2. Modelos de distribución multivariados

2.1. Funciones de distribución y densidad

Autor: Emmanuel Alcalá Google Scholar

Variables aleatorias, espacios muestrales y realizaciones

Suponer que X representa una cantidad desconocida de interés, como la temperatura de una máquina. Si los valores de X son desconocidos o cambian, decimos que X es una variable aleatoria, o **VA**. El conjunto de todos los valores *posibles* se denota como $\mathcal{X}$, llamado espacio muestral o espacio de estados. Un evento es un *sub*conjunto de valores del espacio muestral. Por ejemplo, si $\mathcal{X}$ es el lado de un dado al ser lanzado, $\mathcal{X} = \{1, 2, 3, 4, 5, 6\}$, el evento ``cae 1'' se denota $X = 1$, y ``ver un número impar'' se denota $X \in \{1, 3, 4\}$, el evento ``cae entre 4 y 6'' como $4 \leq X \leq 6$.

Vamos a enlazar todos los conceptos:

Variable Aleatoria X

Pensemos en X como la temperatura de un horno. Aquí, X no es un valor de temperatura específico. En cambio, es una ``regla'' que nos dice cómo interpretar las mediciones de temperatura que podríamos obtener del horno. Técnicamente, una variable aleatoria es una función que asigna un número real a cada resultado en el espacio muestral. En este caso específico, la función es una función de identidad, lo que significa que X asigna el mismo valor a cada resultado en el espacio muestral. Lo anterior podríamos describirlo como $X : \mathcal{X} \to \mathbb{R}$, o como $X(\omega) = \omega$ para todo $\omega \in \mathcal{X}$, en donde ω es un resultado en el espacio muestral.

Una variable aleatoria más compleja puede ser el índice de masa corporal. El Índice de Masa Corporal (BMI, por sus siglas en inglés) es una medida comúnmente utilizada para evaluar la proporción de peso a estatura de un individuo. El BMI se calcula mediante la fórmula:

$$\text{BMI} = \frac{\text{peso en kg}}{(\text{altura en m})^2}$$

En el contexto de la teoría de la probabilidad, podemos considerar el BMI como una Variable Aleatoria X que es una función del peso y la altura de un individuo.

La Variable Aleatoria X asignaría a cada par (w, h) en el espacio muestral un valor del BMI calculado mediante la fórmula mencionada:

$$X(w, h) = \frac{w}{h^2}$$

Espacio Muestral $\mathcal{X}$

Para el ejemplo del horno, el espacio muestral $\mathcal{X}$ podría ser todos los valores que la temperatura del horno podría tomar en una escala específica, digamos, de 100 a 250 grados Celsius: $X = 100, 101, 102, ..., 250$.

Para el ejemplo del BMI, el espacio muestral $\mathcal{X}$ podría ser el conjunto de todas las posibles combinaciones de peso y altura para un grupo de individuos. Por ejemplo:

$$\mathcal{X} = \{(w, h) | 40 \leq w \leq 200, 1.2 \leq h \leq 2.5\}$$

Realización x

Si tomas una lectura real del termómetro en un momento específico y muestra, digamos, 180 grados Celsius, esa medida es una ``realización'' x de la Variable Aleatoria X.

Para el BMI, una realización x de X sería el BMI específico calculado para una combinación particular de peso y altura. Por ejemplo, si alguien tiene un peso de 70 kg y una altura de 1.75 m, la realización sería $x = X(70, 1.75) = 22.86$.

Eventos

Un ``evento'' es un subconjunto específico de posibles temperaturas que podríamos encontrar interesantes. Por ejemplo:

1. El evento ``El horno está por debajo de 150 grados'' se denotaría como $X < 150$.
2. El evento ``El horno está a una temperatura ideal para hornear pan, entre 175 y 200 grados'' podría denotarse como $175 \leq X \leq 200$.
3. El evento ``tener sobrepeso'', que se podría definir como $X(w, h) \geq 25$.

Nota: generalmente, no se hace distinción entre una variable aleatoria y su realización. Por ejemplo, X y x se usan indistintamente para denotar la realización de una variable aleatoria. Así, aunque técnicamente deberíamos denotar la probabilidad de que el horno esté por debajo de 150 grados como $P(X < 150)$, en la práctica, a menudo escribimos $P(x < 150)$.

Definición de probabilidad

¿Qué es una probabilidad?

