Actividad sesión 02

Juan Diego Gómez Chavarro

22/02/2024

1. Probabilidad frecuentista

- Suponga que tienen dos dados balanceados de seis caras cada uno. ¿La probabilidad de obtener un uno en ambos dados será la misma si se lanzan simultáneamente o si se lanzan de manera consecutiva?
- En una baraja de 52 cartas, ¿cuál es la probabilidad de obtener un póker de ases en la primera mano?
- En una baraja de 52 cartas, ¿cuál es la probabilidad de recibir tres cartas cuya suma sea 21?

1.1. Solución item 1

La probabilidad $P(A \cap B)$ de obtener un uno en el primer dado (evento A) y obtener un uno en un segundo dado (evento B) no se ve afectada por el orden de lanzamiento de los dados, ya que los eventos A y B son independientes entre sí.

Por lo tanto, $P(A|B)$ y $P(B|A)$ son iguales a las probabilidades individuales de cada evento, es decir, $P(A)$ y $P(B)$. En este orden de ideas:

$$P(A \cap B) = P(A) \cdot P(B) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$$

1.2. Solución ítem 2

Una mano de póker estándar consiste en 5 cartas extraídas de una baraja de 52 cartas. El número total de formas en que se pueden seleccionar 5 cartas de 52 (sin importar el orden) es:

$$\text{Total de manos posibles} = \binom{52}{5} = \frac{52!}{5!(52-5)!} = 2,598,960 \tag{1}$$

Para obtener un póker de ases, necesitamos exactamente 4 ases y una carta adicional cualquiera.

- Hay 4 ases en la baraja y debemos elegir todos ellos:

$$\binom{4}{4} = 1 \tag{2}$$

- La quinta carta puede ser cualquiera de las 48 cartas restantes (las que no son ases):

$$\binom{48}{1} = 48 \tag{3}$$

Por lo tanto, el número de manos que contienen un póker de ases es:

$$\binom{4}{4} \times \binom{48}{1} = 1 \times 48 = 48 \tag{4}$$

La probabilidad de obtener un póker de ases es el número de casos favorables dividido entre el total de casos posibles:

$$P(\text{póker de ases}) = \frac{48}{2,598,960} \tag{5}$$

Calculamos el valor numérico:

$$P(\text{póker de ases}) \approx 0{,}0000185 \tag{6}$$

Es decir, aproximadamente un 0.00185 % (menos de 2 manos en 100,000).

1.3. Solución item 3

Se muestra a continuación la demostración para calcular la probabilidad de recibir tres cartas cuya suma sea 21 en una baraja de 52 cartas.

El número de formas de escoger 3 cartas (sin importar el orden) es:

$$\binom{52}{3} = \frac{52!}{3!(52-3)!} = 22100.$$

Ahora se busca una tripleta (sin importar el orden) de tres cartas que al ser sumadas den 21:

$$\begin{aligned}
&(10, 10, 1) \quad (10 + 10 + 1 = 21)\\
&(10, 9, 2) \quad (10 + 9 + 2 = 21)\\
&(10, 8, 3) \quad (10 + 8 + 3 = 21)\\
&(10, 7, 4) \quad (10 + 7 + 4 = 21)\\
&(10, 6, 5) \quad (10 + 6 + 5 = 21)\\
&(9, 9, 3) \quad (9 + 9 + 3 = 21)\\
&(9, 8, 4) \quad (9 + 8 + 4 = 21)\\
&(9, 7, 5) \quad (9 + 7 + 5 = 21)\\
&(9, 6, 6) \quad (9 + 6 + 6 = 21)\\
&(8, 8, 5) \quad (8 + 8 + 5 = 21)\\
&(8, 7, 6) \quad (8 + 7 + 6 = 21)\\
&(7, 7, 7) \quad (7 + 7 + 7 = 21)
\end{aligned}$$

Se procede a calcular el numero de formas para cada combinación. teniendo en cuenta que hay cuatro cartas de cada una.

1. $(10, 10, 1)$**:**

   - Elegir 2 cartas de valor 10: $\binom{16}{2} = \frac{16!}{2!(16-2)!} = 120$ formas.
   - Elegir 1 as: $\binom{4}{1} = \frac{4!}{1!(4-1)!} = 4$ formas.

   $$\text{Total: } 120 \times 4 = 480.$$

2. $(10, 9, 2)$**:**

   $$\text{Total: } 16 \times 4 \times 4 = 256.$$

3. $(10, 8, 3)$**:**

   $$\text{Total: } 16 \times 4 \times 4 = 256.$$

4. $(10, 7, 4)$**:**

   $$\text{Total: } 16 \times 4 \times 4 = 256.$$

5. $(10, 6, 5)$**:**

   $$\text{Total: } 16 \times 4 \times 4 = 256.$$

6. $(9, 9, 3)$**:**

   - Elegir 2 cartas de 9: $\binom{4}{2} = 6$ formas.
   - Elegir 1 carta de 3: 4 formas.

