Tarea 8

Dehesa Corona Valeria Carolina. Ochoa Chávez Ana Karen.

4 de octubre de 2018

- Realizar programas para simular las siguientes variables aleatorias: (a) Exponencial(λ)
  (b) BinomialNegativa(r, p)
  (c) Hipergeometrica(N, K, n)
  (d) P areto(m, α)

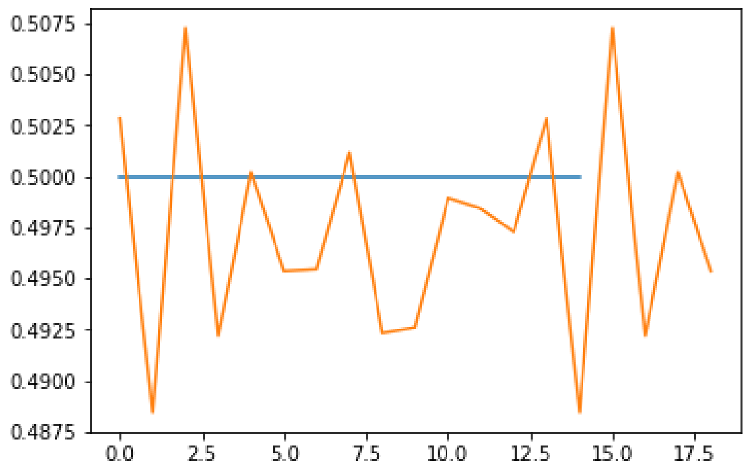
- Para las densidades f = Exp(2), Pareto(2, .05), Pareto(2,1.5), realizar los siguientes programas en cada caso. Imprimir gráficas para cada programa y comentar los resultados para cada caso.

a)Haga un programa que verifique si se cumple o no la Ley Fuerte de los Grandes Números, y grafique los promedios parciales, hasta cierto n.

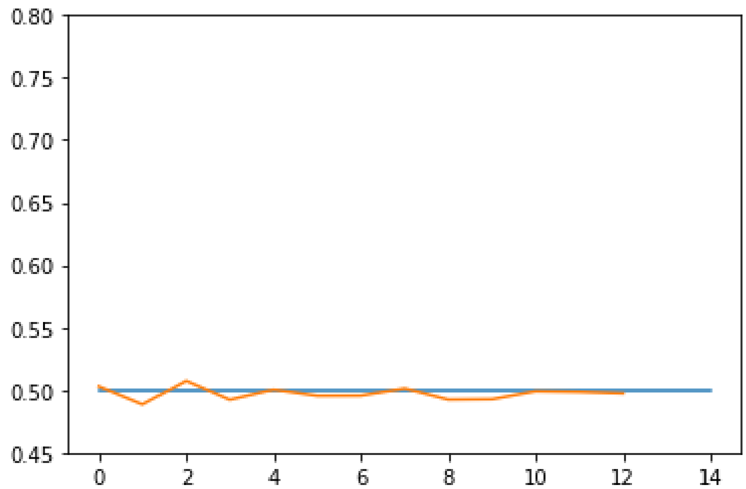
Para verificar esto, realizamos lo siguiente:
La ley Fuerte de los Grandes Números dice, a grandes rasgos, que si tenemos una sucesión X_n de variables aleatorias independientes tales que E(X_n)=μ
$\longrightarrow$Y$_n$=$\frac{X1+X2+X3+....+Xn}{n}$
converge a μ casi seguramente, lo que se traduce en que el promedio de las variables aleatorias converge al valor esperado de la distribución, casi seguramente.
Sabemos que la Esperanza de una exponencial es $1/\lambda$, por lo que, en el caso de la Exponencial(2), la esperanza es 1/2=0.5. Calculamos la suma de las X_i entre n, cuyo valor era fijo, con el programa generador de m variables aleatorias con distribución exponencial, para ver qué valor obteníamos y si este convergía al valor esperado (0.5). Lasmedias que obtuvimos después de realizar varias veces el procedimiento, fueron las siguientes:

0.5028458718562564, 0.4884438486132225, 0.5072611388950747, 0.49218664629496345, 0.5002058281849016, 0.49535758733501234, 0.4954483272882394, 0.5011644355080601, 0.49231963918370075, 0.49260030568353297, 0.4989302226849856, 0.49841231700711514, 0.4972679667935615

Además graficamos los valores anteriores, comparándolos con la media de una exponencial(2), que es 0.5. Obtuvimos el gráfico siguiente:



La línea azul horizontal es la media y la curva naranja son los valores de las medias muestrales. Por el tipo de escala no podríamos decir que realmente convergen a 0.5, pero si vemos la misma gráfica con menos zoom, queda algo de la siguiente manera:



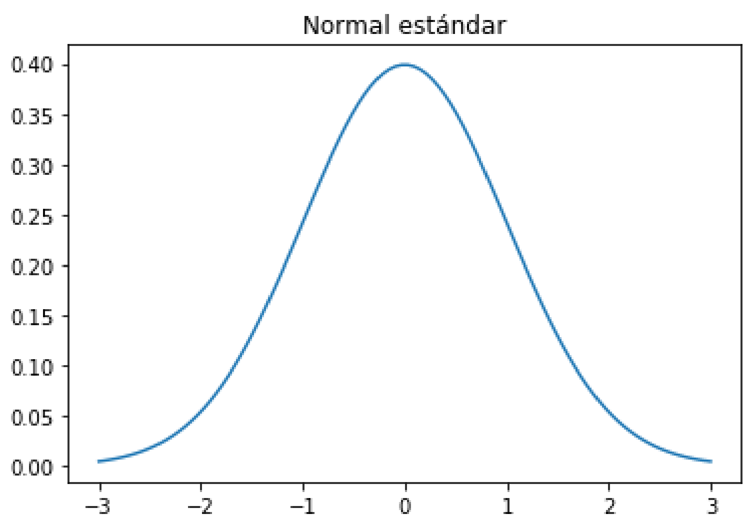
Notamos que los valores están bastante cerca de 0.5, se pasan por milésimas o les falta menos de una centésima para llegar al 0.5 que deseamos. Concluimos así que la distribución exponencial(2) cumple la Ley Fuerte de los Grandes Números.

b) Dado n grande (fijo), hay que realizar un programa que simule m números aleatorios de la forma

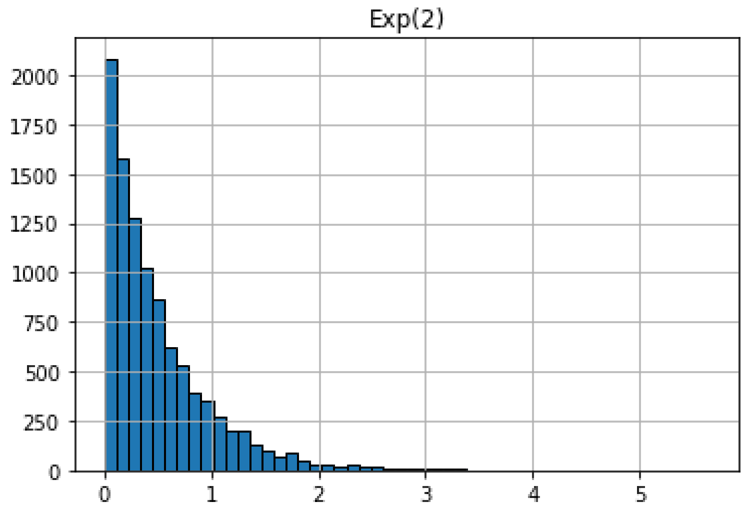
$$\frac{\sum_{i=1}^{n} x_i - n\mu}{\sigma\sqrt{n}} \tag{1}$$

donde cada Xi tiene densidad f. Hacer un histograma de los m números y comparar con la normal estándar.

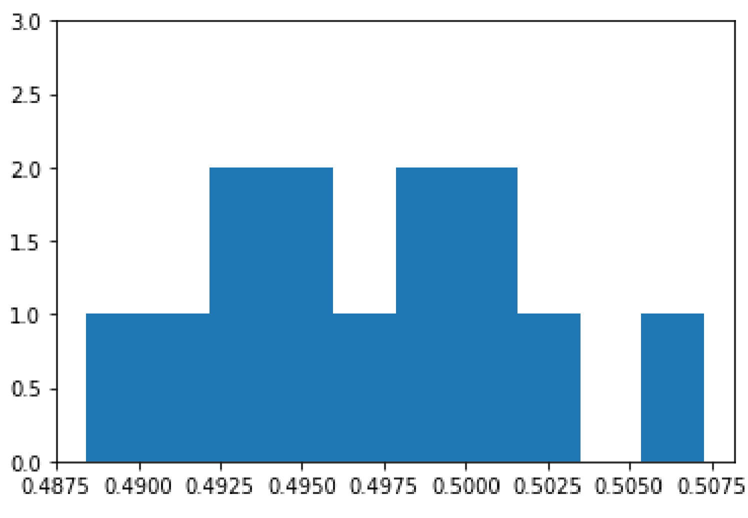
Primero sacamos la gráfica de la curva de la normal estándar, que se ve de la forma siguiente:



Posteriormente, obtuvimos el histograma de la Exp(2) con los m valores pedidos de la forma de la ecuación (1). La imagen del histograma se presenta a continuación.



