Tarea 2: Variables aleatorias múltiples

Katzrin Myrie Berger

Resumen
En el siguiente informe se describe tanto el comportamiento un archivo con dados de variables aleatorias, como también la curva que describe de mejor manera estos datos. Por otra parte se desarrolla un poco el concepto de covarianza, correlación y la constante de pearson.
Mediante la realización de distintas pruebas experimentales fue posible obtener los datos necesarios para la aplicación de las fórmulas que dicta la teoría, y de este modo comparar los resultados experimentales con la probabilidad de los dantos brindados en la asignatura.

Palabras claves
Python — probabilidad — variables aleatorias multiples — función de densidad marginal — función de densidad acumulativa

Escuela de Ingeniería Eléctrica, Universidad de Costa Rica
Correspondencia: Katzrin.Myrie@ucr.ac.cr

Índice

1. Introducción

Las variables aleatorias múltiples a diferencia de las variablea aleatorias simples pasan de un espacio Ordinario, "simple.º tambien conocido como marginal a un espacio çonjunto, este último hace referencia a la probabilidad conjunta de los elementos en el espacio marginal.

Para la obtener todos estos valores y determinantes acerca de las variables aleatorias múltiples es necesariala utilización de un lenguaje de progamación de alto nivel que permita determinar los parámetros probabilísticos necesarios.

La probabilidad representa una forma de determinar o decifrar , en cierta manera, el mundo que nos rodea y las tendecias o movimientos que este puede presentar, para moldear este según nuestro entendimiento.

2. Marco teórico

Se procederá a definir conceptos teóricos de gran relevancia para la realización de la asignatura, para una mayor comprención de los métodos y teoría utilizada en su resolución.

2.1 Variable aleatoria múltiple

Esta se divide en dos según sus carácterísticas:

- Variable aleatoria continua: Es aquella cuyo campo de variación es un intervalo de la recta real, una unión de varia intervalos o la totalida de la recta real, por lo tanto los valores definidos por esta variable son un conjunto infinito no numerable[1]

- Variable aleatoria discreta: una variable es discreta cuando su campo de variación, (dominio de definición), está definido por un conjunto finito o infinito, numerable de valores posibles: es decir, a cada suceso le corresponde un valor[2].

2.2 Funciones de densidad marginal

Las funciones de densidad marginal corresponden a la sumatoria de los datos en el espacio ordinario o marginal de cada variable. si las variables aleatorias son continuas su densidad marginal se representa de la siguiente forma[1]:

$$f(y) = \int_{-\infty}^{\infty} f_{x,y}(x,y)d_x = \frac{d}{d_y}F_y(y) \tag{1}$$

$$f(x) = \int_{-\infty}^{\infty} f_{x,y}(x,y)d_y = \frac{d}{d_x}F_x(x) \tag{2}$$

Si las variables aleatorias son discretas corresponden a la suma matricial en el espacio vectorial correspondiente.[3]

2.3 Función de densidad conjunta

La función de densidad conjunta representa la unión de las funciones de densidad marginal en el espacio correspondiente del mismo[4].

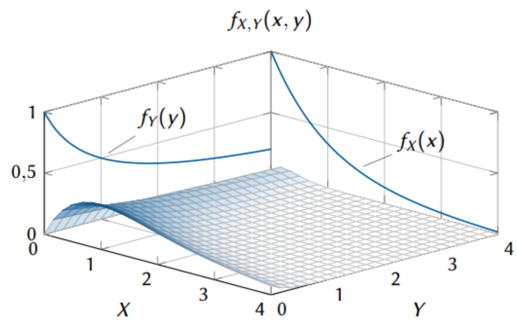

Figura 1. Densidad conjunta

Si y solo sí las variables son independientes la multiplicación de la las densidades marginales tendrán como resultado la función de densidad conjunta, es esque caso la función original[2].

$$f_{(x,y)}(x,y) = f_x(x)F_y(y) \tag{3}$$

Además de las funciones generales se tienen instrumentos que nos ayudan a determinar características de la función:

- Media: Es la presencia de una tendencia determinista, el nivel seleccionado como referencia para el comportamiento de la variable[5].
- Varianza: La varianza es el promedio ponderado de la desviación cuadr·tica entre un punto extraÌdo al azar del soporte de la distribuciÛn y la esperanza matem·tica. La ponderaciÛn que recibe cada punto del soporte de la distribución es igual a la masa de probabilidad en cada punto[5].
- Correlación : Un coeficiente de correlación, mide el grado de relación o asociación existente generalmente entre dos variables aleatorias. No es conveniente identificar correlación con dependencia causal, ya que, si hay una semejanza formal entre ambos conceptos, no puede deducirse de esto que sean análogos.[3]
- Inclinación: la cual hace referencia al promedio de inclinación que tiene la función y su crecimiento[4].
- Covarianza: La covariancia indica en promedio, que tanto se alejan conjuntamente los valores de sus medias respectivas. Se define como[5]:

$$C_{(x,y)} = E[(XX)(YY)] \tag{4}$$

- Constante de pearson : Tiene como objetivo medir la fuerza o grado de asociación entre dos variables aleatorias cuantitativas que poseen una distribución normal bivariada conjunta[3].

3. Resultados

Mediante el uso del programa Jupyter se lograron obtener los siguientes resultados para los puntos respectivos:

- Pregunta 1: En primera instancia se encontró las funciones de densidad marginal. debido a que son variables discretas se realizó la suma los valores en el espacio vectorial correspondiente, seguidamente se realizó la impresión de las densidades marginales, para visualizar la forma que tienen estos datos y así encontar la curva e mejor ajuste.

```
import csv
import seaborn as sns
import scipy as sp
from scipy import stats
from scipy.stats import norm
from scipy.stats import rayleigh


from scipy.optimize import curve_fit
from scipy.stats import kurtosis
from scipy.stats import skew
from matplotlib import pyplot as plt
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import axes3d
from matplotlib import style
import pandas as pd
import numpy as np

#Se definen las variables para leer los datos csv

L = pd.read_csv('xy (1).csv')
P= pd.read_csv('xyp.csv')

#Vectores de posición X y Y

X= np.linspace(5,15,10)
Y= np.linspace(5,25,21)

#Funciones de densidad marginal

fX=np.sum(L, axis=1)
fY=np.sum(L, axis=0)

#Gráficas
fun1= (fX,X)
fun2=(fY, Y)

plt.plot (X, fX)
plt.show()

plt.plot (Y, fY)
plt.show()
```

De lo cual se obtuvieron las siguientes gráficas:

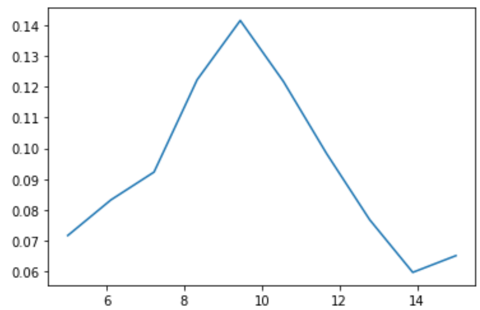

Figura 2. Gráfica de la función marginal de x

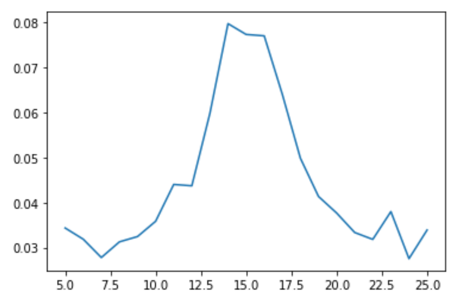

Figura 3. Gráfica de la función marginal de Y

Luego de esto se puede observar que las gráficas se asemejan a la curva normal (de campana) o curva gaussiana por lo que se aplicó el siguiente código se obtuvieron los parámetros de media y desviación:

```
def gaussiana(x, mu, sigma):
    return 1/(np.sqrt (2*np.pi*sigma**2))*
    np.exp(-(x-mu)**2/(2*sigma**2))

print("Los parámetros de  para f(x):")
param,_=(curve_fit(gaussiana,X,fX))
print (param)

print("____________________")

print("Los parámetros de  para f(Y):")

param2,_=(curve_fit(gaussiana,Y,fY))
print (param2)
```