   $$\text{Total: } 6 \times 4 = 24.$$

7. $(9, 8, 4)$**:**

   $$\text{Total: } 4 \times 4 \times 4 = 64.$$

8. $(9, 7, 5)$**:**

   $$\text{Total: } 4 \times 4 \times 4 = 64.$$

9. $(9, 6, 6)$**:**

   - 9: 4 formas.
   - Elegir 2 cartas de 6: $\binom{4}{2} = 6$ formas.

   $$\text{Total: } 4 \times 6 = 24.$$

10. $(8, 8, 5)$**:**

    - Elegir 2 cartas de 8: $\binom{4}{2} = 6$ formas.
    - Elegir 1 carta de 5: 4 formas.

    $$\text{Total: } 6 \times 4 = 24.$$

11. $(8,7,6)$:
$$4 \times 4 \times 4 = 64.$$

12. $(7,7,7)$:
$$\binom{4}{3} = 4.$$

En este orden de ideas, el numero total de formas de obtener estas combinaciones son:

$$480 + 256 + 256 + 256 + 256 + 24 + 64 + 64 + 24 + 24 + 64 + 4 = 1772 \text{ Formas.}$$

La probabilidad P de obtener tres cartas cuya suma sea 21 es:

$$P = \frac{\text{Número de casos favorables}}{\text{Número total de casos}} = \frac{1772}{22100}.$$

en conclusion la probabilidad de obtener tres cartas cuya suma sea 21 es:

$$P \approx 0{,}0802 \quad (8{,}02\,\%).$$

2. Solución enunciado 2

- La probabilidad de tener una enfermedad es del 1
- Si una persona TIENE la enfermedad, el test lo detecta el 95 % de las veces.
- Si una persona esta sana, el test podría detectar la enfermedad un 5 % de las veces.
- Un paciente tomó el test y le dio positivo. ¿Cuál es la probabilidad de que el paciente SI TENGA la enfermedad?

Definimos las siguientes probabilidades:

$$P(\text{enfermo}) = 0{,}01$$

$$P(\text{sano}) = 1 - P(\text{enfermo}) = 0{,}99$$

$$P(\text{Test positivo} \mid \text{enfermo}) = 0{,}95$$

$$P(\text{Test positivo} \mid \text{sano}) = 0{,}05$$

Queremos calcular:

$$P(\text{enfermo} \mid \text{Test positivo}) = \frac{P(\text{Test positivo} \mid \text{enfermo}) \cdot P(\text{enfermo})}{P(\text{Test positivo})}$$

Como no conocemos $P(\text{Test positivo})$, lo calculamos utilizando la ley de probabilidad total:

$$P(\text{Test positivo}) = P(\text{enfermo}) \cdot P(\text{Test positivo} \mid \text{enfermo}) + P(\text{sano}) \cdot P(\text{Test positivo} \mid \text{sano})$$

Sustituyendo los valores:

$$P(\text{Test positivo}) = (0{,}01 \times 0{,}95) + (0{,}99 \times 0{,}05)$$

$$P(\text{Test positivo}) = 0{,}0095 + 0{,}0495 = 0{,}059$$

Ahora aplicamos el teorema de Bayes:

$$P(\text{enfermo} \mid \text{Test positivo}) = \frac{P(\text{Test positivo} \mid \text{enfermo}) \cdot P(\text{enfermo})}{P(\text{Test positivo})}$$

$$P(\text{enfermo} \mid \text{Test positivo}) = \frac{0{,}95 \times 0{,}01}{0{,}059}$$

$$P(\text{enfermo} \mid \text{Test positivo}) = \frac{0{,}0095}{0{,}059} = 0{,}161$$

3. Solución enunciado 3

- Obtenga un muestra de 100 datos de una distribución normal con media CERO y varianza UNO. Para ello implemente la transformación de Box-Muller.
- Grafique el histograma de las muestras para verificar visualmente que son gausianas.
- Aplique el criterio de Kolmogorov-Smirnov a las muestras obtenidas para compararlas primero con una gausiana de referencia, y luego con una exponencial.
- Repita el proceso con una muestra de tan solo 10 datos. Concluya lo observado

