Valor esperado y varianza
Ejercicios
Simulación

Fabiola Vázquez

9 de noviembre de 2020

---

1. Introducción

En la tarea anterior [4] fueron realizados diversos ejercicios del libro de Grinstead y Snell [1]. El objetivo de este trabajo es realizar una simulación de algunos de estos ejercicios, dicha simulación se realiza con el software R [3] en un cuaderno de Jupyter [2].

2. Ejercicios

Ejercicio 1 (Ej. 1, p. 247)**.** *Se saca una carta al azar de una baraja que consiste de cartas numeradas del 2 al 10. Un jugador gana un dólar si el número en la carta es impar y pierde un dólar si el número es par. ¿Cuál es el valor esperado de sus ganancias?.*

Simulación. Se considera la variable `carta` como un valor entre 2 y 10 y se verifica si dicho valor es divisible entre dos y se le asigna una ganancia de -1, en caso contrario se asigna una ganancia de 1. El proceso se repite una cantidad `t` de veces para calcular la media y ver si se aproxima al valor esperado, esto se automatiza con la función `cartas_funcion`.

```
cartas_funcion <- function(t){
    m <- c()
    for (j in 1:t){
    carta<-sample(x = 2:10,size = 1)
        if(carta %% 2==0){
            ganancia <- -1
        } else{ganancia <-1}
        m[j] <- ganancia
    }
return(mean(m))
}
```

El experimento se corre variando `t` en $\{10, 100, 1000, 10000, 100000, 1000000\}$ y para cada uno de los valores se replica 100 veces. En la figura 1 se muestra los gráficos de caja con los datos recopilados, la

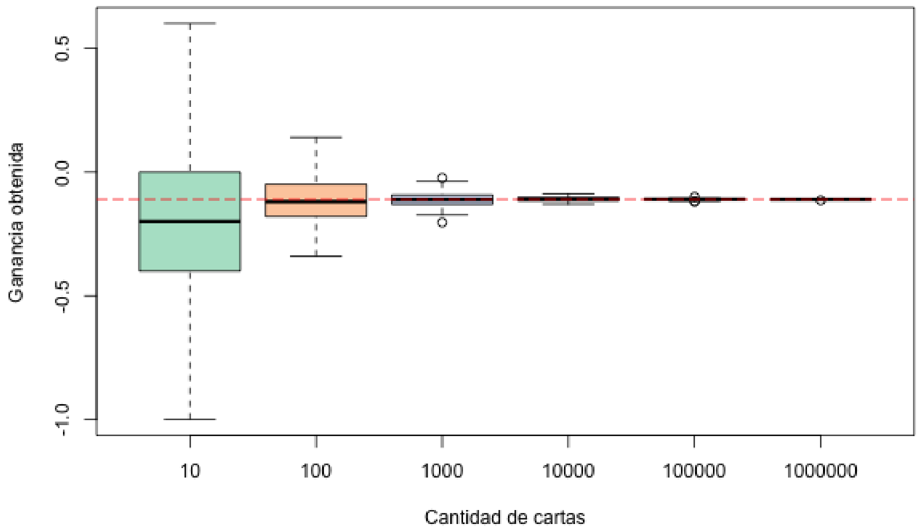

Figura 1: Gráficas de caja con las medias del ejercicio 1.

recta horizontal roja está ubicada en el valor $-\frac{1}{9}$, el cual es el valor encontrado analíticamente. Como se aprecia en dicha gráfica, entre mayor es el valor de `t` la media se aproxima al valor de $-\frac{1}{9}$.

□

Ejercicio 2 (Ej. 6, p.247)**.** *Se lanza un dado dos veces. Sea X la suma de los dos números que aparecen, y sea Y la diferencia de los números (específicamente, el número que aparece primera tirada menos el número de la segunda). Demuestra que $E(XY) = E(X)E(Y)$. ¿Son X e Y independientes?.*

Simulación. Primero, se va a estimar el valor esperado de la variable X, para ello se definen dos variables `dado1` y `dado2` como un valor pseudo aleatorio entre 1 y 6, y la variable `dado` se define como la suma de los otros dos. Se repite una cantidad `t` de veces para calcular la media y ver si se aproxima al valor esperado, esto se automatiza con la función `dados_funcion`.

```
dado_funcion <- function(n){
    m <- c()
    for (j in 1:n){
    dado1 <- sample(x = 1:6, size = 1, replace = TRUE)
    dado2 <- sample(x = 1:6, size = 1, replace = TRUE)
    dado <- dado1+dado2
    m[j] <- dado
    }
return(mean(m))
}
```

