

Moisés Cantón Jara

Fundamentos de Matemáticas

Probabilidad y Estadística en IA

Probabilidad y Estadística en IA

1. Introducción

La Probabilidad y la Estadística son dos pilares fundamentales en el mundo de la inteligencia artificial y el aprendizaje automático. En este tema nos sumergiremos en el fascinante mundo de la incertidumbre y la toma de decisiones basada en datos. La probabilidad, como herramienta para medir y comprender la aleatoriedad, y la estadística, como el arte de extraer conocimiento significativo de los datos, son esenciales para modelar, predecir y tomar decisiones en situaciones complejas y variables. A lo largo de este tema, exploraremos los conceptos clave de probabilidad y estadística, su aplicación en la inteligencia artificial y cómo estas disciplinas se convierten en cimientos sólidos para abordar problemas del mundo real en el ámbito de la IA.

2. Conceptos Fundamentalesⁱ

¿Qué es la incertidumbre?

La incertidumbre es una característica esencial en muchas situaciones que involucran la probabilidad. En nuestro contexto, esta se refiere a la falta de certeza o predictibilidad completa en los resultados de un evento o experimento y es una consecuencia natural de la variabilidad inherente en el mundo real.

Aquí algunas claves para entender la incertidumbre en relación con la probabilidad:

- **Aleatoriedad.** Muchos eventos y fenómenos son inherentemente aleatorios, lo que implica que sus resultados no pueden preverse con una certeza absoluta. La probabilidad es la rama de las matemáticas que se ocupa de cuantificar esta aleatoriedad y proporcionar una medida de la incertidumbre asociada.
- **Modelado de la incertidumbre.** A través de las conocidas como distribuciones de probabilidad y de los cálculos probabilísticos, se puede modelizar cómo se distribuyen los posibles resultados y asignar una probabilidad a cada uno de ellos.
- **Toma de decisiones bajo incertidumbre.** A la hora de tomar decisiones en situaciones de incertidumbre, la probabilidad se convierte en una guía, ya que nos permite tomar estas maximizando las oportunidades y minimizando los riesgos asociados.
- **Estadística e Inferencia.** En estadística, la incertidumbre se aborda mediante el uso de estimaciones e intervalos de confianza. Los datos muestrales se utilizan para hacer inferencias sobre poblaciones más grandes, y la probabilidad se utiliza para evaluar la confiabilidad de estas inferencias.

¿Qué es la precisión?

La precisión es un concepto crítico en el campo de la probabilidad y la estadística, ya que se relaciona directamente con la calidad y confiabilidad de los análisis y las predicciones. Sin embargo, en la práctica, existen desafíos que pueden afectar a la precisión de los resultados. A continuación, se destacan algunos:

- **Datos Ruidosos.** Al recopilar datos en el mundo real, estos pueden contener variaciones aleatorias que pueden distorsionar la señal que tratamos de medir. La presencia de estas variaciones puede disminuir la precisión y dificultar la detección de patrones reales en los datos.

- Cobertura Limitada del Área del problema. En algunos casos, los datos disponibles pueden representar una parte limitada o sesgada del fenómeno a estudiar y conllevar con ello a estimaciones sesgadas y reducir la precisión de las conclusiones.
- Modelos Imperfectos. En estadística, a menudo se emplean modelos matemáticos para representar fenómenos complejos. Estos modelos son simplificaciones de la realidad y, por tanto, no capturan todos los aspectos fundamentales de un problema.

A la hora de abordar estos desafíos, se emplean diversas estrategias que nos permiten mejorar la precisión:

- Limpieza de datos. A la hora de combatir datos ruidosos, se suele realizar una limpieza y preprocesado de datos de manera cuidadosa, donde se eliminan valores atípicos y se interpolan datos faltantes.
- Muestreo Representativo. Para superar la cobertura limitada del área del problema, es crucial realizar un muestreo representativo.
- Mejora de Modelos. Los modelos estadísticos se pueden mejorar continuamente a medida que se obtienen más y más datos y se refina la comprensión del problema.