3.1. Solución

Para obtener una muestra de 100 y 10 datos de una distribución normal con media cero y varianza uno, se utilizó la transformación de Box-Muller. Esta técnica genera pares de variables aleatorias independientes con distribución normal estándar a partir de dos variables aleatorias independientes con distribución uniforme en el intervalo $[0, 1]$.

```python
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import norm, expon, kstest

def boxMuller(n):
    U1 = np.random.uniform(0, 1, size=n)
    U2 = np.random.uniform(0, 1, size=n)
    Z0 = np.sqrt(-2 * np.log(U1)) * np.cos(2 * np.pi * U2)
    Z1 = np.sqrt(-2 * np.log(U1)) * np.sin(2 * np.pi * U2)
    return Z0, Z1

Z0, Z1 = boxMuller(500)
Z_100 = np.concatenate([Z0, Z1])

Z0, Z1 = boxMuller(5)
Z_10 = np.concatenate([Z0, Z1])
```

Figura 1: Muestras utilizando Box-muller

Para verificar la forma de la distribución de las muestras generadas, se construyeron histogramas para las muestras de 100 y 10 datos, respectivamente. Se observa que la distribución de 100 muestras tiene una forma cercana a la gaussiana, mientras que la de 10 muestras presenta mayor variabilidad y menor semejanza a la distribución esperada.

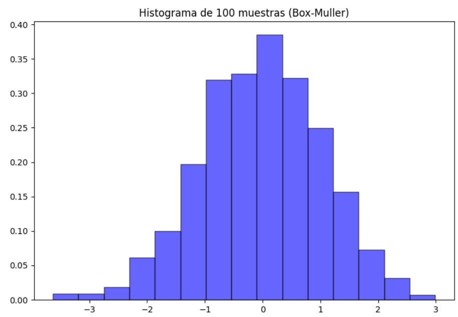

Figura 2: Histograma de 100 muestras.

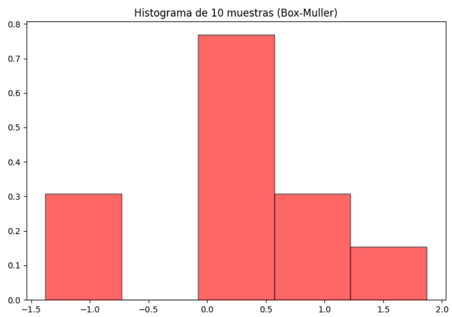

Figura 3: Histograma de 10 muestras.

Para evaluar cuantitativamente si las muestras generadas provienen de una distribución normal, se aplic'o la prueba de Kolmogorov-Smirnov. Se compararon las muestras tanto con una distribución normal de referencia como con una distribuci'on exponencial para verificar cuán bien se ajustan.

```python
def kolmogorov_smirnov_test(Z, cdf):
    Z = np.sort(Z)
    n = len(Z)
    # Se calcula la función teórica en cada punto
    cdf_teorica = np.array([cdf(x) for x in Z])
    # Límite derecho: valor de la función empírica justo después del salto
    emp_der = np.arange(1, n + 1) / n
    # Límite izquierdo: valor de la función empírica justo antes del salto
    emp_izq = np.arange(0, n) / n
    d_der = np.abs(emp_der - cdf_teorica)
    d_izq = np.abs(cdf_teorica - emp_izq)
    d_n = np.max(np.concatenate([d_der, d_izq]))
    return d_n

d_n_100_norm = kolmogorov_smirnov_test(Z_100, norm.cdf)
d_n_100_expon = kolmogorov_smirnov_test(Z_100, expon.cdf)

d_n_10_norm = kolmogorov_smirnov_test(Z_10, norm.cdf)
d_n_10_expon = kolmogorov_smirnov_test(Z_10, expon.cdf)

print("RESULTADOS")
print("D_n_100_norm  : ", d_n_100_norm)
print("D_n_100_expon : ", d_n_100_expon)
print("D_n_10_norm   : ", d_n_10_norm)
print("D_n_10_expon  : ",d_n_10_expon)
```

Figura 4: cálculo de Kolmogorov-Smirnov

Los resultados obtenidos se presentan a continuación:

```
RESULTADOS
D_n_100_norm  :  0.023264122990833636
D_n_100_expon :  0.489
D_n_10_norm   :  0.4117003301047814
D_n_10_expon  :  0.2740400959013576
```

Figura 5: Resultados KS

Para la muestra de 100 datos, los resultados de la prueba de Kolmogorov-Smirnov muestran que los valores generados se ajustan bien a una distribución normal, mientras que para la muestra de 10 datos la diferencia es mayor. Esto se debe a que, con menos datos, hay más variabilidad y es más difícil determinar con precisión la distribución real. Esto muestra que usar muestras más grandes ayuda a obtener resultados más confiables en este tipo de pruebas.