DIPLOMADO EN
CIENCIA DE DATOS -
COHORTE 2024

Curso: Fundamentos de Probabilidad

Bienvenido al curso *"Fundamentos de Probabilidad"*. Durante este curso, nos sumergiremos en los conceptos esenciales de la teoría de la probabilidad, abordando temas como probabilidad, probabilidad condicional, variables aleatorias, distribuciones de probabilidad, distribuciones continuas, esperanza matemática y entropía.

A lo largo de este curso de Fundamentos de Probabilidad, exploraremos ejemplos que ayudarán a clarificar los conceptos presentados, al mismo tiempo que proporcionarán una comprensión concreta de las capacidades introductorias de esta disciplina. El curso está diseñado para guiar gradualmente a los estudiantes a través de los aspectos fundamentales en la teoría de la probabilidad, junto con el uso de métodos numéricos básicos que amplían las capacidades de análisis en Python como lenguaje de programación.

Contenidos:

1. **Introducción a Probabilidad**
- Sistemas Complejos
1. **Probabilidad**
- Introducción y espacios muestrales
- Evento
- Provabilidad
- Regla Aditiva de la Probabilidad
- Media de todo el especio muestral
- Probabilidad del complemento de un evento
1. Probabilidad Condicional
- Regla multiplicativa de la probabilidad
- Independencia
1. Variables Aleatoria
- Variable Aleatoria
- Calculo de frecuencias
- Función de Probabilidad de una variable discreta
- Cálculo de probabilidades

1. Distribuciones de Probabilidad
- Interpretación de la función de probabilidad
- Distribución de Bernoulli
- Distribución Binomial
1. Distribuciones Continuas
- Distribución Gaussiana
- De Binomial a Gaussiana
- Distribución de Poisson
1. Variables Categóricas
2. Esperanza Matemática de una variable
- Esperanza de la distribución Binomial
- Esperanza de la distribución de Poisson
- Varianza y desviación estándar de una variable aleatoria
- Desviación Estándar
1. Entropía
- Entropía de una variable aleatoria
- Entropía de Shannon

---

1. Introducción a Probabilidad

Introducción a Probabilidad

Cómo interpretar la ciencia de datos en nuestra realidad...

Recordemos: La ciencia de datos se ocupa del análisis, interpretación, visualización y utilización de datos para tomar decisiones basadas en información.

Podemos hacer una analogía con una de las ramas de la física que utiliza conceptos estadísticos para describir el comportamiento de sistemas compuestos por muchas partículas: la mecánica estadística.

En Sistemas Complejos (debidos a muchos objetos):

Analogías:

1. **Gran cantidad de datos (sistemas microscópicos)**: En la ciencia de datos, trabajamos con grandes volúmenes de datos, que pueden compararse a los numerosos sistemas microscópicos en un conjunto canónico. Cada punto de datos o

conjunto de datos puede considerarse como un sistema individual con su propia "energía" (información o características únicas).

2. **Análisis estadístico (equilibrio térmico)**: Así como un conjunto canónico utiliza herramientas estadísticas para describir el comportamiento de sistemas en equilibrio térmico, la ciencia de datos utiliza métodos estadísticos y de aprendizaje automático para analizar y extraer patrones o tendencias de grandes conjuntos de datos. Este proceso puede considerarse análogo al establecimiento de "equilibrio" en el sentido de que buscamos comprender las propiedades generales de los datos, a pesar de la variabilidad o "fluctuaciones" individuales.
3. **Propiedades macroscópicas (insights y decisiones)**: En mecánica estadística, a pesar de la complejidad y variabilidad de cada sistema microscópico, las propiedades macroscópicas del conjunto pueden predecirse y entenderse. De manera similar, en la ciencia de datos, buscamos comprender las propiedades generales de los datos (tendencias, patrones, correlaciones) para tomar decisiones informadas o ganar insights, a pesar de la complejidad y variabilidad de los datos individuales.
4. **Temperatura fija (hipótesis o marco de trabajo)**: La temperatura fija en un conjunto canónico representa las condiciones controladas o constantes bajo las cuales se estudian los sistemas. En ciencia de datos, esto puede compararse con el marco teórico, hipótesis o algoritmos específicos que aplicamos para analizar los datos. Estas "condiciones" guían nuestra exploración y análisis de los datos, permitiéndonos extraer información significativa.

2. Probabilidad

2.1. Introducción y espacios muestrales

Introducción

Partamos del hecho de pensar que nuestro mundo es... INCERTIDUMBRE. Absolutamente todo en nuestra realidad física tiene una incertidumbre. Es por esta razón que existe el campo de estudio de la Probabilidad: para dar un acercamiento objetivo de la realidad mediante postulados y leyes matemáticas que describen los componentes de un sistema que están sometidos a incertidumbre.

Iniciemos este fascinante mundo explorando los conceptos más fundamentales!

Espacio muestral

La siguiente imagen representa un ejemplo de un `espacio muestral`, el cual denotaremos como M. Cada objeto o bola es un elemento del espacio muestral. Esto significa que este espacio muestral tiene N elementos. Se supone que cada objeto puede identificarse de manera única. En este ejemplo hemos usado un identificador $1, 2, \ldots, N$, para cada uno de los elementos del espacio muestral, donde cada objeto tiene un atributo de color: azul, rojo o gris.

```python
#Para crear nuestro espacio muestral aleatorio:

import random

# Definir el espacio muestral como una lista de resultados posibles
espacio_muestral = ['cara', 'cruz', 'neutro']

# Seleccionar un resultado aleatorio del espacio muestral
resultado = random.choice(espacio_muestral)

print(resultado)
```

```
cruz
```

```python
#Para crear nuestro espacio muestral discreto:

import numpy as np

# Definir un espacio muestral como un array de numpy, por ejemplo,
lanzar un dado
espacio_muestral = np.array([1, 2, 3, 4, 5, 6])

# Seleccionar un resultado aleatorio del espacio muestral
resultado = np.random.choice(espacio_muestral)

print(resultado)
```

```
2
```

```python
#Para crear nuestro espacio muestral continuo:

# Generar 1000 muestras de una distribución normal con media 0 y
desviación estándar 1

muestras = np.random.normal(loc=0, scale=1, size=1000)

print(muestras[:10])
```

```
[ 1.52602448 -0.08044835  0.26894374 -0.50699429 -0.03495826
0.84669873
  0.42992717 -0.87352896  0.53557042  0.01357136]
```

2.2. Evento

Evento

Un evento es cualquier subconjunto del espacio muestral. El lector interesado puede verificar, si lo desea, que el espacio muestral M tiene exactamente 2^N subconjuntos.