Además, se debe validar la precisión de los resultados y compararlos con las predicciones, a través de medidas de calidad como el error cuadrático medio R^2 .

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

donde

- y_i es el valor observado para la i -ésima observación
- \hat{y} es el valor predicho para la i -ésima observación
- \bar{y} es la media del valor observado para la i -ésima observación
- n es el número de observaciones

¿Qué es la expectativa?

La expectativa o valor esperado es otro concepto fundamental en la teoría de la probabilidad, pues representa el valor promedio o medio que se espera de una variable aleatoria en un conjunto de posibles resultados. Normalmente, se denota como $E[X]$, donde X es la variable aleatoria de interés.

La fórmula general para calcular la expectativa de una variable aleatoria discreta X es:

$$E[X] = \sum_x x * P(X = x) = x_1 p_1 + x_2 p_2 + \dots + x_n p_n$$

donde x son los posibles valores de X y $P(X=x)$ es la probabilidad de que X tome el valor x .

Tomando esto, tendríamos, por ejemplo, para el lanzamiento de un dado que el valor promedio o esperado para la puntuación sería.

$$E[X] = \frac{1}{6} (1 + 2 + 3 + 4 + 5 + 6) = 3.5$$

En cambio, para el caso de una variable aleatoria continua, la expectativa se calcularía a través de una integral en lugar de una suma.

¿Qué es la varianza? ¿y la covarianza?

Varianza y covarianza son dos conceptos clave para cuantificar la dispersión y las relaciones entre variables aleatorias o conjuntos de datos. A continuación, se explica cada uno de ellos:

La varianza es una medida de la dispersión o variabilidad de una variable aleatoria. Indica cuán dispersos están los valores de la variable en torno a su valor esperado (media) y se calcula a través de la siguiente fórmula:

$$Var(X) = \mathbb{E}[(X - \mathbb{E}(X))^2]$$

donde $\mathbb{E}[X]$ es el valor esperado de X . Esta fórmula representa la media de los cuadrados de las diferencias entre cada valor de X y su valor esperado.

Nota: La raíz cuadrada positiva de la varianza se conoce como la desviación estándar (σ en poblaciones o s en muestras). La desviación estándar es una medida más interpretable que tiene la misma unidad que la variable aleatoria.

Mientras, la covarianza es una medida que cuantifica la relación lineal entre dos variables aleatorias e indica si las dos variables tienden a aumentar o disminuir juntas. Una covarianza positiva indica una relación directa, mientras que una covarianza negativa indica una relación inversa, y se calcula tal que:

$$Cov(X, Y) = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))]$$

donde $\mathbb{E}[X]$ y $\mathbb{E}[Y]$ son los valores esperados de X e Y .

3. Reglas y Teoremas Esenciales de Probabilidad

Regla Suma

Si tenemos dos eventos A y B que son mutuamente excluyentes (es decir, no pueden ocurrir al mismo tiempo), la probabilidad de que ocurra A o B es la suma de sus probabilidades individuales.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

donde

- $P(A)$ es la probabilidad de que ocurra A
- $P(B)$ es la probabilidad de que ocurra B
- $P(A \cap B)$ es la probabilidad de que ocurra A y B

Regla Producto

Para dos eventos independientes A y B (es decir, la ocurrencia de uno no afecta la probabilidad del otro), la probabilidad de que ambos eventos ocurran es el producto de sus probabilidades.

$$P(A \cap B) = P(A) \times P(B)$$

Teorema de Bayesⁱⁱ

Esencial para muchos aspectos del aprendizaje automático, el Teorema de Bayes relaciona la probabilidad posterior de un evento A dado otro evento B con la probabilidad prior de A , la probabilidad de B dado A , y la probabilidad marginal de B .

$$P(A | B) = \frac{P(B | A) \times P(A)}{P(B)}$$

donde:

- $P(A|B)$ es la probabilidad posterior de A dado B.
- $P(B|A)$ es la probabilidad de B dado A.
- $P(A)$ y $P(B)$ son las probabilidades marginales de A y B respectivamente.