Definición: La probabilidad es una medida que describe la posibilidad de que ocurra un evento en relación con la totalidad de posibles eventos. Se expresa numéricamente entre 0 y 1, donde 0 indica imposibilidad y 1 indica certeza.

Matemáticamente: Sea $ S $ un espacio muestral, es decir, el conjunto de todos los posibles resultados de un experimento, y $ A $ un evento específico que pertenece a $ S $. La probabilidad $ P(A) $ de que ocurra $ A $ se define como:

$$P(A) = \frac{\text{número de casos favorables para } A}{\text{número total de casos en } S}$$

Algunas propiedades de las $P()$ son:

1. Si $\emptyset$ es un conjunto nulo, entonces $\Pr(\emptyset) = 0$.
2. Si $A \subset B$, entonces $\Pr(A) \leq \Pr(B)$.
3. Si A^c denota el complemento de A, entonces $\Pr(A^c) = 1 - \Pr(A)$.
4. Si $A \cap B = \emptyset$ denota la intersección nula de A y B, entonces $\Pr(A \cup B) = \Pr(A) + \Pr(B)$, Pr es aditiva para eventos disjuntos.
5. De otra manera, para eventos arbitrarios A, B, $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$.
6. Dos eventos A y B son independientes si $\Pr(A \cup B) = \Pr(A)\Pr(B)$. Por ejemplo, lanzar una moneda dos veces. Cada lanzamiento es independiente del otro. Si queremos saber la probabilidad de que en ambos lanzamientos salga caras, esa probabilidad sería $\Pr(H_1 \cap H_2) = (1/2)(1/2)$.

Probabilidades empíricas

Podemos ilustrar varios conceptos de probabilidad utilizando un arreglo en dos dimensiones. Por ejemplo, en la siguiente imagen (adaptado de MML-6.2)

```python
from IPython.display import display, HTML
image_path =  "img/joint_dist_tikz.png"
display(HTML('<div style="text-align: center;"><img src="{}" width="30%"/></div>'.format(i
```

```
<IPython.core.display.HTML object>
```

muestra dos variables aleatorias bivariadas, X (que puede tomar valores $x_j, j = 1, 2, \ldots, M$) y Y (que puede tomar valores $i = 1, 2, \ldots, L$).

- c_j: conteo marginal en x_j

- r_i: conteo marginal en y_i
- n_{ij}: casos en la celda x_j, y_i

La probabilidad (*conjunta*) de que $X = x, Y = y$ se define como

$$p(x, y) \triangleq \Pr(X = x, Y = y)$$

$p(x, y)$ se lee ``la probabilidad de x y y'', en ocasiones escrito usando notación de teoría de conjuntos como $p(x \cap y)$. Cada entrada (cada celda) en la tabla se calcula contando la cantidad de casos que (x_j, y_i) suceden conjuntamente

$$\Pr(X = x_j, Y = y_i) = \frac{n_{ij}}{N} \tag{1}$$

En la tabla, si contamos todos los casos de la fila i tenemos la cantidad total de veces que y_i sucede, y lo llamamos r_i. Si contamos la columna j tenemos la cantidad total de veces que x_j sucede, y que llamamos c_j.

La *probabilidad marginal* de x, es decir, la probabilidad que $X = x$ sin importar el valor de Y se escribe como $p(x)$ y se calcula como sigue

$$p(X = x_j) = p(x_j) = \frac{c_j}{N} \tag{2}$$

El número de casos en la columna j es la suma de casos en cada celda *en esa columna*, por lo que $c_j = \sum_i n_{ij}$ (dejando fija i). Por lo tanto,

$$p(x_j) = \sum_{i=1}^{L} p(x_j, y_i) \tag{3}$$

A la ecuación (3) se le conoce como la *regla de la suma*. Notar que se fija la columna j y se itera sobre las filas $i = 1, 2, 3$, por lo que por cada columna j tenemos una probabilidad $p(x_j)$, que llamamos *la distribución marginal* de x.