Al igual que el experimento anterior, este se corre variando `t` en $\{10, 100, 1000, 10000, 100000, 1000000\}$ y para cada uno de los valores se replica 100 veces. La figura 2 muestra los gráficos de caja con los datos obtenidos, la recta roja está ubicada en el valor 7, el cual es el valor de $E(X)$, encontrado analíticamente. Como se aprecia en dicha gráfica, entre mayor es el valor de `t` la media se aproxima al valor esperado encontrado. Similarmente, se estima el valor de $E(Y)$, para ello se utiliza la función `dados_resta`.

```
dado_resta <- function(n){
```

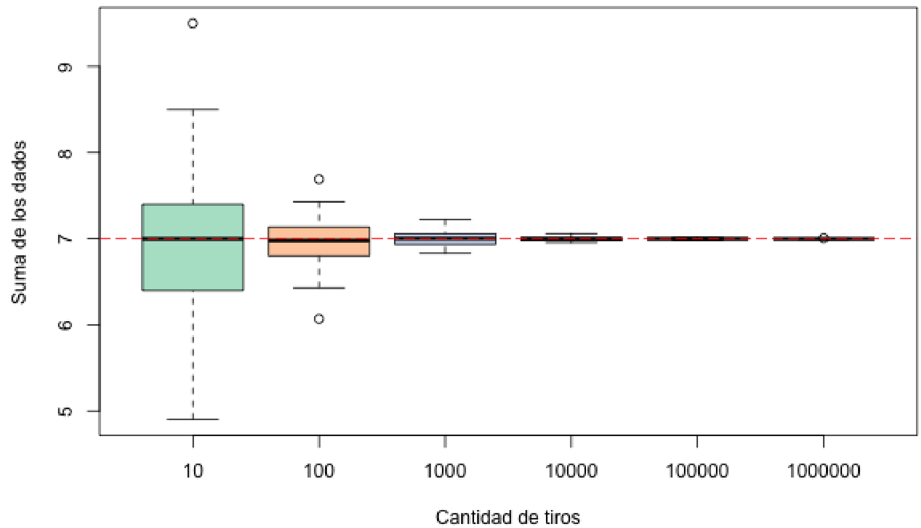

Figura 2: Gráficas de caja con las medias de la suma de dados del ejercicio 2.

```
    m <- c()
    for (j in 1:n){
    dado1 <- sample(x = 1:6,size = 1,replace = TRUE)
    dado2 <- sample(x = 1:6,size = 1,replace = TRUE)
    dado <- dado1-dado2
    m[j] <- dado
    }
return(mean(m))
}
```

La figura 3 muestra los gráficos de caja obtenidos, donde se puede apreciar que a mayor valor de `t` la media de aproxima a 0, el cual es el valor encontrado analíticamente. Por último se estima el valor esperado de la variable XY, lo cual se automatiza con la función `dado_multiplicacion`.

```
dado_multiplicacion <- function(n){
    m <- c()
    for (j in 1:n){
    dado1 <- sample(x = 1:6,size = 1,replace = TRUE)
    dado2 <- sample(x = 1:6,size = 1,replace = TRUE)
    dado <- dado1^2-dado2^2
    m[j] <- dado
    }
return(mean(m))
}
```

La figura 4 muestra los gráficos de caja obtenidos, donde se puede apreciar que a mayor valor de `t` la media se aproxima a 0, el cual es el valor encontrado analíticamente.

□

Ejercicio 3 (Ej. 15, p.249). *Una caja contiene dos balones dorados y tres plateados. Se te permite tomar sucesivamente balones de la caja al azar. Ganas un dólar cada vez que tomas un balón dorado y pierdes*

