Probabilidad aplicada y simulación estocástica

Práctica # 8 Ley de los grandes números

14/Octubre/2021

alumnolcd42

A continuación dos versiones de la ley de los grandes números:

1. Versión débil

Teorema.(Ley débil de los grandes números)
Sean $X_1, X_2, ...$ variables aleatorias independientes e idénticamente distribuidas (iid) con media μ. Entonces

$$\frac{1}{N}\sum_{i=1}^{N} X_i \xrightarrow{p} \mu$$

2. Versión fuerte

Teorema.(Ley fuerte de los grandes números)
Sean $X_1, X_2, ...$ variables aleatorias independientes e idénticamente distribuidas (iid) con media μ. Entonces

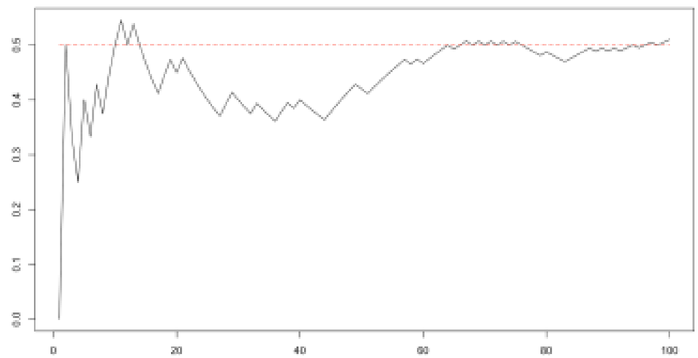
$$\frac{1}{N}\sum_{i=1}^{N} X_i \xrightarrow{c.s} \mu$$

Observación: La *ley débil* establece la convergencia en probabilidad y la *ley fuerte* dice que la convergencia es casi segura (c.s). La ley fuerte implica la ley débil.

3. Ejercicios

3.1. Tiempo de Convergencia

Considera las variables $Bernoulli(p)$ y $Pareto(a, b)$. Realiza con parámetros de tu elección y con $N = 100, 1000$, una gráfica de los promedios parciales; agrega también una línea punteada en $y = \mu$. Como la siguiente figura:



Para una $Bernoulli(0{,}5)$, ¿Con qué n, el error absoluto es menor $1e - 5$? Para una $Pareto(1{,}98, 6)$ ¿Con qué n, el error absoluto es menor $1e - 3$?
¿A qué crees que se debe que la Pareto tarda más en converger?

3.2. Estimación

Se tiran 7 dados idénticos, justos e independientes.
Sea X la suma de los dados.
Usen la Ley de los Grandes Números para estimar $P(X = k)$, $7 \leq X \leq 42$
¿Cómo aseguras que tu aproximación es buena?

Nota: No olvides poner el número de alumno en Moodle, y si desean poner su nombre que sea empezando por el apellido paterno pues así esta en la lista.

4. Solución ejercicios

4.1. Tiempo de convergencia

Lo primero que vamos a hacer es una función en la cual se calculan n veces las variables aleatorias *Bernoulli* y *Pareto* para después dividir la suma de todos los resultados entre el número de veces que se hizo, por lo que vamos a simular la **Ley de los grandes números** por lo que se esperaría que tiendan a una constante mientras mayor es la n, por lo tanto también vamos a hacer una gráfica para ver a que número tiende mientras mayor es la n así mismo vamos a graficar su media, ya que esperamos que tiendan hacia allí.

```
Ejer1 <- function(n,p,err){
  y = c()
  sum = 0
  for (i in 1:n){
    sum = sum + rbern(1,p)
    y[i] = sum/i
    if (y[i]-p <= err){
      N = i
    }
  }
  plot(y,type="l",main="Sumas parciales Bernoulli",xlab="Ensayos",ylab="Sum par")
  abline(h=p,col="blue")
  return(N)
}
res = Ejer1(1000,0.8,1e-5)
res
```

Figura 1: Código para ver cuando converge *Bernoulli*

```
Ejer2 <- function(n,a,b,err){
  y = c()
  sum = 0
  for (i in 1:n){
    sum = sum + rPareto(1,a,b)
    y[i] = sum/i
    if (y[i]-(a*b)/(b-1) <= err){
      N = i
    }
  }
  plot(y,type="l",main="Sumas parciales Pareto",xlab="Ensayos",ylab="Sum par")
  abline(h=(a*b)/(b-1),col="blue")
  return(N)
}
```