Si consideramos los casos en los que $X = x_j$, la *fracción* de casos en los que $Y = y_i$ se escribe como $p(Y = y_i | X = x_j)$, probabilidad que es conocida como *probabilidad condicional* de $Y = y_i$ *dado* que $X = x_j$. Se obtiene encontrando la fracción de puntos en una columna j que caen en una celda (i, j) dada

$$p(y_i | x_j) = \frac{n_{ij}}{c_j} \tag{4}$$

Notar que c_j es la cantidad total de valores para los cuales $X = x_j$, y n_{ij} es una celda particular. Con las ecuaciones (1), (2) y (4) podemos derivar la siguiente relación:

$$
\begin{aligned}
p(x_j, y_i) &= \frac{n_{ij}}{N} = \frac{n_{ij}}{c_j}\frac{c_j}{N} \\
&= p(y_i|x_j)p(x_j)
\end{aligned}
\tag{5}
$$

La ecuación (5) es conocida como *regla del producto*.

Eliminando la notación excesiva, podemos sintetizar las dos reglas como sigue

$$
\begin{aligned}
\textbf{regla de la suma} \qquad & p(x) = \sum_y p(x, y) \\
\textbf{regla del producto} \qquad & p(x, y) = p(y|x)p(x)
\end{aligned}
$$

Considerando la simetría $p(x, y) = p(y, x)$ y que $p(y, x) = p(x|y)p(y)$, encontramos la siguiente relación

$$
\begin{aligned}
p(y|x)p(x) &= p(x|y)p(y) \\
p(y|x) &= \frac{p(x|y)p(y)}{p(x)}
\end{aligned}
\tag{6}
$$

Con

$$
\begin{aligned}
&1. f(x, y) \geq 0 \\
&2. \sum_{x,y} f(x, y) = 1
\end{aligned}
$$

La ecuación (6) es llamada *Teorema de Bayes*, y tiene una importancia en ML y estadística que no puede ser exagerada.

```
import pandas as pd
import numpy as np
from IPython.display import display, HTML
from tabulate import tabulate
import matplotlib.pyplot as plt
from scipy.stats import *

# Crear distribución conjunta {X,Y}
```

```python
data = {'Y': np.tile(np.arange(0, 4), 3),
        'X': np.repeat(np.arange(0, 3), 4)}

data_for_sampling = pd.DataFrame(data)
data_for_sampling['wgts'] = [0, 3, 9, 3, 2, 18, 18, 2, 3, 9, 3, 0]

# Tomar una muestra aleatoria de la distribución conjunta X,Y, usando wgts como pesos
np.random.seed(452)
sampled_data = data_for_sampling.sample(n=10000, replace=True, weights='wgts')

# Crear una tabla de frecuencia
freq_table = pd.crosstab(sampled_data['Y'], sampled_data['X'], rownames=['Y/X'])

# Mostrar la tabla en formato HTML
display(HTML(tabulate(freq_table, headers='keys', tablefmt='html')))
# X en las columnas, Y en las filas
```

Y/X	0	1	2
0	0	275	416
1	446	2570	1338
2	1251	2523	396
3	462	323	0

```python
freq_table.sum(axis=1)
```

```
Y/X
0     691
1    4354
2    4170
3     785
dtype: int64
```

Ejercicios a mano

1. La segunda tabla muestra la tabla de frecuencias de la muestra total, ¿cuál es la probabilidad $p(y = 0, x = 1)$?
2. ¿Cuál es la probabilidad $p(x = 1 \mid y = 0)$?
3. Encontrar $F(X \leq 1, Y \leq 2)$

Para el ejercicio 2

```
# Obtener la distribución conjunta de cada celda
xt_prob = freq_table / freq_table.sum().sum()

def highlight_cells(x, row=0, col=1):
    df = pd.DataFrame('', index=x.index, columns=x.columns)
    df.loc[row, col] = 'background-color: red; color: black'
    return df

styled_table = xt_prob.style.apply(highlight_cells, axis=None)
display(styled_table)
```

Table 2

X Y/X	0	1	2
0	0.000000	0.027500	0.041600
1	0.044600	0.257000	0.133800
2	0.125100	0.252300	0.039600
3	0.046200	0.032300	0.000000

La celda en rojo corresponde a la probabilidad

$$p(y=0, x=1) = \frac{325}{10000}$$

Para el ejercicio 3

Ahora, para calcular la probabilidad condicional $p(x=1|y=0)$ debemos obtener la probabilidad marginal de y, o $p(y=0)$.

```
# Calculate row sums
row_sums = freq_table.sum(axis=1)

# Calculate p(y)
xt_py = freq_table.copy()
xt_py['p(y)'] = [f"{row_sum}/{freq_table.sum().sum()}={row_sum/freq_table.sum().sum():.4f}

# Style the table
def highlight_column(x):
    df = pd.DataFrame('', index=x.index, columns=x.columns)
```

```
    df['p(y)'] = 'background-color: #d4d1d1; color: black'
    return df

styled_table = xt_py.style.apply(highlight_column, axis=None)
display(styled_table)
```