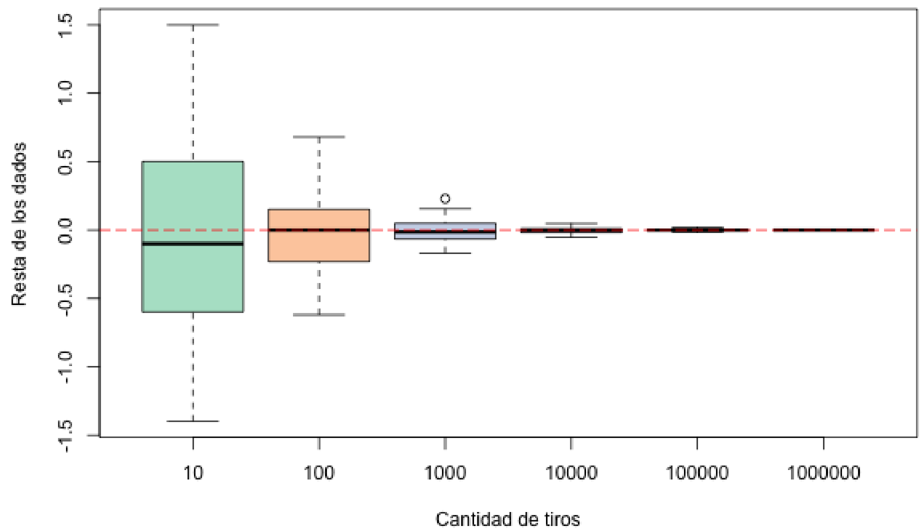

Figura 3: Gráficas de caja con las medias de la resta de dados del ejercicio 2.

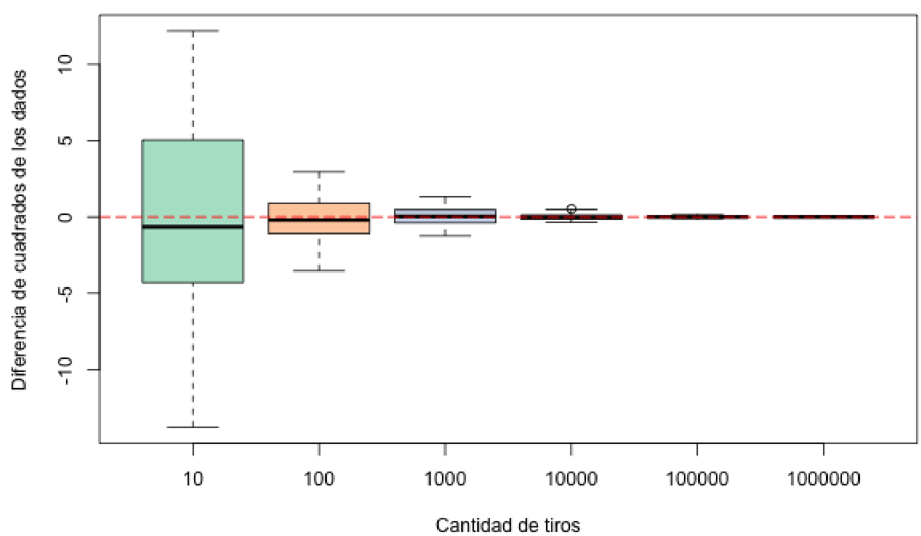

Figura 4: Gráficas de caja con las medias de la multiplicación de dados del ejercicio 2.

un dólar cada vez que tomas un balón plateado. Después de tomar, el balón no es reemplazado. Demuestra que, si tomas hasta que estás por delante por un dólar o hasta que no hay más balones dorados, este es un juego favorable.

Simulación. Para simular este comportamiento del juego se utiliza la función `balones`, en la cual se representa los balones dorados con 1, ya que es la ganancia que se obtiene al sacar dicho balón y del mismo modo, se representa con 1 a la extracción de un balón plateado.

En la figura 5 se muestran las medias obtenidas al aumentar las réplicas del experimento. Como se aprecia en dicha figura, entre mayor número de réplicas la media se aproxima al valor de 0.2 que es el que se obtuvo analíticamente.

```
balones <- function(n) {
    gan <- c()
    for(j in 1:n){
        ganancia=0
        dorados=0
        balon<-sample(c(1,1,-1,-1,-1))
        for (i in 1:5){
            if(balon[1]==1){
                ganancia<-1
                break

            }
            ganancia=ganancia+balon[i]
            if(balon[i]==1){
                dorados=dorados+1
            }

            if(ganancia==1){
                break
            }
            if(dorados==2){
                break
            }
        }
         gan[j] <- ganancia
    }
    return(mean(gan))
}
```