Figura 2: Código para ver cuando converge *Pareto*

Primero vamos a hacer la variable *Bernoulli* con $p = 0{,}5$ para $n = 100$ y $n = 1000$ con lo que obtenemos las siguientes gráficas

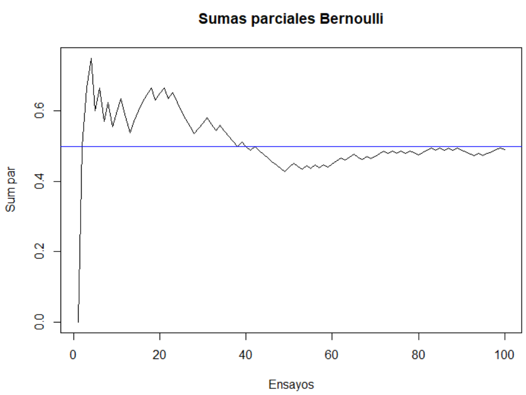


Figura 3: Convergencia de *Bernoulli* para $n = 100$

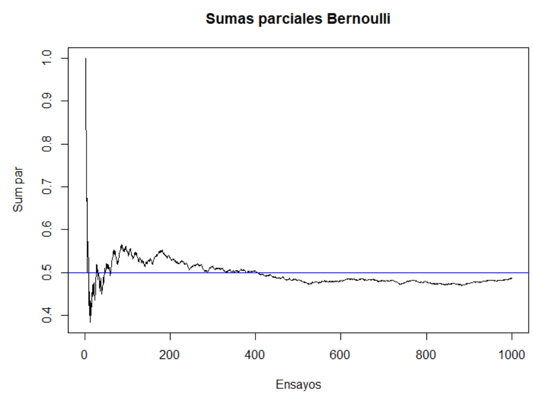


Figura 4: Convergencia de *Bernoulli* para $n = 1000$

Como podemos ver a partir de las gráficas cuando n es pequeña hay una gran diferencia entre los valores obtenidos contra los esperados, pero conforme n va aumentando la gráfica se acerca más a lo que esperaríamos, y ahora por la ley de los grandes números suponemos que cuando n sea todavía mayor a lo graficado ,la recta de μ con la de sumas parciales si irían acercando más.

Ahora queremos encontrar la n con la que obtenemos que el error entre la media y las sumas parciales es menor a 1 e-5, esto lo podemos hacer con la misma función con la que creamos la gráfica, ya que le pedimos que nos regrese el n mínimo con lo que obtuvimos el siguiente resultado.Tenemos que una *Bernoulli(0.5)* el error es menor a lo deseado en 792(este resultado cambia por cada vez que se llama, pero esta fue la media de los intentos). Por lo que podemos decir que se requieren 660 intentos para que haya una diferencia menor a 1 e-5, la cual es una muy buena aproximación.
Ahora haremos el mismo proceso pero para una variable *Pareto* con $a = 1{,}98$ y $b = 6$