Consideremos ahora cuatro eventos (subconjuntos) especiales de M:

1. *azul*: el subconjunto de bolas azules;
2. *gris*: el subconjunto de bolas grises.
3. *pares*: el subconjunto de bolas pares.
4. *impares*: el subconjunto de bolas impares. Extra: *pares azules*: el subconjunto de bolas pares azules.

Prueba: Para verificar que un espacio muestral de (N) elementos tenga (2^N) subconjuntos (eventos), se puede emplear el principio de la teoría de conjuntos y la combinatoria. Este fenómeno se explica a través del concepto de las partes de un conjunto, también conocido como el conjunto potencia. El conjunto potencia de cualquier conjunto es el conjunto de todos sus posibles subconjuntos, incluyendo el conjunto vacío y el conjunto en sí.

Para entender paso a paso por qué un espacio muestral de (N) elementos tiene (2^N) subconjuntos, podemos describirlo así:

1. **Elemento Individual**: Para cada elemento en el espacio muestral, hay dos opciones: o el elemento está presente en un subconjunto, o no lo está.

2. **Combinaciones de Elementos**: Dado que cada elemento del espacio muestral puede estar presente o no en un subconjunto, y considerando que el espacio muestral tiene (N) elementos, entonces para cada elemento hay 2 opciones. Esto nos da un total de (2 x 2 x 2 x ... x 2) (con (N) factores de 2), lo cual es igual a (2^N).

3. **Conjunto Potencia**: Matemáticamente, el conjunto de todos los posibles subconjuntos de un conjunto se llama su conjunto potencia. Si el conjunto tiene (N)

elementos, entonces el tamaño de su conjunto potencia (el número de subconjuntos posibles) es (2^N).

Ejemplo: Si tienes un espacio muestral con 3 elementos (por ejemplo, {A, B, C}), los posibles subconjuntos serían: {}, {A}, {B}, {C}, {A, B}, {A, C}, {B, C}, {A, B, C}. Aquí puedes ver que hay (2^3 = 8) subconjuntos, lo que coincide con nuestra explicación.

Esta propiedad se puede verificar empíricamente para espacios muestrales pequeños mediante la enumeración de todos los posibles subconjuntos. Para espacios muestrales grandes, se puede demostrar matemáticamente o mediante programación, usando algoritmos que generen todos los posibles subconjuntos de un conjunto dado y luego contando el número total de subconjuntos generados para asegurar que coincide con (2^N).

```
#Para nuestro espacio muestral aleatorio:

import random

# Definir el espacio muestral
espacio_muestral = ['cara', 'cruz']

# Definir el evento deseado
evento = 'cara'

# Realizar N experimentos
N = 1000
eventos_favorables = sum(random.choice(espacio_muestral) == evento for
_ in range(N))

# Calcular la probabilidad del evento
probabilidad = eventos_favorables / N
print(f"Probabilidad de obtener {evento}: {probabilidad}")
```

```
Probabilidad de obtener cara: 0.486
```

```
#Para nuestro espacio muestral discreto:
import numpy as np

# Definir el espacio muestral
espacio_muestral = np.array([1, 2, 3, 4, 5, 6])

# Realizar N experimentos
N = 1000
resultados = np.random.choice(espacio_muestral, size=N)

# Definir el evento deseado (resultado par)
eventos_favorables = np.sum(resultados % 2 == 0)

# Calcular la probabilidad del evento
probabilidad = eventos_favorables / N
print(f"Probabilidad de obtener un número par: {probabilidad}")
```

```
Probabilidad de obtener un número par: 0.489
```

```python
#Para nuesrro espacio muestral continuo:
import numpy as np

# Generar muestras de una distribución normal estándar
N = 1000
muestras = np.random.normal(loc=0, scale=1, size=N)

# Definir el evento deseado (muestra < 0)
eventos_favorables = np.sum(muestras < 0)

# Calcular la probabilidad del evento
probabilidad = eventos_favorables / N
print(f"Probabilidad de que una muestra sea menor que 0:
{probabilidad}")
```

```
Probabilidad de que una muestra sea menor que 0: 0.46
```

2.3. Probabilidad

Probabilidad

La teoría de probabilidad se creó para poder medir los subconjunto de una espacio muestral. El concepto de medida en cada caso está asociado a la naturaleza del experimento que se desea modelar (representar de manera abstracta).

En el ejemplo de las bolas presentado arriba, la medida que usaremos será la proporción entre el número de elementos de un evento y el número de elementos del espacio muestral M.

Esto significa que:

$$\begin{array}{ll}
\text{Prob}[\text{azul}) & ¿Bolas_N azules/N \\
\text{Prob}[\text{rojo}) & ¿Bolas_N rojas/N \\
\text{Prob}[\text{gris}) & ¿Bolas_N grises/N \\
\text{Prob}[\text{pares}) & ¿Pares_N/N \\
\text{Prob}[\text{impares}) & ¿Impares_N/N
\end{array}$$

Observe además que $\text{Prob}[M)=1$.

En las siguientes secciones presentamos las propiedades esenciales de la probabilidad.

Son conceptos relativamente sencillos pero muy importantes.

2.4. Regla aditiva de la probabilidad

Regla aditiva de la probabilidad

La probabilidad de la unión de dos eventos (subconjuntos) disyuntos (que no tiene intersección) es la suma de la probabilidad (medida) de cada uno de ellos. En símbolos, si A y B son eventos disyuntos de M, entonces

$$\mathrm{Prob}[A \cup B] = \mathrm{Prob}[A] + \mathrm{Prob}[B].$$

Por ejemplo, observe que:

Para N= 20, con Azules=5 , Rojos=7, pares = 10 \

$$\mathrm{Prob}[\mathrm{azul} \cup \mathrm{rojo}] = 5/20 + 7/20 = 12/20.$$

Lo anterior conduce a la regla aditiva general la cual dice que

$$\mathrm{Prob}[A \cup B] = \mathrm{Prob}[A] + \mathrm{Prob}[B] - \mathrm{Prob}[A \cap B].$$

En el ejemplo se tiene entonces que

$$\mathrm{Prob}[\mathrm{azul} \cup \mathrm{pares}] = 5/20 + 10/20 - 3/20 = 12/20$$

Nota: Un conjunto disyunto se refiere a dos o más conjuntos que no tienen elementos en común. Es decir, la intersección de estos conjuntos es vacía.