□

Ejercicio 4 (Ej.18, p.249). *Exactamente una de seis llaves similares abre una cierta puerta. Si pruebas las llaves, una después de otra, ¿cuál es el número esperado de llaves que deberá probar antes de tener éxito?.*

Simulación. Para este ejercicio, se utiliza un ordenamiento pseudo al azar del vector $(1, -1, -1, -1, -1, -1)$, donde la llave que sí abre la puerta se representa con un 1 y las demás con un -1, después se contabilizan los intentos antes del éxito. La función `llave` guarda los intentos obtenidos en cada iteración y al final regresa la media de dichos intentos.

```
llave <- function(n) {
    int <- c()
    for (i in 1:n){
```

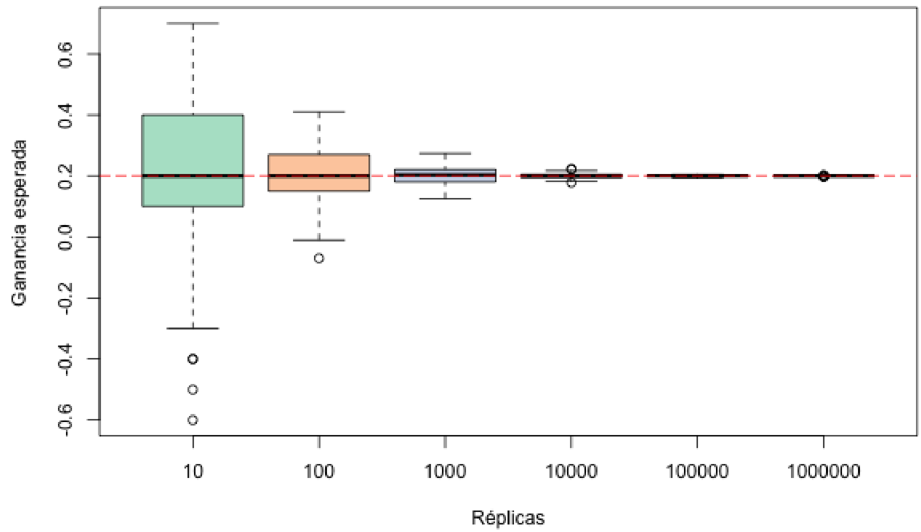

Figura 5: Gráficas de caja con las medias del ejercicio 3.

```
        intentos=0
        llaves<-sample(c(1,-1,-1,-1,-1,-1))
        for (j in 1:6){
            if (llaves[j]==1){
                break
            } else {intentos=intentos+1}
    }
     int[i] <- intentos
        }
     return(mean(int))
}
```

En la figura 6 se muestran las medias obtenidas al aumentar el número de réplicas, y dichas medias se aproximan al valor 2.5 que es el obtenido de forma analítica. □

Ejercicio 5 (Ej. 19, p.249)**.** *Se tiene un examen de opción múltiple. Un problema tiene cuatro posibles respuestas, y exactamente una es correcta. Se le permite al estudiante escoger un subconjunto de las cuatro posibles respuestas como su respuesta. Si escoge un conjunto que contiene la respuesta correcta, el estudiante recibe tres puntos, pero pierde un punto por cada respuesta incorrecta en el subconjunto escogido. Demuestra que si solo adivina un subconjunto de manera uniforme y aleatoria, su puntuación esperada es cero.*

Simulación. Se representa la solución correcta como un 1 y a una solución incorrecta como -1, entonces las posibles respuestas se representa con el vector $(1, -1, -1, -1)$. El estudiante puede elegir 0,1,2,3 o 4 respuestas, entonces se usa `n` como un número pseudo aleatorio entre 0 y 4 y se eligen `n` elementos del vector inicial. Si 1 forma parte del subconjunto, gana 3 puntos. Este proceso de elección se automatiza con la función `examen`, la cual repite el proceso `k` veces para obtener el promedio del puntaje obtenido.

```
examen <- function(k){
   m <- c()
   for (j in 1:k){
       respuestas <- c(1,-1,-1,-1)
```

