

Tarea 2: Variables aleatorias

Timna Belinda Brown Ramírez

Resumen

Palabras claves

Python — probabilidad — variable aleatoria — función acumulativa — función de densidad

Escuela de Ingeniería Eléctrica, Universidad de Costa Rica

Correspondencia: timna.brown@ucr.ac.cr

Índice

Introducción	1
1 Marco teórico	1
2 Resultados	1
2.1 Solicitado	1
2.2 Parte extra	2
3 Conclusiones	2
4 Recomendaciones	2
5 Código implementado	2
5.1 Paquetes	2
5.2 Definiciones	2
5.3 Lectura de base de datos	3
5.4 Calculo de momentos	3
5.5 Histograma de data	3
5.6 Curva de mejor ajuste	4
5.7 Frecuencia relativa y probabilidad	4
5.8 Transformación de X	4
5.9 Histograma de la transformación	4
5.10 Densidad de la transformación	4
5.11 Código completo	4
6 Gráficas resultantes	4
Referencias	5
7 Anexos	6

Introducción

Esta tarea corresponde a la aplicación de los conceptos obtenidos en la documentación que engloba el tema de variables aleatorias.

1. Marco teórico

Se presenta el análisis de la *variable aleatoria* la cual es una función, se caracteriza por ser determinista. Este análisis es realizado mediante diferentes paquetes los cuales se mostrarán más adelante. Es importante recalcar que toda la

teoría empleada a nivel lógico matemático fue obtenida del curso.

Aspectos importantes a definir:

Python es un lenguaje de programación interpretado cuya ideología recae en la legibilidad de su código [1]. Debido a que es un lenguaje interpretado se considera de alto nivel. Entre sus ventajas principales se encuentran:

1. Es muy útil en scripting
2. Posee muchas herramientas para análisis de datos y tratamiento de señales.
3. Python es un lenguaje portable.
4. Es capaz de utilizar bibliotecaslibres de inteligencia digital, las cuales en su mayoría poseen una mejor velocidad en reacción.

NumPy es definido como una expansión de las funciones de Python las cuales son de código abierto y permiten implementar funciones que manipulan de una manera más fácil para el usuario matrices y vectores permitiendo ampliar las herramientas que se poseen para manejar funciones matemáticas más complejas [2].

2. Resultados

2.1 Solicitado

A nivel de resultados se obtienen los siguientes valores:

- La **media** obtenida desde **pandas** es de: 37.56722880
- La **varianza** obtenida desde **pandas** es de: 376.49325
- El **sesgo** obtenido desde **pandas** es de: 0.63111066
- La **curtosis** obtenida desde **pandas** es de: 0.24508930
- La **frecuencia relativa** perteneciente al intervalo [12, 54] es de: 122060.06310
- La **probabilidad** obtenida mediante las funciones importadas para este rango es de: 0.73173631

- La **media de la transformación**: 5.90474171
- La **desviación de la transformación**: 1.626435351

Con lo que respecta al significado de los momentos tenemos que permiten resumir el comportamiento de las variables de manera general. Considerando la forma del histograma se analiza el modelo de Rayleigh en donde:

1. Varianza medida de la dispersión de la función al rededor del promedio.
2. Desviación estándar es el cuadrado de la varianza.
3. Sesgo (Inclinación) refiere a la tendencia que tiene una descripción completa de la variable aleatoria. Si es cero la pdf es simétrica, negativa tiende a la izquierda y positivo tiende a la derecha.
4. Curtosis (Medida de abultamiento) Se denomina como k_x , si $k_x < 0$ es achatada es decir como si tanta cima. Si $k_x > 0$ es prominente como si fuese una montaña, sería empinada.

Respecto al intervalo para calcular la probabilidad y la densidad se considera que mi carnet es B61254 dadas las instrucciones se toma el intervalo [12,54]. Por otro lado, tenemos que la forma de calcular la transformación se da a raíz de que se poseen datos del archivo csv son X. Se debe realizar $Y = \sqrt{X}$ que es la transformación.