En el contexto de la probabilidad, cuando se dice que dos eventos son mutuamente excluyentes o disyuntos, significa que la ocurrencia de uno de los eventos excluye automáticamente la ocurrencia del otro. Por lo tanto, no pueden ocurrir al mismo tiempo.

Por ejemplo: si consideramos el lanzamiento de un dado de seis caras, los eventos "obtener un número par" y "obtener un número impar" son mutuamente excluyentes, ya que un resultado específico no puede ser ambos, par e impar al mismo tiempo. En este caso, no hay ningún elemento (resultado del dado) que pertenezca a ambos conjuntos (par e impar), lo que los hace conjuntos disyuntos.

2.4.1. Ejercicio

Ejercicio

Que piensa de la siguiente afirmación. ¿Verdadero o falso? Justifique su respuesta.

Si A y B son conjuntos disyuntos, entonces $\mathrm{Prob}[A \cap B] = 0$.

```
# Demostración:

# Definir dos conjuntos disyuntos
A = {1, 2, 4}
B = {4, 5, 6}
```

```python
# Calcular la intersección de A y B
A_inter_B = A.intersection(B)

# Verificar si la intersección es el conjunto vacío y calcular la probabilidad
prob_A_inter_B = 0 if len(A_inter_B) == 0 else "No aplicable, ya que A y B no son disyuntos"

print(f"Intersección de A y B: {A_inter_B}")
print(f"Probabilidad de A ∩ B: {prob_A_inter_B}")
```

```
Intersección de A y B: {4}
Probabilidad de A ∩ B: No aplicable, ya que A y B no son disyuntos
```

2.5. Medida de todo el espacio muestral

Medida de todo el espacio muestral

Vamos a denotar por $\varnothing$ al conjunto vacío, es decir un conjunto que no tiene elementos.

En nuestro ejemplo tenemos que

$$M = \text{azul} \cup \text{rojo} \cup \text{gris}$$

.

Además se tiene que

$$\begin{aligned} \text{azul} \cap \text{rojo} &¿ \varnothing \\ \text{azul} \cap \text{gris} &¿ \varnothing \\ \text{gris} \cap \text{rojo} &¿ \varnothing \end{aligned}$$

Se dice en esta situación que los conjuntos son `mutuamente excluyentes`. De acuerdo con la regla aditiva tenemos que

$$\text{Prob}[M] = \text{Prob}[\text{azul}] + \text{Prob}[\text{rojo}] + \text{Prob}[\text{gris}] = 5/20 + 7/20 + 8/20 = 1\,.$$

Esta es una propiedad general de la probabilidad. El espacio muestral siempre tiene medida de probabilidad 1.

Además observe que si se tienen eventos disyuntos entre sí (mutuamente excluyentes), cuya unión es el espacio muestral, entonces la probabilidad de la unión de todos esos eventos tiene probabilidad 1.

2.6. Probabilidad del complemento de un evento

Probabilidad del complemento de un evento

El complemento de un evento A se denotará por A^c. Este simplemente el conjunto de elementos del espacio muestral que están por fuera de A. Entonces, es inmediato que $M=A \cup A^c$. Por lo que

$$Prob[A^c]=1-Prob[A].$$

Una consecuencia inmediata de esta propiedad es que como $M^c=\varnothing$, porque el espacio muestral contiene a todos los elementos, entonces $Prob[\varnothing]=0$.

En nuestro ejemplo $impares^c=pares$. Entonces $Prob[\text{impares}]=1-9/20=11/20$.

3. Probabilidad Condicional

Probabilidad Condicional

El concepto de probabilidad condicional es de vital importancia en el estudio del aprendizaje profundo y la inteligencia artificial.

Como el nombre parece indicar, es trata de calcular la probabilidad de un evento sujeto a una restricción. En realidad es así y la restricción normalmente está asociada con otro evento.

Para ilustrar el asunto, supongamos que se pregunta por la probabilidad que una bola extraída sea par, dado que la bola es azul.

Se observa entonces, que se da una información antes de calcular la probabilidad de ser par. Esta información corresponde al evento *azul*. Escribiremos

$$\text{Prob}(\text{par} \vee \text{azul})$$

Para hacer el cálculo correcto, se procede de la siguiente manera: Primero se reduce el espacio muestra a *azul*. En el ejemplo se tiene que

$$\text{azul}=\{5,7,8,10,16\}.$$

Ahora que se ha restringido el espacio muestral a *azul*, se calcula la probabilidad de interés. En este caso *par*. Observe entonces que

$$\text{Prob}(\text{par} \vee \text{azul})=3/5,$$

porque en el evento *azul* que tiene 5 elementos hay 3 *pares*.

Puede verificarse que

$$\text{Prob}(\text{par} \vee \text{azul})=\frac{\text{Prob}[\text{par} \cap \text{azul})}{\text{Prob}[\text{azul})}$$

Esta es una regla general, que se enuncia a si. Si A y B son eventos del espacio muestral M, entonces se define $\mathrm{Prob}\left[A \vee B\right)$ como

$$\mathrm{Prob}\left[A \vee B\right) = \frac{\mathrm{Prob}\left[A \cap B\right)}{\mathrm{Prob}\left[B\right)}$$

3.1. Regla multiplicativa de la probabilidad

Regla multiplicativa de la probabilidad

De la definición de la probabilidad condicional $\mathrm{Prob}\left[A \vee B\right)$ se desprende que

$$\mathrm{Prob}\left[A \cap B\right) = \mathrm{Prob}\left[B\right) \times \mathrm{Prob}\left[A \vee B\right)$$

3.1.1. Ejemplo

Ejemplo

Con nuestro ejemplo supongamos que se pregunta por la probabilidad de obtener una bola par azul en un experimento.

La solución es sencilla, por que ya hemos obtenido que $\mathrm{Prob}\left(\mathrm{par} \vee \mathrm{azul}\right) = 3/5$, y $\mathrm{Prob}\left[\mathrm{azul}\right) = 5/20$. por lo tanto

$$\mathrm{Prob}\left[\mathrm{par} \cap \mathrm{azul}\right) = \mathrm{Prob}\left[\mathrm{azul}\right) \times \mathrm{Prob}\left[\mathrm{par} \vee \mathrm{azul}\right) = 5/20 \times 3/5 = 3/20$$

Esto está de acuerdo con la ilustración de los evento del espacio muestral exhibidos arriba.