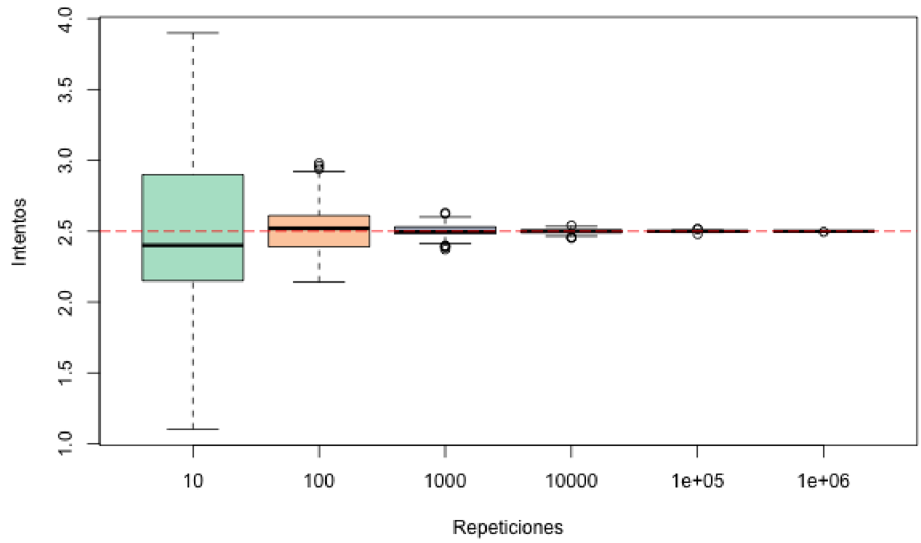

Figura 6: Gráficas de caja con las medias del ejercicio 4.

```
        n <- sample(0:4,1)
        eleccion <- sample(respuestas,n)
        puntaje <- 0
        if (n !=0){
            for (i in 1:n){
                if (eleccion[i]==1){
                    puntaje = puntaje + 3
                } else {puntaje = puntaje - 1}
            }
        }
        m[j] <- puntaje
    }
    return(mean(m))
}
```

La figura 7 muestra las medias obtenidas y como al aumentar el número de repeticiones, la media se aproxima al 0, que es el valor obtenido analíticamente. □

Referencias

[1] Charles M. Grinstead and J. Laurie Snell. *Introduction to Probability.* 2006.

[2] Thomas Kluyver, Benjamin Ragan-Kelley, Pérez, et al. Jupyter notebooks—a publishing format for reproducible computational workflows. In *Positioning and Power in Academic Publishing: Players, Agents and Agendas: Proceedings of the 20th International Conference on Electronic Publishing*, page 87. IOS Press, 2016.

[3] R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2020.

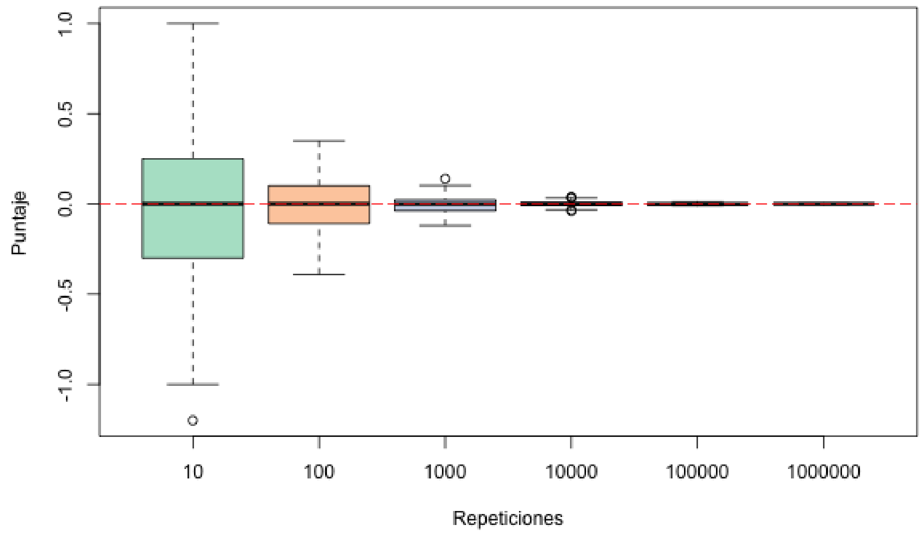

Figura 7: Gráficas de caja con las medias del ejercicio 5.

[4] Fabiola Vázquez. Valor esperado y varianza. https://github.com/fvzqa/Metodos-Probabilisticos/blob/master/Tarea9/Tarea9.pdf.