A pesar de que cada vez que corríamos el programa, obteníamos valores distintos (puesto que se generaban n valores aleatorios distribuidos exponencialmente), el histograma nos salía de la misma forma. Vemos que así como está expresada la función, no parece tener una distribución normal, sin embargo, en la imagen siguiente se muestra el histograma de la distribución de las medias, obtenido con el arreglo de las medias del inciso anterior, y notamos que el gráfico es aproximadamente normal.



Tanto para f=Pareto(2, .05) como para f=Pareto(2, 1.5) es prácticamente imposible realizar la prueba para el Teorema de Límite Central con los conocimientos de Probabilidad 1, ya que para obtener los datos numéricos de la ecuación (1) del presente documento, es necesario utilizar la desviación estándar de la distribución, pero como se demostrará en estadística 1, la varianza no existe para la distribución Pareto cuando $\alpha \leq 2$ y por lo tanto, la desviación estándar tampoco existe. En nuestro caso estamos trabajando con Paretos donde $\alpha = 0{,}05$ y $\alpha = 1{,}5$, respectivamente. Para poder resolver este problema, deberíamos de ocupar la Varianza Muestral, conocimiento que será adquirido hasta el curso de Inferencia Estadística I.

c) El teorema de Glivenko-Cantelli dice lo siguiente. Dada una sucesión $X_1, X_2, ...$ de v.a.i.i.d. definamos la distribución empírica de la muestra $\{X_1, ..., X_n\}$ por:

$$F_n(t) = \frac{1}{n} \sum 1_{[x_i,\infty)}(t), t \in R \tag{2}$$

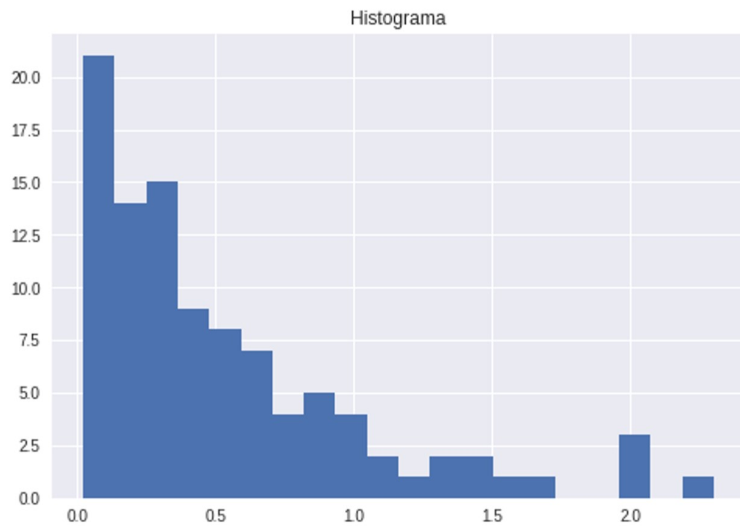
Entonces ocurre que:

$$sup_{t \in R} |F_n(t) - F(t)| \xrightarrow{P} 0 \tag{3}$$

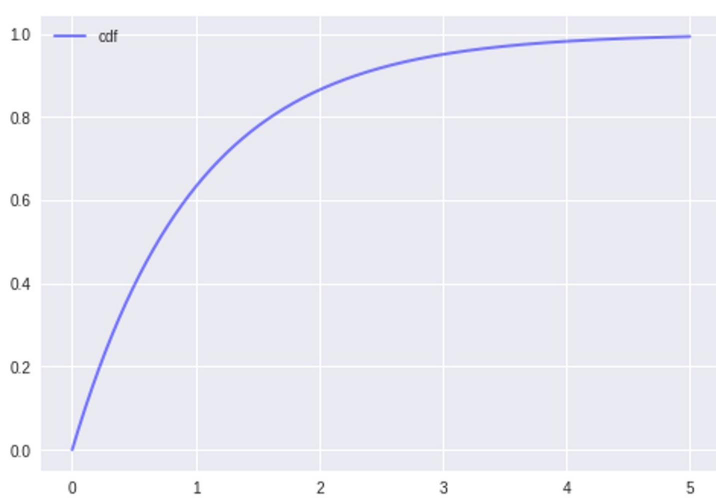
Verificar el Teorema de Glivenko-Cantelli, al simular n números aleatorios para un número n grande. Comparar de dos formas:
Hacer un histograma con los datos y comparar con la densidad f. Construir la distribución empírica y comparar con la distribución F.