3.2. Independencia

Independencia entre eventos

Dos eventos A y B del espacio muestral M se dicen independientes si

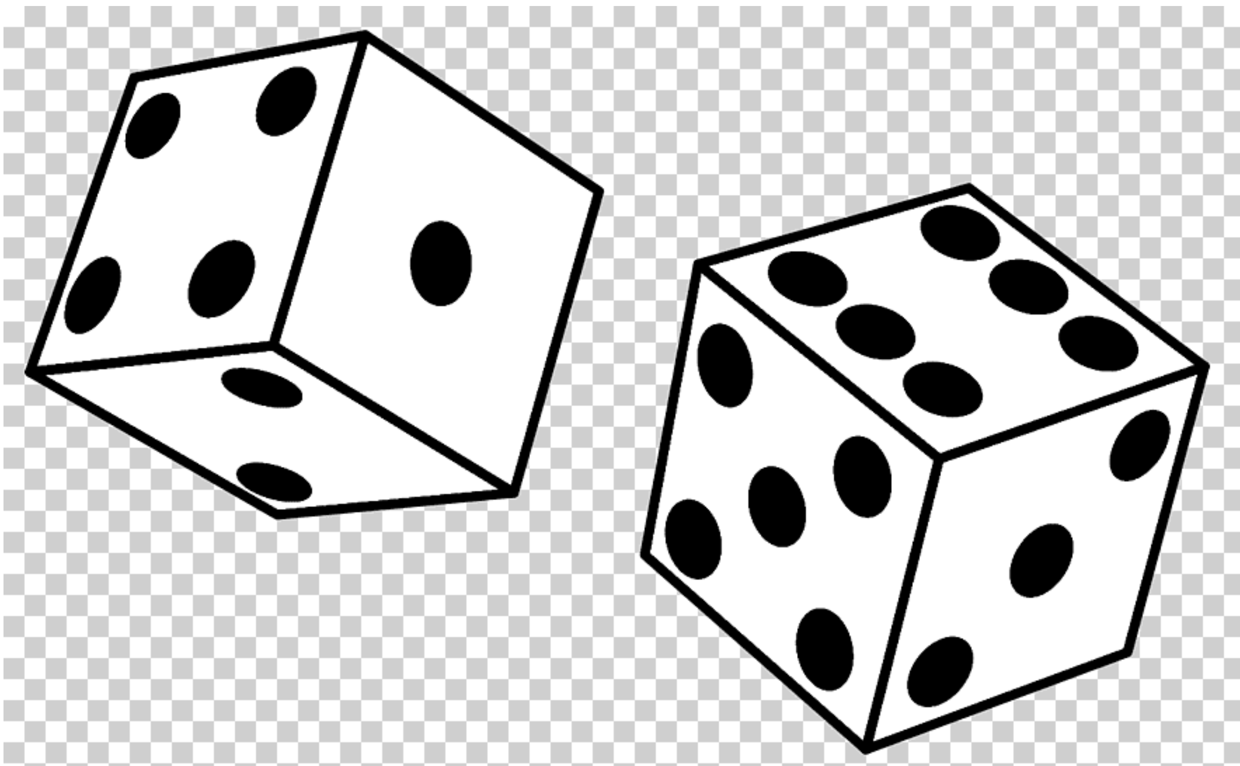
$$\mathrm{Prob}\left[A \cap B\right) = \mathrm{Prob}\left[A\right) \times \mathrm{Prob}\left[B\right).$$

Esta definición es bastante técnica, pero intuitivamente puede entenderse como que la ocurrencia de un evento no afecta la ocurrencia del otro. Observe que en este caso se tiene que

$$\mathrm{Prob}\left[A \vee B\right) = \mathrm{Prob}\left[A\right).$$

4. Variables Aleatorias

Variables Aleatorias



Como ya sabe del ejercicio hay 36 posibles resultados. Es decir el espacio muestral M tiene 36 elementos y 2^{36} eventos.

Supongamos que (x, y) es un elemento de M. Observe que (x, y) es en realidad una pareja de números. Por ejemplo $(3, 4)$. Definamos ahora la siguiente función:

$$f(x, y) = x + y,$$

es decir la función f simplemente calcula la suma de los dos números. Como el espacio muestral tiene 36 elementos, entonces hay 36 posibles resultados, aunque no todos son diferentes.

La función f se llama una **variable aleatoria**. Veamos algunos posibles valores que puede tomar la función:

$$\begin{array}{ll} f(1,1) & ¿2 \\ f(3,2) & ¿5 \\ f(4,3) & ¿7 \\ f(3,4) & ¿7 \\ f(6,5) & ¿11 \end{array}$$

4.1. Variable Aleatoria

Variable Aleatria

Dado un espacio muestral M, es una función que asigna a los elementos del espacio muestral un número real.

En el espacio muestral de los dos dados de seis lados, se pueden definir muchas variables aleatorias. Considere por ejemplo la variable aleatoria definida por $g(x, y) = x \times y$.

#####4.1.1. Ejercicio

```
# Considere el ejemplo del espacio muestral de los dos dados de seis caras, no cargados.

# libreria numérica
import numpy as np

# Construir la tabla completa de los posibles resultados de de las funcioens  f  y  g

# Construir tensor 36 x 2
M = np.zeros((36,2))
k = 0
for i in range(1,7):
    for j in range(1,7):
        M[k,] = (i,j)
        k+=1


# Proponer otra variable aleatoria definida sobre el espacio muestral M $.
f = np.zeros(36)
for k in range(36):
    f[k] = M[k,0] + M[k,1]


# presenta los resultados
for k in range(36):
    print('f(',M[k,0],',',M[k,1],') =', f[k])
```

4.2. Cálculo de frecuencias

Cálculo de frecuencias

Observe que en el ejemplo anterior, la variable aleatoria toma once(11) posibles valores diferentes:

$$f=\{2,3,4,5,6,7,8,9,10,11,12\}$$

.

Entonces tenemos lo siguiente: 36 posibles valores, pero solamente 11 diferentes. El siguiente fragmento de código muestra como obtener los posibles valores diferentes que toma la variable aleatoria f.

```
print(np.unique(f))
```

La siguiente función muestra un camino de como calcular la tabla de frecuencias de un conjunto de datos, puestos en una lista de Python.

```
# Función Python para contar los elementos la fecuencia
# de los elementos de una lista, usando un diccionario

def CountFrequency(my_list):

    # Creating an empty dictionary
    freq = {}
    for item in my_list:
        if (item in freq):
            freq[item] += 1
        else:
            freq[item] = 1
    return freq
```

Asegúrese de entender completamente el código. Vamos a probar la función con dos ejemplos. Primero calculamos la frecuencia de los valores de la variable aleatoria f.

```
frec = CountFrequency(f)

for key, value in frec.items():
    print ("% d : % d"%(key, value))

frec
```

Podemos manipular directamente los valores del diccionario. Por ejemplo para calcular la frecuencia de 7, que ya sabemos es 6, se escribe

```
frec[7]
```

En un segundo ejemplo, consideremos el conjunto $ W= {A,B,B,C,A,A }$

```
W = ['A','B','B','C','A','A']
frec = CountFrequency(W)
frec['B']
```

4.3. Función de Probabilidad de una variable numérica discreta

---

Función de Probabilidad de una variable numérica discreta

Ahora que hemos aprendido a calcular tablas de frecuencias de conjuntos de datos organizados en una lista, vamos a introducir el concepto clave de función de probabilidad. En esta sección consideramos variables aleatorias numéricas discretas como la función f asociada al espacio muestral del lanzamiento de dos dados no cargados.