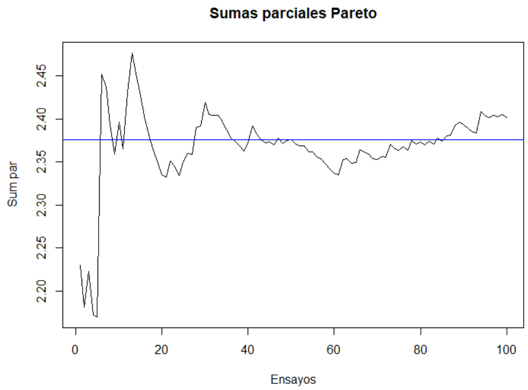


Figura 5: Convergencia de *Pareto* para $n = 100$

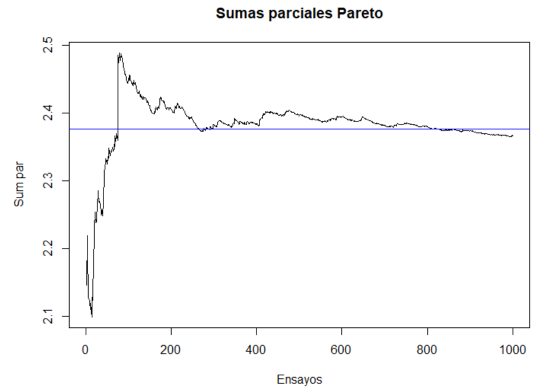


Figura 6: Convergencia de *Pareto* para $n = 1000$

A partir de las gráficas logramos observar que a n pequeñas hay una gran oscilación alrededor de su media la cual viene dada por $\mu = \frac{ab}{b-1}$ pero conforme esta va aumentando tiende mucho más a su media, esto se logra ver bastante bien para los últimos valores de la gráfica con $n = 1000$ en donde parece que el error es bastante pequeño, y de nuevo podemos suponer que por la ley de los grandes números este error ira disminuyendo conforme aumentamos el número de intentos.

Ahora queremos encontrar la n con la que obtenemos que el error entre la media y las sumas parciales es menor a 1 e-6, esto lo podemos hacer con la misma función con la que creamos la gráfica, ya que le pedimos que nos regrese el n mínimo con lo que obtuvimos el siguiente resultado.Tenemos que una *Pareto(1.98,6)* el error es menor a lo deseado en 9149(este resultado cambia por cada vez que se llama, pero esta fue la media de los intentos), por lo que podemos decir que se requieren 660 intentos para que haya una diferencia menor a 1 e-6, la cual es una muy buena aproximación.

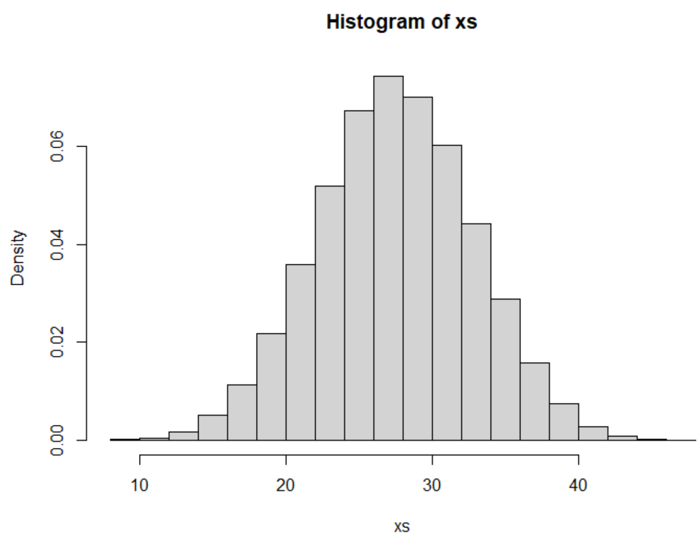
Notamos que la *Pareto* se tarda más en converger la *Bernoulli* esto se puede deber a como estan definidas las variables, ya que en una distribución de *Bernoulli(0.5)* todos los valores estan muy cercanos en el intervalo (0,1), pero en cambio en la de *Pareto(1.98,6)* hay mayor varianza en los valores obtenidos, y en un mayor intervalo.Por lo que esta se tardara más en converger a la media.

4.2. Estimación

Ahora queremos tirar 7 dados y sumar sus caras, para estimar cual es la probabilidad de que se sume cierto número.
Para eso hacemos el sigueinte programa

```
Ejer3 <- function(n,k){
  xs <- c()
  for (i in 1:n){
    xs[i] <- sum(sample(1:7,7,replace=T))
  }
  res = table(xs)/n
  print(res[k])
}
Ejer3(1000,"32")
```

En la cual le vamos a dar el número que estamos buscando y nos va a decir la probabilidad de que este caiga, sabemos que por la ley de los grandes números mientras mayor sea la n más va a converger a la media,por lo cual vamos a obtener mejores resultados a mayor sea la n por lo cual la haremos igual a 100000.
Esperaríamos que hay números que sean más probables que otros por ejemplo el 42 solo se logra si todos los dados salen 6, en cambio hay varias opciones para que salgan números como 34.
Para ver si esto es cierto vamos a hacer un histograma de los valores posibles



Podemos ver que en efecto los valores centrales salen muchas más veces que los extremos, de el histograma se parece mucho al que se obtendría con una variable normal.

Por ultimo diríamos que la aproximación es bastante buena ya que se utilizo una n bastante grande.