Table 3

X	0	1	2	p(y)
Y/X				
0	0	275	416	691/10000=0.0691
1	446	2570	1338	4354/10000=0.4354
2	1251	2523	396	4170/10000=0.4170
3	462	323	0	785/10000=0.0785

Luego, la probabilidad condicional será la probabilidad conjunta $p(y = 0, x = 1)$ dividida por la probabilidad marginal $p(y = 0)$

$$p(x = 1|y = 0) = \frac{p(y = 0, x = 1)}{p(y = 0)} = \frac{325/10000}{725/10000} = \frac{13}{29}$$

En general, la distribución condicional $p(x|y = 0)$ se puede obtener como sigue

```
# Filter rows where Y == 0
filtered_data = sampled_data[sampled_data['Y'] == 0]

# Count occurrences of each X value
count_data = filtered_data['X'].value_counts().reset_index()
count_data.columns = ['X', 'n']

# Calculate the conditional probability p(x|y=0)
count_data['p(x|y=0)'] = count_data['n'] / count_data['n'].sum()

# Display the table (you can further style this table if you'd like)
display(HTML(tabulate(count_data, headers='keys', tablefmt='html')))
```

	X	n	p(x|y=0)
0	2	416	0.602026
1	1	275	0.397974

Para el ejercicio 4.

$F(X \leq 1, Y \leq 2) = p(x = 0, y = 0) + p(x = 0, y = 1) + p(x = 0, y = 2) + p(x = 1, y = 0) + p(x = 1, y = 1) + p(x = 1, y = 2) =$

```
# Create a pivot table
xt = pd.pivot_table(sampled_data, values=None, index='Y', columns='X', aggfunc='size', fil

# Normalize to get probabilities
xt_prob = xt / xt.values.sum()

display(HTML(xt_prob.to_html()))
```

X	0	1	2
Y			
0	0.0000	0.0275	0.0416
1	0.0446	0.2570	0.1338
2	0.1251	0.2523	0.0396
3	0.0462	0.0323	0.0000

La distribución condicional de $x|y = y$ puede verse totalmente diferente que la distribución marginal $p(x)$. En la siguiente imagen, tomada de Bishop (PRML-1.1), se ilustra la distribución de dos variables X y Y. X puede tomar 9 valores posibles, y Y 2. La distribución marginal de X, la distribución independientemente de qué valores tome Y, tiene una forma más achatada (más amplia, con mayor varianza) que la distribución condicional a $Y = 1$.

Definición de Funciones de Masa de Probabilidad (PMF)

Si $\mathcal{X}$ es finito o contable infinito (cada miembro de $\mathcal{X}$ puede asociarse con un elemento de $\mathbb{N}$), entonces X es una variable aleatoria *discreta*. Una función de masa de probabilidad (PMF, por sus siglas en inglés) es una función que describe la probabilidad de que una variable aleatoria discreta tome un valor específico. En otras palabras, para cada posible resultado en el espacio de muestras, la PMF asigna una probabilidad.

Matemáticamente, la PMF $ P(X = x) $ es una función que satisface las siguientes condiciones:

1. **Positividad**: $ P(X = x) 0 $ para todo $ x $.
2. **Normalización**: $ _{x } P(X = x) = 1 $, donde $ $ es el espacio de muestras.

Definición de Funciones de Densidad de Probabilidad (PDF)

Si $X \in \mathbb{R}$ es una función real, se denomina *variable aleatoria continua*, y es más natural especificar que X se encuentre en un intervalo, e.g,, $a \leq X \leq b$, dado que en este caso no se puede crear un conjunto *contable* de valores posibles que X puede tomar, pero podemos crear un conjunto finito de *intervalos* en la linea real, y asociar los eventos de X que están en esos intervalos.