Como observamos antes, la variable aleatoria f es discreta (particularmente es finita), debido a que toma únicamente 11 posibles valores numéricos. Por otro lado sabemos que el resultado del experimento de lanzar los dados puede arrojar 36 posibles resultados. La **función de probabilidad de la variable aleatoria** *f*, que notaremos p_f se define como la probabilidad de obtener cada posible valor de *f*.

Entonces como vimos arriba tenemos que la función de probabilidad de la variable aleatoria f esta definida por extensión de la siguiente forma:

$$\begin{aligned}
p_f(2) & \text{¿} 1/36 \\
p_f(3) & \text{¿} 2/36 \\
p_f(4) & \text{¿} 3/36 \\
p_f(5) & \text{¿} 4/36 \\
p_f(6) & \text{¿} 5/36 \\
p_f(7) & \text{¿} 6/36 \\
p_f(8) & \text{¿} 5/36 \\
p_f(9) & \text{¿} 4/36 \\
p_f(10) & \text{¿} 3/36 \\
p_f(11) & \text{¿} 2/36 \\
p_f(12) & \text{¿} 1/36
\end{aligned}$$

Podemos obtener una imagen de la función, utilizando un histograma. Un histograma es un gráfico de una función de probabilidad de una variable numérica. El siguiente código Python muestra como construir un histograma de la variable aleatoria *f*.

```
import matplotlib.pyplot as plt
plt.hist(f, bins=11,density=True)
plt.title('Funcion de probabilidad de la V.A. $f$')
plt.xlabel('Valores de  $f$')
```

```
plt.ylabel('Probabilidad')
plt.show()
```

4.4. Cálculo de algunas probabilidades

Cálculo de algunas probabilidades

Dado que la variable aleatoria f es numérica, podemos calcular la probabilidad de diferentes eventos del espacio muestral, basados en el valor de f. Veamos algunos ejemplos.

$$\begin{array}{ll} \text{Prob}[f \text{ es par}] & ¿18/36 \\ \text{Prob}[f \leq 6] & ¿15/36 \\ \text{Prob}[f > 10] & ¿3/36 \\ \text{Prob}[f \text{ es par} \vee f < 5] & ¿4/6 \end{array}$$

5. Distribuciones de Probabilidad

5.1. Interpretación de la función de probabilidad

Interpretación de la función de probabilidad

Supongamos que el experimento de lanzar se repite muchas veces. Digamos 100 veces. Como la probabilidad de obtener digamos $f=7$ es $6/36$, entonces lo que se espera que ocurra es que en los 100 lanzamientos se obtenga un valor cercano a

$$\text{Número de veces que se espera que ocurra}\{f=7\} = 6/36 \times 100 \approx 17.$$

Por supuesto, no necesariamente el resultado será 17. Pero si un número cercano. Por ejemplo no esperamos que no ocurra ninguna vez o que ocurra todas las veces el resultado $\{f=7\}$.

4.5.1.Ejemplos de variables numéricas discretas

5.2. Variable de Bernoulli (Distribución de Bernoulli)

Variable de Bernoulli (Distribución de Bernoulli)

Una variable aleatoria f es Bernoulli, si solamente toma dos posibles valores, los cuales por convención son $\{0,1\}$. Es común llamar al resultado 1 como *éxito* y a cero como *fallo*. Observe que en este caso se tiene que la función de probabilidad es dada por

$$p_f(1) \quad ¿\pi$$
$$p_f(0) \quad ¿1-\pi\,,$$

en donde $\pi = \mathrm{Prob}(f=1)$. En este caso, se puede escribir la función en una forma más compacta como

$$p_f(x) = \pi^x (1-\pi)^{1-x}, x=0,1.$$

Asegúrese de entender la fórmula anterior.

Un ejemplo es el siguiente. En el experimento del lanzamiento dados no cargados, definimos la variable aleatoria g como sigue:

$$g = \begin{pmatrix} 1, \text{si la suma de los dos dados es par} \\ 0, \text{si la suma de los dos dados es impar} \end{pmatrix}$$

Verifique que en este caso $\pi = 18/36 = 1/2$.

5.3. Distribuciones discretas (Binomial)

Distribuciones discretas (Binomial)

Recordando que la distribución binomial está dada por:

$$P(k,n;p) = \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

donde $P(k,n;p)$ representa la probabilidad de obtener k éxitos de n intentos con posibilidad **binaria** (por ejemplo, lanzamientos de moneda).

`Ejemplo`: la probabilidad de obtener 4 caras a partir de 10 lanzamientos consecutivos de moneda, está dada por (tomando $p=0.5$, por lo tanto $1-p=0.5$):

$$P(k=4, n=10; p=0.5) = \binom{10}{4} \left(\frac{1}{2}\right)^{10} = \frac{10!}{4!\,6!} \left(\frac{1}{2}\right)^{10}$$

Ahora, la probabilidad de obtener k o menos éxitos a partir de n intentos está dada por la distribución acumulada:

$$C(k,n;p) = \sum_{i=0}^{k} P(i,n;p) = \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i}$$

Por convención entendemos que:

$$C(k=3, n=6; p=0.5) = P(k \leq 3, n=6, p=0.5)$$

`Ejemplo`: la probabilidad de obtener 3 o menos caras a partir de 6 lanzamientos consecutivos está dada por (tomando $p=0.5$, por lo tanto $1-p=0.5$):

$$P(k \leq 3, n=6; p=0.5) = \sum_{i=0}^{3} \binom{6}{i} \left(\frac{1}{2}\right)^6$$

$$P(k \leq 3, n=6; p=0.5) = \left(\frac{1}{2}\right)^6 \sum_{i=0}^{3} \binom{6}{i}$$

$$P(k \leq 3, n=6; p=0.5) = \left(\frac{1}{2}\right)^6 \left\{ \binom{6}{0} + \binom{6}{1} + \binom{6}{2} + \binom{6}{3} \right)$$

Consideremos la variable aleatoria g definida arriba. Pero ahora vamos a considerar que hacemos el experimento digamos $N=3$ veces y contamos el número de veces en que $g=1$. Por otro lado vamos a suponer que que hay un sesgo en el experimento (una especie de dados cargados y se verifica que $\mathrm{Prob}[g=1]=0.6$. Por lo tanto se tiene que $\mathrm{Prob}[g=0]=0.4$.