De lo que se obtuvieron los siguientes parámetros de media y desviación respectivamente:

- Los parámetros de ajuste para f(x):
  [9.53068453 3.25802246]
- Los parámetros de ajuste para f(Y):
  [15.1527004 6.32588545]

- Pregunta 3: Hallar los valores de correlación, covarianza y coeficiente de correlación (Pearson) para los datos y explicar su significado. con los valores esncontrados anteriormente se buscan los siguienten valores:

```
##correlación
P['correl'] = (P['x']*P['y'])*P['p']
#Hacer la suma
correlacion=np.sum(P['correl'] , axis=0)
print("La correlación es : ")

print(correlacion)
print("_______________________________________________")

### covarianza

YM=np.mean(Y)
XM=np.mean(X)


P['cova'] =(P['x']-XM )*(P['y']-YM)*P['p']
#Hacer la suma
covarianza=np.sum(P['cova'] , axis=0)
print("La covarianza  es : ")
print(covarianza)

print("______________________________________________")


####coeficiente de pearson


P['pearson'] =(P['cova']/(3.25802246*6.32588545))
coefp=np.sum(P['pearson'] , axis=0)
print("La constante de pearson :")
print(coefp)
print("__________________________________________")
```

De lo cual se obtuvieron los siguientes valores :

La correlación es : 149.54281

La covarianza es : 0.06481000000000009

La constante de pearson : 0.0031446087908018183

- Pregunta 4: Graficar las funciones de densidad marginales (2D), la función de densidad conjunta (3D).

```
from matplotlib.pyplot import *
from pylab import *
from mpl_toolkits.mplot3d.axes3d import A3D
from matplotlib import style


x1 = X
plt.plot (X, fX)
G1=1/(np.sqrt (2*np.pi*3.25802246**2))**
np.exp(-(x1-9.53068453)**2/(2*3.25802246**2))
plt.plot ( x1, G1, color='#AA1111')
plt.show()

y1=Y
plt.plot (Y, fY)
G2=1/(np.sqrt (2*np.pi* 6.32588545**2))**
np.exp(-(y1-15.1527004)**2/(2* 6.32588545**2))
plt.plot ( y1, G2, color='#AA1111')

plt.show()


# Creamos la figura
fig = plt.figure(figsize=(8,7))
# Datos en array bi-dimensional
x = X
y = Y
x,y=np.meshgrid(x,y)
z = 1/(np.sqrt (2*np.pi*3.25802246**2))**
np.exp(-(x-9.53068453)**2/(2*3.25802246**2))*
1/(np.sqrt (2*np.pi* 6.32588545**2))*np.exp(-
*2/(2* 6.32588545**2))
# Agrrgamos un plano 3D
ax1 = fig.add_subplot(1,1,1,projection='3d')
# plot_wireframe nos permite
#agregar los datos x, y, z.
#Por ello 3D es necesario que los
#datos esten contenidos
#en un array bi-dimensional
ax1.plot_wireframe(x, y, z)

# Mostramos el gráfico
show()
```

De lo cual se obtienen las siguientes gráficas:

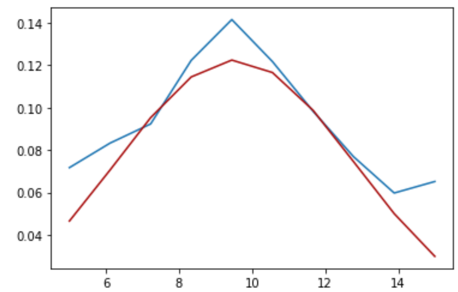

Figura 4. Función de densidad f(x) vs curva gaussiana

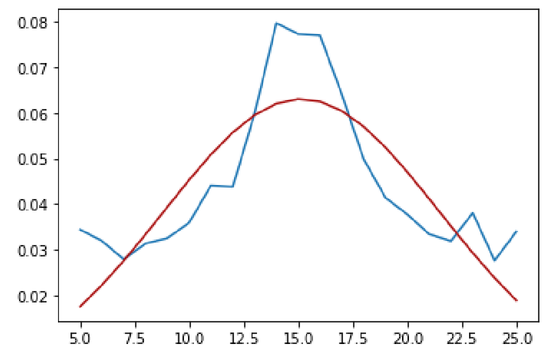

Figura 5. Función de densidad f(y) vs curva gaussiana

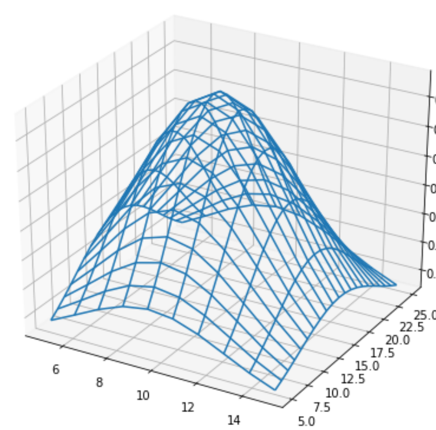

Figura 6. Función de densidad conjunta 3D

Pregunta 2: Asumir independencia de X y Y. Analíticamente, ¿cuál es entonces la expresión de la función de densidad conjunta que modela los datos? Se procedío a realizar una búsqueda bibliográfica y se notó que al ser vaiables discretas las función de densidad conjunta se define como la multiplicación de que sucedan los intervalos uno primero que el otro por lo tanto :

$$f(x,y) = f(x) * f(y) \tag{5}$$

Se hace que la comprobación teórico con los parámetros encontrados:

- $f_{(x)}(x) = \frac{e^{\frac{(x-u_x)^2}{2*\sigma_x^2}}}{\sigma\sqrt{2\pi}}$

- $f_{(y)}(y) = \frac{e^{\frac{(y-u_y)^2}{2*\sigma_y^2}}}{\sigma\sqrt{2\pi}}$

Por lo que al final queda la siguiente expresión con los datos obtenidos con anterioridad:

- $f_{(x,y)}(x,y) = \frac{e^{\frac{(x-9{,}53)^2*3{,}25^2+(y-15{,}1527004)^2*6{,}32^2}{2*3{,}25^2*6{,}32^2}}}{6{,}32*3{,}25*2*\pi}$

4. Conclusiones

De los resultados obtenidos se observan las siguientes conclusiones:

- se logró encontara los momentos de la función para así encontarar más información de la misma .

- Se logró realizar un curve fit y obtener los parámetros de la curva de mejor ajuste, además graficar en 2D las curvas de densidad marginal y en 3D la curva de densidad conjunta.

- Se profundizó en las funciones de densidad de las variables aleatorias multiples, las transformaciones que pueden tener las variables aleatorias, además de como se puede obtener los resultados.

Referencias

[1] M. Alperin, "Introducción al análisis estadístico de datos geológicos," Master's thesis, Universidad Nacional de La Plata, 2017.

[2] F. A. Cálderón, "Funciones que dan momentos," Master's thesis, Universidad de Costa Rica, 2020.

[3] L. F. Restrepo B and J. A. GonzÃ¡lez L, "De Pearson a Spearman," *Revista Colombiana de Ciencias Pecuarias*, vol. 20, pp. 183 – 192, 06 2007.

[4] F. A. Cálderón, "transformaciones de una variable aleatoria," Master's thesis, Universidad de Costa Rica, 2020.

[5] F. A. Cálderón, "Momentos de las variables aleatorias," Master's thesis, Universidad de Costa Rica, 2020.