Si el conjunto de posibles valores que una variable aleatoria $ X $ puede tomar es un intervalo (o una unión de intervalos) en la recta real, entonces $ X $ es una variable aleatoria *continua*. Una función de densidad de probabilidad (PDF, por sus siglas en inglés) es una función que describe cómo se distribuyen las probabilidades a lo largo de los posibles valores de una variable aleatoria continua. A diferencia de una PMF, donde se asigna una probabilidad directamente a un valor específico, la PDF asigna una densidad. Para obtener una probabilidad real de que $ X $ caiga en un intervalo específico, se integra la PDF sobre ese intervalo.

Matemáticamente, la PDF $ f(x) $ de una variable aleatoria continua $ X $ satisface las siguientes condiciones:

1. **No Negatividad**: $ f(x) 0 $ para todo $ x $.
2. **Normalización**: $ _{-∞}^{∞} f(x) , dx = 1 $. Esto significa que el área total bajo la curva de la PDF (desde el infinito negativo hasta el infinito positivo) es 1.
3. **Probabilidad en un intervalo**: La probabilidad de que $ X $ esté en el intervalo $ [a, b] $ es $ _{a}^{b} f(x) , dx $.

Nota: Dado que estamos tratando con variables continuas, la probabilidad de que $ X $ tome un valor específico es, en teoría, cero. Es decir, $ P(X = x) = 0 $ para cualquier $ x $. En su lugar, trabajamos con probabilidades en intervalos, como se indicó anteriormente.

PMFs y PDFs frecuentemente usadas

Las PMFs más usadas son:

1. **Bernoulli**: $p(x) = \begin{cases} 1-p & \text{si } x = 0 \\ p & \text{si } x = 1 \end{cases}$
2. **Binomial**: $p(x) = \binom{n}{x} p^x (1-p)^{n-x}$
3. **Poisson**: $p(x) = \frac{\lambda^x}{x!} e^{-\lambda}$
4. **Geométrica**: $p(x) = (1-p)^{x-1} p$

La distribución de Bernoulli es un caso especial de la distribución binomial, cuando $n = 1$.

```
image_path = "img/pmf.png"
display(HTML('<div style="text-align: center;"><img src="{}" width="1000"/></div>'.format(
```

<IPython.core.display.HTML object>

Las PDFs más usadas son:

1. **Uniforme**: $f(x) = \begin{cases} \frac{1}{b-a} & \text{si } a \leq x \leq b \\ 0 & \text{en otro caso} \end{cases}$
2. **Normal**: $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
3. **Exponencial**: $f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$

Distribución acumulada (cdf)

La CDF es estrictamente creciente, y tiene límites 0 y 1.

Sea $A = (X \leq a)$, $B = (X \leq b)$ y $C = (a < X \leq b)$ en donde $a < b$. Tenemos que B es la unión de A y C, es decir, B toma todos los valores que toman A y C. Dado que A y C son mutuamente excluyentes, la probabilidad de B es

$$\Pr(B) = \Pr(A) + \Pr(C) \tag{7}$$

Y la probabilidad de estar en el intervalo C como

$$\Pr(C) = \Pr(B) - \Pr(A) \tag{8}$$

Definimos la distribución de probabilidad acumulada como

$$P(x) \triangleq \Pr(X \leq x)$$

Con esto, podemos calcular la probabilidad de estar en un intervalo como

$$\Pr(a < X \leq b) = P(b) - P(a) \tag{9}$$

La cdf gráficamente tiene esta forma, cuyo eje Y se lee ``probabilidad de que X tome valores igual o menor a x''. Notar que esto es una suma acumulada (o integral) que termina en 1.

CDF para Variables Discretas

Definición: Para una variable aleatoria discreta X, la Función de Distribución Acumulativa $F(x)$ es la probabilidad de que X tome un valor menor o igual a x. La CDF es una función escalonada, ya que aumenta solo en los puntos donde la función de masa de probabilidad (PMF) es positiva.

Matemáticamente:

$$F(x) = P(X \leq x) = \sum \_k \leq xP(X = k)$$

Dónde:

- La suma se extiende sobre todos los valores de k que son menores o iguales a x.
- $F(x)$ es una función escalonada, ya que aumenta solo en los puntos donde la función de masa de probabilidad (PMF) es positiva.

Interpretación: Si tienes una variable aleatoria discreta que representa, por ejemplo, el número de veces que un evento ocurre, la CDF en un punto específico x te dice la probabilidad acumulada de que ese evento ocurra (x) o menos veces.