2.2 Parte extra

Se considera la integral del intervalo de [4,8] de la raíz de x. Para esto se calcula me grafica y se muestra en la imagen de **Densidad de la transformada de datos.csv** Tenemos que:

$$\int_4^8 \sqrt{x} dx \quad (1)$$

Por lo que sería el área bajo la curva en este rango figura del **Histograma de la transformada de datos.csv**

3. Conclusiones

El comportamiento respecto a los valores encontrados y esperados se considera apto dentro del posible margen de error debido a que se están trabajando con una base de datos de números flotantes. Además existen diversas herramientas en las cuales se pueden realizar estos cálculos como graficar las diferentes funciones.

4. Recomendaciones

Se recomienda investigar sobre como funcionan las diferentes definiciones existentes en las bibliotecas. Dado que estas permiten ahorrar tiempo y manipular gran cantidad de datos de manera eficiente.

5. Código implementado

5.1 Paquetes

A lo largo del desarrollo de la tarea se utilizaron los siguientes paquetes:

```
from __future__ import division
import csv
from pylab import *
import matplotlib.pyplot as plt
from sklearn import *
from sklearn.preprocessing
import PolynomialFeatures
from numpy import *
import numpy as np
import math
from collections import OrderedDict
import decimal
import scipy.stats as stats
import pandas as pd
import matplotlib.mlab as mlab
from scipy.stats import norm
from scipy.stats import rayleigh
```

5.2 Definiciones

Con el fin de realizar ciertos cálculos se crearon las siguientes definiciones:

```
def media_arit_lista(lista):
    n = len(lista) + 1
    sumatoria = 0
    for i in lista:
        sumatoria = sumatoria
        + float(i)
    return sumatoria/n

def varianza_lista(lista):
    n = len(lista) + 1
    sum_var = 0
    for j in lista:
        sum_var = sum_var
        + (media_arit_lista(lista)
        - float(j))**2
    return sum_var/n
```

```
def desvia_estan_lista(lista):
    desvi = varianza_
    lista(lista)**(1/2)
    return desvi

def intervalo(data):
    inter = []
    for k in data:
        redondeado =
        round(float(k))
        # Rango de 12 a 54
        if (12<=redondeado)
        and
        (redondeado<=54):
            inter.append(
                float(k))
    return inter
```

5.3 Lectura de base de datos

```
with open("datos.csv", "r") as csv_file:
    # Leyendo cada celda y separandola
    # con coma para poder
    # interpretar los datos
    csv_reader = csv.reader
    (csv_file, delimiter=',')
    # Se salta la primera
    linea del CSV
    next(csv_reader)
    # Arreglo para
    #guardar los datos aleatorios
    data = []
    # Recorre todas las
    #filas dentro del archivo cvs
    for filas_completas_data
    in csv_reader:
        # Se guardan
        #los datos en
        # un arreglo
        data.append(
            float(
                filas_completas_data[0]))
```

5.4 Calculo de momentos

```
mu, sigma = stats.rayleigh.fit(data)
desv = stats.rayleigh(mu, sigma)
media_pandas, var_pandas,
inclinacion_pandas,
curtosis_pandas =
desv.stats(moments='mvsk')
print("La media
obtenida desde pandas
es de: ", media_pandas)
print("La varianza
obtenida desde pandas
es de: ", var_pandas)
print("El sesgo obtenido
desde pandas
es de: ", inclinacion_pandas)
print("La curtosis obtenida
desde pandas es de: ", curtosis_pandas)
media = media_arit_lista(data)
varianza = varianza_lista(data)
desviacion_std = desvia_estan_lista(data)
```

5.5 Histograma de data

```
histograma = plt.hist(data,30,
density=True)
plt.savefig('Histograma de data ')
# # Limpia el area de graficacion
plt.cla()
```

5.6 Curva de mejor ajuste

```
N = len(data) + 1
escala = media_pandas/np.sqrt(np.pi/2)
V_norm_hist = escala *
np.sqrt(-2* np.log
(np.random.uniform(0, 1, N)))
fig, ax = plt.subplots(1, 1)
num_bins = 30
_binvalues, bins, _patches =
plt.hist(V_norm_hist,
bins=num_bins,
density=False, rwidth=1,
ec='blue', label='Histograma')
x = np.linspace(bins[0], bins[-1], 120)
binwidth = (bins[-1] -
bins[0]) / num_bins
escala = V_norm_hist.mean() /
np.sqrt(np.pi/2)
plt.plot(x, rayleigh(loc=0,
scale=escala).pdf(x) *
len(V_norm_hist)*binwidth,
lw=5, alpha=0.6,
label='Rayleigh pdf')
plt.legend()
plt.savefig('Curva de mejor ajuste')
# # Limpia el area de graficacion
plt.cla()
```

