades asociadas a dichos valores.

Cuadro 9: Cardinalidad de los subconjunto con sus posibles valores y probabilidades

Elementos de Subconjuntos	**Posibles Valores de x**	$P(X = x)$
0	0	1/16
1	3	1/16
	-1	3/16
2	2	3/16
	-2	3/16
3	1	3/16
	-3	1/16
4	0	1/16

Sustituyendo en la ecuación 1:

$$\begin{aligned} E[X] &= 3 \times \left(\tfrac{1}{16}\right) - 1 \times \left(\tfrac{3}{16}\right) + 2 \times \left(\tfrac{3}{16}\right) - 2 \times \left(\tfrac{3}{16}\right) + 1 \times \left(\tfrac{3}{16}\right) - 3 \times \left(\tfrac{1}{16}\right) \\ E[X] &= 0. \end{aligned} \tag{9}$$

Entonces si solo adivina un subconjunto de manera uniforme y aleatoria, su puntuación esperada es cero.

Para la comprobación y mejor entendimiento del resultado obtenido se realiza una simulación de la variable aleatoria X, la misma es realizada por el código 5 a continuación.

```
resp = c(0,0,1,0)
dos =numeric()
esperado = numeric()
for (k in c(1:1000)){
  sub = numeric()
  uno=sample(resp,1)
  if(uno == 1){
    sub=c(sub, 3)
  }else {
    sub=c(sub, -1)
  }
```

```
  d=sample(resp,2)
  dos=c(d)
  for (i in c(1:2)) {
    if(dos[i] == 1){
      sub=c(sub, 2)
    }else {
      sub=c(sub, -2)
    }
  }
  t=sample(resp,3)
  tres=c(t)
  for (i in c(1:3)) {
    if(tres[i] == 1){
      sub=c(sub, 1)
    }else {
      sub=c(sub, -3)
    }
  }

  esperado= c(esperado,sub)
}
```

Tarea10.R

Los valores para X son almacenados en un *dataframe*, como se muestra en el cuadro **??** de la página **??**, al aplicarle la función *summary* de R al *dataframe* se obtiene los siguientes valores.

summary.txt

```
summaryesperado
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
 -3.000  -3.000  -2.000  -1.335  -1.000   3.000 
```

Cuadro 10: Fragmento

Experimentos	x
1	3
2	-2
3	-2
4	-3
5	-3
6	-3
7	3
8	-2
9	-2
10	1
5996	-2
5997	-2
5998	-3
5999	-3
6000	1

En los valores obtenidos se puede observar que la media es igual a -2 por lo que se puede decir que el valor promedio es negativo por lo que se puede decir que va a obtener cero en el examen. El código general se encuentra disponible en el repositorio. https://github.com/Albertomnoa/Tareas

Referencias

[1] Grinstead, Charles M., Snell, J. Laurie. *Introduction to Probability.* American Mathematical Society, 2006.

[2] R Core Team. R: Un lenguaje y un entorno para la informática estadística, 2020.