CDF para Variables Continuas

Definición: Para una variable aleatoria continua X, la Función de Distribución Acumulativa $F(x)$ es la probabilidad de que X tome un valor menor o igual a x.

Matemáticamente:

$$F(x) = P(X \leq x) = \int \_{-}\infty^x f(t)\, dt$$

Dónde:

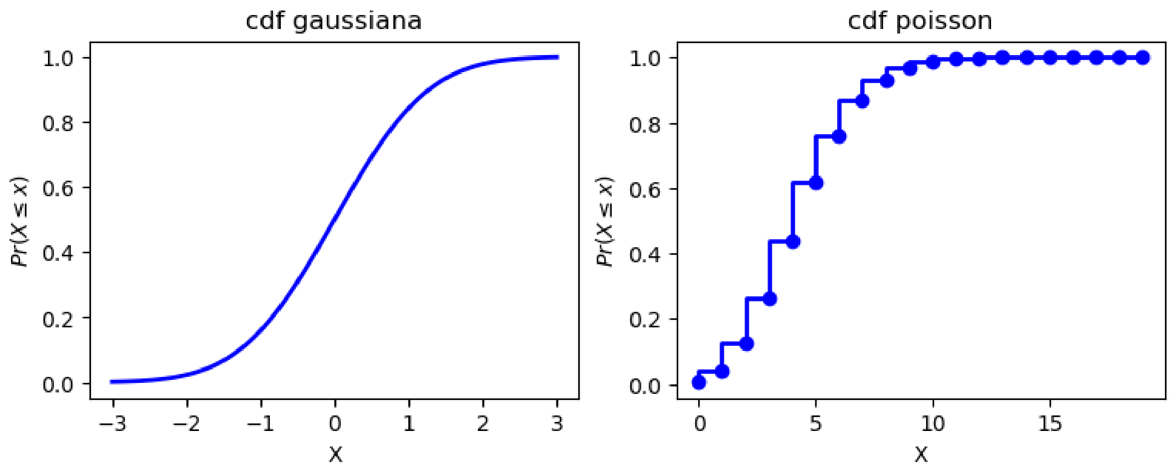
- $f(t)$ es la función de densidad de probabilidad (PDF) de X.
- $F(x)$ es una función continua y no disminuyente.

Interpretación: Si tienes una variable aleatoria continua, como la altura de las personas en una población, la CDF en un punto específico x te dice la probabilidad acumulada de que una persona seleccionada al azar tenga una altura de x o menos.

Para ambas variables, discretas y continuas, la CDF siempre satisface:

1. $0 \leq F(x) \leq 1$
2. $F(-\infty) = 0$ y $F(\infty) = 1$
3. $F(x)$ es no decreciente, es decir, si $a < b$, entonces $F(a) \leq F(b)$.

```python
# Setting up plot dimensions
fig, axes = plt.subplots(1, 2, figsize=(9, 3))

# PDF Gaussiana (Normal)
x = np.linspace(-3, 3, 101)
y = norm.pdf(x, 0, 1)

# CDF Gaussiana
y_cdf = norm.cdf(x, 0, 1)
axes[0].plot(x, y_cdf, color='blue', linewidth=2)
axes[0].set_title('cdf gaussiana')
axes[0].set_xlabel('X')
axes[0].set_ylabel(r"$Pr(X \leq x)$");

# CDF para la distribución de Poisson

# PDF Poisson
x = np.arange(0, 20)
y = poisson.pmf(x, 5)

# CDF Poisson
y_cdf = poisson.cdf(x, 5)
axes[1].step(x, y_cdf, color='blue', linewidth=2)
axes[1].scatter(x, y_cdf, color='blue')
axes[1].set_title('cdf poisson')
axes[1].set_xlabel('X')
axes[1].set_ylabel(r"$Pr(X \leq x)$");
```



Relación entre función de densidad y función de distribución acumulativa

Podemos definir la función de densidad de probabilidad, o *pdf*, como la derivada de la cdf

$$p(x) \triangleq \frac{d}{dx} P(x)$$

Alternativamente, en vez de $p(x)$ y $P(x)$ se suele usar, para VA continuas, $f(x)$ y $F(x)$ respectivamente.