Entonces obtenemos una variable aleatoria que llamaremos q y diremos que q es una variable Binomial. La función de probabilidad de la variable aleatoria q es dada por extensión como sigue. La variable q toma 4 posibles valores, $q=\{0,1,2,3\}$

Veamos como calcular las probabilidades de los valores de q. La siguiente tabla muestra como calcular tales probabilidades.

Valor	Experimentos	probabilidad cada experimento	probabilidad para este valor de f	total
0	000	$0.4 \times 0.4 \times 0.4$	0.064	0.064
1	100	$0.6 \times 0.4 \times 0.4$	0.096	
1	010	$0.4 \times 0.6 \times 0.4$	0.096	
1	001	$0.4 \times 0.4 \times 0.6$	0.096	0.288
2	110	$0.6 \times 0.6 \times 0.4$	0.144	
2	011	$0.4 \times 0.6 \times 0.6$	0.144	
2	101	$0.6 \times 0.4 \times 0.6$	0.144	0.432
3	111	$0.6 \times 0.6 \times 0.6$	0.216	0.216

Entonces se tienen que

$$\begin{array}{ll} p_q(0) & ¿\,0.064 \\ p_q(1) & ¿\,0.288 \\ p_q(2) & ¿\,0.432 \\ p_q(3) & ¿\,0.216 \end{array}$$

El siguiente código muestra como calcular la función de probabilidad de la variable Binomial q y como obtener un gráfico de la función.

```
import numpy as np
from scipy.stats import binom
import matplotlib.pyplot as plt
```

```
N, p  = 3, 0.6
p_q = np.zeros(4)

for k in range(4):
    p_q[k] = binom.pmf(k,N,p)

print('p_q=',p_q)
label = ['0','1','2','3']
plt.bar(label, p_q, color = 'orange', edgecolor='blue',width=1)
plt.title('Funcion de probabilidad distribución $Binomial(3,0.6)$')
plt.xlabel('Valores de  de la variable aleatoria')
plt.ylabel('Probabilidad')
plt.show()
```

Note que la función de probabilidad de Bernoulli es un caso especial de la Binomial con $N=1$.

Existe una formula general para la función de probabilidad de una variable Binomial. Si se supone que se hacen N experimentos de Bernoulli realizados de manera independientes con la misma probabilidad de éxito (obtener 1), y se anota el número de unos (*éxitos*) obtenidos, entonces la probabilidad de obtener k exitos es dada por

$$p_{bin}[k]=\binom{N}{k}\pi^k(1-\pi)^{N-k},$$

en donde $\binom{N}{k}$ es el símbolo combinatorio y π es la probabilidad de éxito de cada experimento de Bernoulli.

Nota.

Recuerde que el combinatorio $\binom{N}{k}$ es el número de grupos diferentes de tamaño k que se pueden formar teniendo N elementos. Consulte cualquier libro de probabilidad para los detalles matemáticos.

5.4. Ejercicios

Calcula a mano las siguientes probabilidades (tomando $p=0.5$, por lo tanto $1-p=0.5$):

1. Probabilidad de obtener 3 caras a partir de 12 lanzamientos de moneda. R = 0,054
2. Probabilidad de obtener 5 o menos caras a partir de 10 lanzamientos de moneda. R = 0.62
3. Probabilidad de obtener menos de 6 caras a partir de 10 lanzamientos de moneda. R = 0.62

Calcula a mano las mismas probabilidades anteriores pero considerando ahora $p=0.3$.

1. R = 0.24

2. R = 0.95

5.5. Bonus en Python

```
from math import factorial

# definición de la distribución binomial:

def my_binomial(k, n, p):
  return factorial(n)/(factorial(k)*(factorial(n-k)))*pow(p,k)*pow(1-
p, n-k)
```

Usando la función `my_binomial()`, definida previamente, verifica el cálculo de todas las probabilidades del punto anterior.

Ejemplo:

$$P(k \leq 3, n=6, p=0.5)$$

Se traduce en :

```
total = 0
for n in range(4):
  total += my_binomial(i,6,0.5)

print(total)


def c_binomial(k, n, p):
  total = 0
  for i in range(k+1):
    total += my_binomial(i,n,p)
  return total
print(c_binomial(5,10,0.5))
```

6. Distribuciones Continuas

6.1. Distribución Gausiana

Distribuciones Gaussiana

Recordemos que la distribución de probabilidad normal o gaussiana está dada por:

$$P(X) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2\right)$$

donde:

- μ: media de la distribución
- σ: desviación estandar de la distribución

`Ejemplo`: considerando una variable aleatoria que sigue una distribución normal con media $\mu=4$ y desviación estándar $\sigma=0.3$, la probabilidad de que dicha variable tome el valor de 0.2 está dada por:

$$P(0.2)=\frac{1}{0.3\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{0.2-4}{0.3}\right)^2\right)$$

Lo cual en Python se traduce en:

```
from scipy.stats import norm

norm(mu, sigma).pdf(X)
```

Así también, la distribucion de probabilidad acumulada correspondiente está dada por:

$$C(X)=\int_{x<¿X} P(x)\,dx=\int_{-\infty}^{X} P(X)\,dX$$

teniendo en cuenta que $Dom(X)=(-\infty,\infty)$.

`Ejemplo`: considerando una variable aleatoria que sigue una distribucion normal con media $\mu=4$ y desviación estándar $\sigma=0.3$, la probabilidad de que dicha variable tome el valor de 0.2 o menos está dada por:

$$C(0.2)=\int_{x<¿0.2} P(x)\,dx=\int_{-\infty}^{0.2}\left(\frac{1}{0.3\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{X-4}{0.3}\right)^2\right)\right)dX$$

La cual se calcula en Python como:

```
from scipy.stats import norm

norm(mu, sigma).cdf(X)
```

Es importante recordar que la función de probabilidad acumulada de la distribución gaussiana no se puede calcular de forma exacta, ya que la integral anterior no tiene una expresión cerrada conocida. Es decir, los métodos de integración conocidos no funcionan para resolver esta integral.