5.7 Frecuencia relativa y probabilidad

```
numeros1254 = 0
intervalo1254 = intervalo(data)
# print(intervalo1254)
for i in range(N):
    # print(i)
    for j in intervalo1254:
        numeros1254 =
            numeros1254 + j
frecuencia_relativa1254 = numeros1254/N
print("\n \n La frecuencia relativa
perteneiente al intervalo
[12, 54] es de: ",
frecuencia_relativa1254)
# # # Calculando la probabilidad
probabilidad =
desv.cdf(54) - desv.cdf(12)
print( 'La probabilidad
obtenida mediante las
funciones importadas
para este rango es de: ',
probabilidad)
```

5.8 Transformación de X

```
sqrt_x = 0
Y_trasnfor= []
for values in data:
    sqrt_x = float(values)**(1/2)
    Y_trasnfor.append(sqrt_x)
```

5.9 Histograma de la transformación

```
histograma = plt.hist(Y_trasnfor,30)
plt.savefig('Histograma transformacion')
# # Limpia el area de graficacion
plt.cla()
```

5.10 Densidad de la transformación

```
media_y = media_arit_lista(Y_trasnfor)
print("La media de la
transformacion:", media_y)
desviacion_y = desvia_estan_lista(Y_trasnfor)
print("La desviación de
la transformacion", desviacion_y)
distribucion_y = norm(loc=media_y,
scale=desviacion_y)
x_transf = np.linspace(distribucion_y.ppf(0.001),
distribucion_y.ppf(0.999), 120)
plt.hist(Y_trasnfor,density=True,
label='Transformacion')
plt.plot(x_transf, distribucion_y.pdf(x_transf),
'r-', label='PDF teorico')
plt.legend()
plt.savefig('Funcion de densidad
de la transformacion')
```

5.11 Código completo

Código con más anotaciones en: https://github.com/brown9804/Modelos_Probabilisticos/tree/master/Variables_Aleatorias

6. Gráficas resultantes

Mediante la herramienta de python se obtienen las siguientes gráficas:

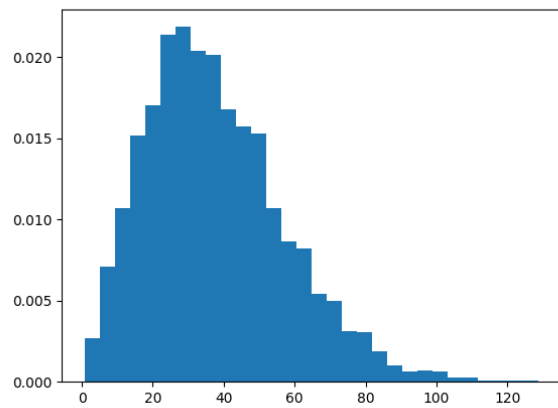


Figura 1. Histograma de la data de datos.csv

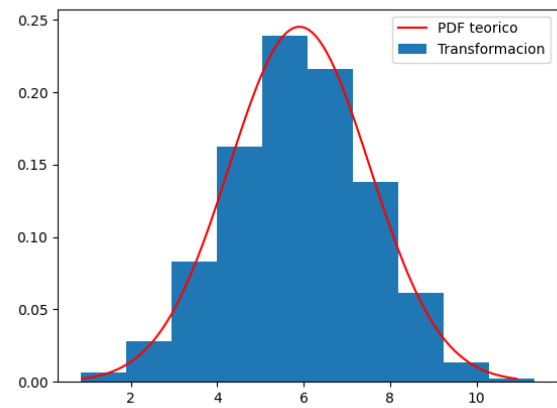


Figura 4. Densidad de la transformada de datos.csv

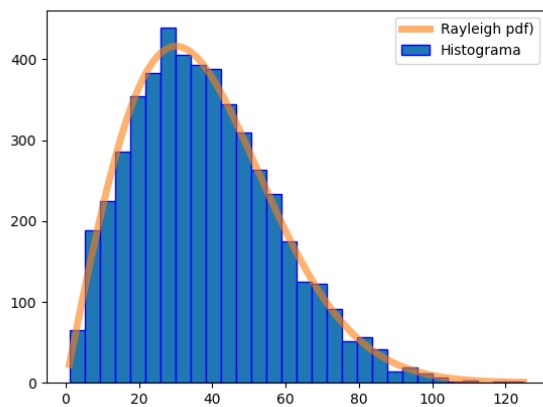


Figura 2. Curva de mejor ajuste de la data de datos.csv

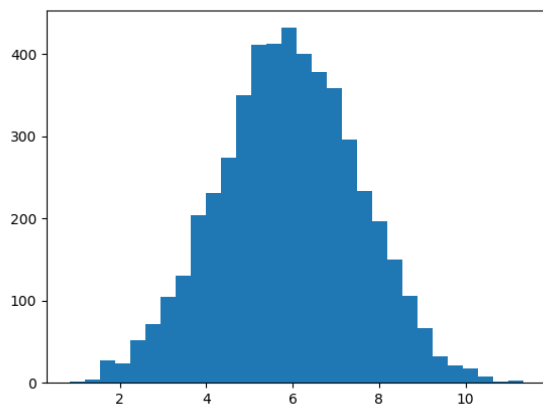


Figura 3. Histograma de la transformada de datos.csv

Referencias

- [1] J Knowlton. Python. Anaya Multimedia-Anaya Interactiva, 2009.
- [2] NumPy project and community. Numpy, Obtenido el 3 de junio del 2020.

7. Anexos

Situación En el archivo datos.csv está el registro de mediciones de una variable física.

Asignaciones (5 %) Crear un usuario de GitHub. Para comprender más del uso de Git y GitHub, puede revisar el tutorial Git0 - Introducción a Git, disponible en Mediación Virtual o en <https://github.com/fabianabarca/git0/blob/master/Git0.md>. Debe indicar en el reporte cuál es su nombre de usuario de GitHub.

(10 %) Hacer un clon del repositorio disponible en github.com/fabianabarca/mpss donde están los tutoriales PyX para Python ejecutables en Jupyter (trabajo en progreso). Ver Ejemplo 2 de Git0. Para verificar este punto, debe dar “Star” a este repositorio desde su usuario de GitHub.

(10 %) Hacer “fork” del repositorio fabianabarca/git0. Ahí, seguir las instrucciones de modificación del archivo himnovaciones.txt. Debe hacerse al menos un “pull request” para la modificación colectiva del archivo.

(15 %) Graficar un histograma.

(15 %) Con los datos contenidos en datos.csv, encontrar la mejor curva de ajuste y graficar la curva de ajuste encontrada.

(10 %) Encontrar la probabilidad en el intervalo $[a, b]$ en el modelo encontrado y contrastarlo con la frecuencia relativa de los elementos de datos.csv que están en realidad en ese mismo intervalo. En el carné A12345, $a = \min(23, 45)$ y $b = \max(23, 45)$. Para este ejemplo, el intervalo es $[23, 45]$. Si fueran iguales, reemplazar por $(12, 45)$.

(10 %) Calcular los cuatro primeros momentos y comentar sobre el “significado” de cada uno y la correspondencia con la gráfica observada y los valores teóricos.

(10 %) Si los valores de datos.csv son X y pasa por la transformación $Y = \sqrt{X}$, graficar el histograma de Y . Oferta de inicio de época lluviosa: puntos extra (generosos) si encuentran la expresión para función de densidad de Y , por medio de la deducción analítica. Pero llame ya.

(15 %) Crear un reporte que contenga los resultados anteriores utilizando la plantilla: <https://www.overleaf.com/read/chzmkkhycgvy>

Resultados Enviar el archivo [carné].pdf, incluyendo en el texto los fragmentos de código relevantes para verificar su trabajo (no hay que enviar ningún archivo de código fuente). Debe ser un documento bien comentado, claro y conciso.