Y viceversa: la cdf como la integral de la pdf. Dada la pdf, podemos calcular la probabilidad de una variable continua en un intervalo finito como sigue

$$\Pr(a < X \leq b) = \int_a^b p(x) dx = P(b) - P(a) \tag{10}$$

La función de densidad de probabilidad debe satisfacer

$$\begin{aligned} &1)\ f(x) \geq 0,\ \text{ para toda } x,\ -\infty < x < \infty \\ &2)\ \int \_{-\infty}^{\infty} f(x) \mathrm{d}x = 1 \end{aligned}$$

Ejemplo con distribución normal

La pdf de una normal está dada por

$$\mathcal{N}(x \mid \mu, \sigma^2) = p(x) = \frac{1}{\sigma\sqrt{2\pi}} e \left\{ \frac{-(x-\mu)^2}{2\sigma^2} \right\} \tag{10}$$

que es gobernada (es decir, su forma y sus valores quedan completamente especificados) por μ, la media o el valor central, y σ^2, llamada la varianza (cuya raíz cuadrada es la desviación estándar σ, y está en las unidades de x y μ).El recíproco de la varianza es llamado *precisión*, $\tau = 1/\sigma^2$.

En la ecuación (10), el término $\frac{1}{\sigma\sqrt{2\pi}}$ es una constante de *normalización*, que asegura que $p(x)$ sume 1, por lo que

$$\int_{-\infty}^{\infty} p(x) \mathrm{d}x = 1$$

Las siguientes expresiones se cumplen en (10). El valor esperado de x es

$$\mathbf{E}[x] = \int_{-\infty}^{\infty} p(x) x \mathrm{d}x = \mu$$

Y el valor esperado de x^2

$$\mathbf{E}[x^2] = \int_{-\infty}^{\infty} p(x) x \mathrm{d}x^2 = \mu^2 + \sigma^2$$

La varianza está dada por $\mathrm{var}[x] = \mathbf{E}[x^2] - \mathbf{E}[x]^2$, por lo que $\mathrm{var}[x] = \sigma^2$.

```python
x = np.random.normal(4, 3, 10000)
print(x.mean())
print(np.mean(x**2) - np.mean(x)**2)
print(np.var(x))
print(np.std(x).round(2))
```

```
3.9881330919624323
9.149508765663048
9.149508765663052
3.02
```

En la siguiente figura a la izquierda se representa una variable aleatoria normal con media 0 y desviación estándar de 1, conocida como distribución normal estándar, y representada como $\mathcal{N}(\mu = 0, \sigma = 1)$ o simplemente $\mathcal{N}(0, 1)$. Para la densidad de probabilidad normal, en Python usamos `scipy.stats.norm.pdf(x, loc=0, scale=1)`, que nos retorna la densidad en `x`. `norm.`, que nos retorna la densidad en `x`.

A la derecha se representa el área equivalente en la cdf, usando `scipy.stats.norm.cdf(x, loc=0, scale=1)`.

```
# Setting up plot dimensions
fig, axes = plt.subplots(1, 2, figsize=(13, 4))

# PDF Gaussiana (Normal)
x = np.linspace(-3, 3, 101)
y = norm.pdf(x, 0, 1)

axes[0].plot(x, y, color='blue', linewidth=2)
axes[0].set_title('pdf gaussiana (normal)')
axes[0].set_ylabel('p(x)')

from_x, to_x = -1.5, 1.5
sx = np.concatenate([[from_x], np.arange(from_x, to_x, 0.01), [to_x]])
sy = np.concatenate([[0], norm.pdf(np.arange(from_x, to_x, 0.01)), [0]])

axes[0].fill(sx, sy, color='#B3B3FF')
axes[0].text(-1, norm.pdf(0, 0) / 2, r"$\int_{-1,5}^{1.5} p(x) dx$", fontsize=12)
axes[0].text(-1, norm.pdf(0, 0) / 3, "= P(1.5) - P(-1.5)", fontsize=12)

# CDF Gaussiana
y_cdf = norm.cdf(x, 0, 1)
axes[1].plot(x, y_cdf, color='blue', linewidth=2)
axes[1].set_title('cdf gaussiana')
axes[1].set_xlabel('X')
axes[1].set_ylabel(r"$Pr(X \leq x)$")

sy_cdf = np.concatenate([[0], norm.cdf(np.arange(from_x, to_x, 0.01)), [0]])
axes[1].fill_between(sx, sy_cdf, color='#B3B3FF')

plt.show()
```