6.1.1. Ejercicios (bloque 1)

Considerando una variable aleatoria que sigue una distribución normal con media $\mu=4$ y desviación estándar $\sigma=0.1$, calcula las siguientes probabilidades (usando Python):

1. $P(X=4)=0$
2. $P(X=-10)=0$
3. $P(X=10)=0$
4. $P(X<¿4)=0.5$
5. $P(X>¿4)=0.5$

```
from scipy.stats import norm

print(norm(4,0.1).cdf(4))
```

6.2. De la binomial a la gaussiana

De la binomial a la Gaussiana

En la clase 8 vimos como generar secuencias aleatorias de experimentos binomiales (aquí), donde cada experimento era lanzar un cierto número de monedas.

¿Qué sucede si el número consecutivo de monedas que lanzamos en cada experimento (trial) es muy largo?

La función `generate_binomial_trials()` nos muestra lo que sucede si graficamos los resultados de muchos experimentos de lanzar 100 monedas en cada intento, con pyplot:

```
import numpy as np
from numpy.random import binomial
import matplotlib.pyplot as plt

def generate_binomial_trials(trials=1000, coin_toss=100):
  '''
  el resultado de esta funcion es generar un conjuntos
  de experimentos binomiales (trials) y de cada uno obtener
  las cantidades de exitos en cada secuencia (e.j. lanzar monedas).

  * trial: es una secuencia de <coin_toss> lanzamientos de moneda

  * coin_toss: es el numero de monedas lanzadas en cada trial
  '''
  arr = []
  for _ in range(trials):
    arr.append(binomial(coin_toss, 0.5))
  values, dist = np.unique(arr, return_counts=True)

  return values, dist, np.array(arr)

values, dist, arr = generate_binomial_trials(100000)
plt.bar(values, dist)
```

Ejemplo

```
# Sabiendo que la distribución es binomial, podemos calcular la media
y desviación estandar de la siguiente manera
# mu = número de intentos (coin_toss) * p
# sigma = math.sqrt(coin_toss*p*(1-p))
# Fuente:
https://stats.libretexts.org/Bookshelves/Introductory_Statistics/Book
%3A_Statistics_Using_Technology_(Kozak)/
05%3A_Discrete_Probability_Distributions/
5.03%3A_Mean_and_Standard_Deviation_of_Binomial_Distribution

mu = 100 * 0.5
sigma = (100*0.5**2)**0.5
normal = norm(mu, sigma)

X = np.arange(30,70,0.1)
Y = [normal.pdf(x) for x in X]

plt.plot(X, Y)
plt.bar(values, dist/100

# Otra solución, en caso de que no tengamos la certeza de que la
distribución es efectivamente binomial
# Para esto toca modificar la función original para que retorne el los
datos completos

mu = arr.mean()
sigma = arr.std()
normal = norm(mu, sigma)

X = np.arange(30,70,0.1)
Y = [normal.pdf(x) for x in X]

plt.plot(X, Y)
plt.bar(values, dist/len(arr))
```

Si todo salió bien, habrás notado que en efecto una distribución normal se ajusta perfectamente a los datos. Esto se conoce como el **teorema del límite central**, el cual establece que en muchas situaciones conocidas, cuando variables aleatorias independientes se combinan, su total tiende a seguir una distribución normal cuando el número de variables que se combinan es muy grande $n \to \infty$.

6.3. Variable Poisson (Distribución de Poisson)

Variable Poisson (Distribución de Poisson)

Esta distribución es utilizada en problemas de conteo. Por ejemplo el número de bacterias por unidad de área o volumen encontradas en un cultivo micro biológico.

Una variable Poisson puede tomar teóricamente valores enteros entre cero e infinito. La función de probabilidad en este caso está dada por

$$p_{poi}(k) = \frac{e^{\lambda}\lambda^{k}}{k!}, k = 0, 1, \ldots$$

El valor λ se interpreta como la cantidad promedio de elementos encontrados por unidad de área, volumen, etc.

7. Variables Categóricas

Variables Categóricas

En esta sección vamos a suponer que la variable aleatoria no asigna valores numéricos a los elementos del espacio muestral. En lugar de eso, asigna etiquetas. Usualmente dice que en este caso la variable es en realidad un objeto aleatorio. No haremos esa distinción en este desarrollo, pero si seremos cuidadosos en tener en cuenta que la variable no es numérica.

Por facilidad usaremos de nuevo el ejemplo de los dados no cargados. Definimos la siguiente variable aleatoria que llamaremos X. Para este ejercicio vamos a distinguir el color de los dados. Entonces en la pareja (x, y), x corresponde al valor del dado rojo. adicionalmente y representa el valor del dado azul. Entonces

$$X((x, y)) = \begin{pmatrix} R, \text{si } x > y \\ A, \text{si } x < y \\ B, \text{si } x = y \end{pmatrix}$$

Observe que en realidad el color del dado es indiferente para la definición de la variable aleatoria

Entonces, la función de probabilidad en este caso es dada por

$$\begin{aligned} p_X(R) & ¿15/36 \\ p_X(A) & ¿15/36 \\ p_X(B) & ¿6/36 \end{aligned}$$

Aún podemos hacer un gráfico de la función de probabilidad p_X como sigue.

```
import numpy as np
import matplotlib.pyplot as plt
```

```
p_X = np.array([15/36,15/36, 6/36])
label = ['R','A','B']

plt.bar(label,p_X, color = 'red', edgecolor='black',width=1)
plt.title('Funcion de probabilidad  $p_X$')
plt.xlabel('Etiquetas de  $p_X$')
plt.ylabel('Probabilidad')
plt.show()
```

Obviamente en este caso no hay cantidades numéricas, por lo que el gráfico puede obtenerse de distintas formas, cambiando el orden de la etiquetas. Sin embargo, aún podemos hacer algunos cálculos de probabilidad.

Por ejemplo, la probabilidad de obtener R o A es dada por $30/36$.

8. Esperanza Matemática de una variable aleatoria numérica

Esperanza Matemática de una variable aleatoria numérica

La esperanza matemática de una variable aleatoria numérica discreta es el promedio ponderado de todos sus posibles valores. La ponderación como se puede sospechar es dada por la función de probabilidad de la variable.

La definición se basa en el hecho que la función de probabilidad de la variable aleatoria representa la frecuencia relativa teórica de cada resultado particular.