Recordemos que la EDF es una variable aleatoria. Podemos apreciar mejor a qué se refiere el teorema observando qué ocurre a la ECDF para n simulaciones desde:

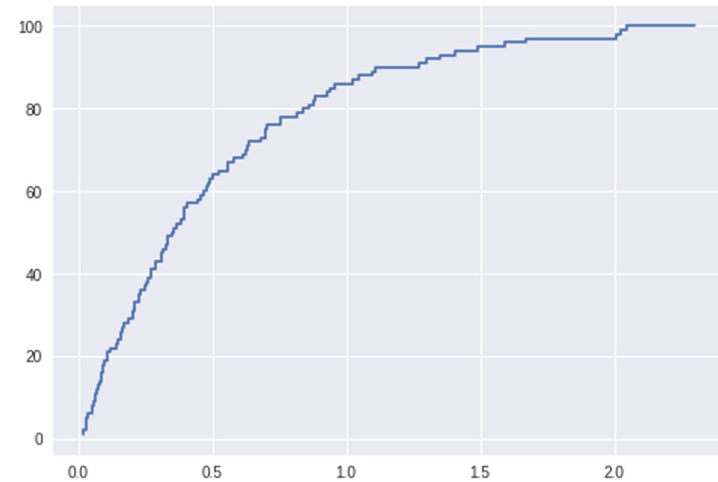
- Exp(2), mientras n aumenta.
- Uniforme(0,1), mientras n aumenta.



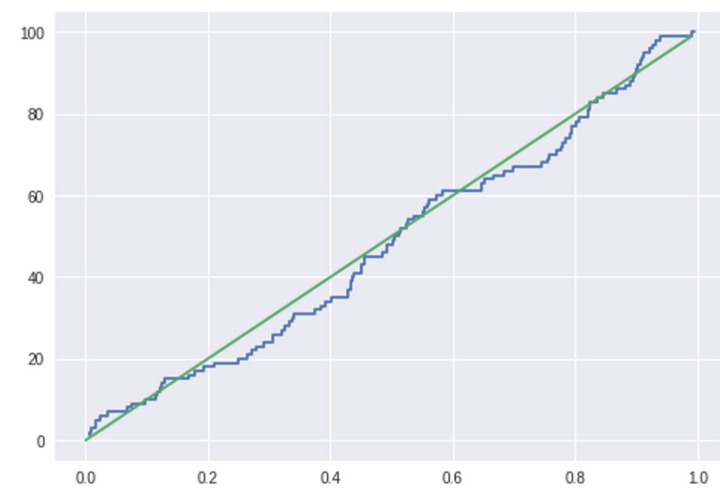
HISTOGRAMA DE LA EXPONENCIAL



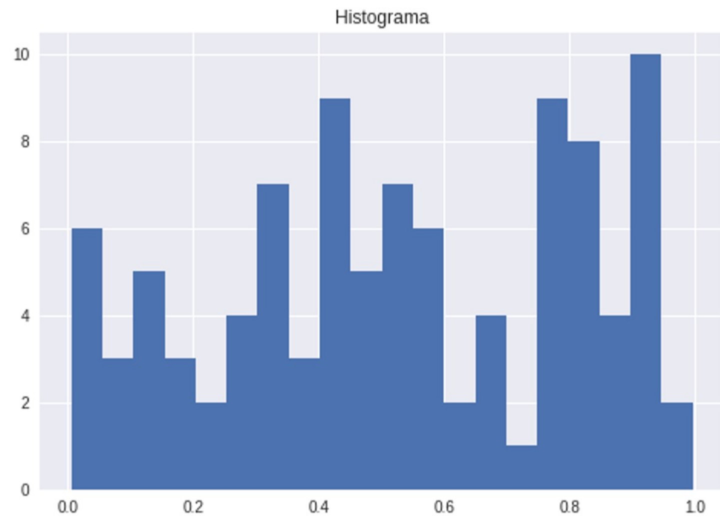
DISTRIBUCIÓN EXPONENCIAL



FUNCIÓN DISTRIBUTIVA EMPÍRICA EXPONENCIAL



FUNCIÓN DISTRIBUTIVA Y DISTRIBUTIVA EMPÍRICA UNIFORME



HISTOGRAMA UNIFORME

Por lo que es observable que conforme n es más grande, F_n se va pareciendo más y más a la DF $F(t)$. Se mantendrá sin importar el valor de $F(t)$