Se tiene que si $X=\{x_1,x_2,\ldots\}$ es el conjunto de valores de la variable aleatoria X y si la función de probabilidad de X es dada por $p_X=\{p_1,p_2,\ldots\}$, en donde $p_X[x_i]=p_i$, entonces la esperanza matemática de X se denota por $E[X]$ y es dada por

$$E[X]=p_1x_1+p_2x_2+\ldots=\sum_i p_ix_i$$

En el ejemplo de los dados se tiene que

$$E[f]=1/36[2\times1+3\times2+4\times3+5\times4+6\times5+7\times6+8\times5+9\times4+10\times3+11\times2+12\times1]=7.$$

Puede tambien calcular la esperanza de f con numpy como sigue.

```
np.mean(f)
```

8.1. Esperanza de la distribución Binomial

Esperanza de la distribución Binomial

En el caso del ejemplo Binomial de la moneda cargada que se lanza tres veces tenemos que

$$E[f]=0\times 0.064+1\times 0.288+2\times 0.432+3\times 0.216=1.8.$$

Se puede verificar que si una variable aleatoria X tiene distribución (función de probabilidad) Binomial, $\mathrm{Bin}(N,p)$, entonces

$$E[X]=N\,p.$$

Puede consultar cualquier texto de probabilidad para verificar el resultado. El siguiente código muestra como calcular la esperanza para el caso $\mathrm{Bin}(3,0.6)$, con Python.

```
from scipy.stats import binom
binom.expect(args=(3,0.6))
```

8.2. Esperanza de la distribución Poisson

Esperanza de la distribución Poisson

Una variable aleatoria con distribución $\mathrm{Pois}(\lambda)$ tiene esperanza matemática λ.

8.3. Varianza y desviación estándar de una variable aleatoria

Varianza y desviación estándar de una variable aleatoria

La varianza de una variable aleatoria mide su nivel de predictibilidad. Dicho en otras palabras, que tan dispersos son los valores de la variable aleatoria en relación con la esperanza matemática. Denotemos la esperanza de X por μ_X. Técnicamente la varianza de una variable aleatoria se define por

$$Var[X]=E[X-\mu_X]^2=\sum_i p_i(x_i-\mu_X)^2.$$

8.4. Desviación estándar

Desviación estándar

La desviación estándar de una variable aleatoria se denota σ_X y es simplemente la raíz cuadrada de la varianza, es decir,

$$\sigma_X = \sqrt{Var[X]}.$$

Para entender el porque la varianza mide el nivel de predictibilidad de una variable aleatoria consideremos una variable Bernoulli X. Escribimos $X \sim \mathrm{Ber}(\pi)$. En este caso puede verificarse que la varianza es dada por

$$Var(X) = \pi(1-\pi).$$

El siguiente gráfico muestra la varianza de las distribuciones Bernoulli con valorea distintos del parámetro π.

```
import numpy as np
import matplotlib.pyplot as plt

pi = np.linspace(0,1,100)
var = pi*(1-pi)
plt.plot(pi,var)
plt.title('Funcion de varianza de distribciones Bernoulli')
plt.xlabel( '$\pi$')
plt.ylabel('Var')
plt.show()
```

Se observa entonces que la máxima varianza se alcanza para el caso $\pi=0.5$. Es porque en este caso, es mas difícil predecir el resultado del experimento, dado que la probabilidad de *acierto* y *fallo* son la misma. Pero a medida que el valor π está cerca de cero o uno, la varianza desciende hasta cero. Por ejemplo si $\pi=0.9$, casi siempre se obtendrá *éxito*. Esto significa que la variable es más predictible para valores muy altos o muy bajos de π. **A menor varianza mayor precisión para predecir el resultado del experimento y viceversa**.

9. Entropía

9.1. Entropía de una variable aleatoria

Entropía de una variable aleatoria

Similar a la varianza, la entropía de una variable aleatoria mide su grado de predictibilidad desde el punto de vista de la teoría de información de Shanon. Dada una variable aleatoria discreta $X = \{x_1, x_2, \ldots\}$ con función de probabilidad $p_X = \{p_1, p_2, \ldots\}$, la entropía de X se define por

$$H(X) = -\sum_i p_i \log p_i = -E[\log P_X]$$

Es usual utilizar la base 2 o la base de los logaritmos Neperianos. En el primer caso, la unidad de medida de la entropía de denomina bit. En el caso de los logaritmos Neperianos, la unidad de medida se acostumbra a llamar nat. En realidad pasar de una base a otra es simplemente un cambio de escala. Estructuralmente miden lo mismo en distintas unidades.

Como en el caso de la varianza ilustramos el concepto para la distribución Bernoulli.

```
import numpy as np
import matplotlib.pyplot as plt

pi = np.linspace(0.0000001,0.999999,100)
H = -(pi*np.log(pi) + (1-pi)*np.log(1-pi))
plt.plot(pi,H)
plt.title('Entropía de distribuciones Bernoulli')
plt.xlabel( '$\pi$')
plt.ylabel('Entropía')
plt.show()
```

Como se observa la entropía mide de la misma forma que la varianza pero en una escala diferente. Por otro lado observe que en el cálculo de la entropía no se utilizan los valores que toma la variable aleatoria. Solamente se requiere la función de probabilidad de la variable.

En palabras se dice que **la entropía no depende de la escala de la variable**. Solamente de su estructura de probabilidad. Por su parte la varianza si es dependiente de la escala de la variable aleatoria.

Aunque el contenido de información proporciona información interesante sobre eventos específicos $ x $, en algunos casos nos gustaría conocer el contenido de información de una distribución de probabilidad. En este asunto, lo más razonable sería estimar la información esperada. Esta estimación se conoce como entropía.

9.2. Entropía de Shannon

Entropía de Shannon

Supongamos que X es una variable aleatoria discreta. Le entropía (de Shanon) de la variable aleatoria X, o lo que es lo mismo, la entropía de la distribución asociada a X se define por

$$H(X)=\sum_{x\in\Omega} P(X=x)I(x)=-\sum_{x\in\Omega} P(X=x)\log P(X=x)=-\sum_i p_i\log p_i.$$

1. En ocasiones, la entropía de Shanon para una variable aleatoria $ X $ con probabilidades $ p_1, \ldots, p_M $ se denota $ H (p_1, \ldots, p_M) $. Por ejemplo, la entropía de una distribución de Bernoulli a veces se denota $H(p,q)=H(p,1-p)$.
2. Algunos autores llaman a $H(X)$ como *incertidumbre*.
3. $H(X)$ Es una esperanza. La entropía es solo la **sorpresa media** de la variable aleatoria $H(X)$.

Créditos

Docente: Andrés Felipe Florez O.

Universidad Pedagógica y Tecnológica de Colombia - *Diplomado en Data Science - Cohorte I - 2024*