Aplicación: Teorema de Bayes

Supongamos que sabemos lo siguiente sobre una población:

1. La probabilidad de que una persona tenga cáncer (sin tener en cuenta si fuma o no) es del 10%, es decir, $P(\text{Cancer}) = 0.10$
2. El 30% de la población es fumadora, es decir, $P(\text{Fumador}) = 0.30$
3. De las personas que tienen cáncer, el 50% son fumadoras, es decir, $P(\text{Fumador} | \text{Cancer}) = 0.50$

Queremos encontrar la probabilidad de que una persona tenga cáncer dado que es fumadora, es decir, $P(\text{Cancer} | \text{Fumador})$. En nuestro caso, si A representa "tener cáncer" y B representa "ser fumador". Sustituyendo en la expresión del teorema de Bayes nuestros datos, obtenemos:

$$P(\text{Cancer} | \text{Fumador}) = \frac{P(\text{Fumador} | \text{Cancer}) * P(\text{Cancer})}{P(\text{Fumador})} = \frac{0.50 * 0.10}{0.30} = 0.167$$

por lo tanto, la probabilidad de que una persona tenga cáncer dado que es fumadora es del 16.7%, lo que significa que, si seleccionamos al azar a una persona fumadora de la población, hay un 16.7% de probabilidad de que tenga cáncer.

4. Métricas de Evaluación para Clasificadores

Precisión (o Accuracy, en inglés)

Tal y como comentamos en la introducción, la precisión mide la proporción de predicciones correctas sobre el total de predicciones realizadas por el modelo para una casuística donde todas las clases tienen la misma importancia y están balanceadas en el conjunto de datos.

Si estamos analizando un fenómeno, donde solo se puede obtener dos resultados o clases posibles, la precisión, se calcularía tal que:

$$Precision = \frac{TP + TN}{TP + FP + TN + FN}$$

donde TP y TN son los verdaderos positivos y negativos, mientras que FP y FN solo los falsos positivos y negativos.

Sensibilidad (o Recall y True Positive Rate, en inglés)

La sensibilidad nos indica la proporción de verdaderos positivos, respecto a todos los resultados positivos. Se calcularía como sigue y es útil cuando es importante detectar todos los resultados positivos.

$$\text{Sensibilidad} = \frac{TP}{TP + FN}$$

Especificidad (o False Negative Rate, en inglés)

La especificidad nos indica la proporción de verdaderos negativos, respecto a todos los resultados positivos. Se calcularía como sigue y es útil cuando es importante detectar todos los falsos negativos.

$$\text{Especificidad} = \frac{TN}{TN + FP}$$

Precisión Positiva (o Positive Predictive Value, en inglés)

La precisión positiva mide la proporción de verdaderos positivos respecto a todas las predicciones positivas, lo cual es útil cuando se quiere evaluar la precisión de las predicciones positivas.

$$\text{Positive Predictive Value} = \frac{TP}{TP + FP}$$

Precisión Negativa (o Negative Predictive Value, en inglés)

$$\text{Negative Predictive Value} = \frac{TN}{TN + FN}$$

Matriz de Confusión

La matriz de confusión muestra de manera visual detallada los verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos de un modelo.

Si suponemos por simplicidad como hemos ido haciendo, en los ejemplos anteriores, tendremos la siguiente matriz:

$$\begin{array}{cc} & \text{Realidad} \\ & \begin{array}{cc} + & - \end{array} \\ \text{Predicción} & \begin{array}{cc} + & - \\ - & \end{array} \begin{bmatrix} TP & FP \\ FN & TN \end{bmatrix} \end{array}$$

F1 Score

El F1 Score es la media armónica de la precisión y la sensibilidad, lo cual es útil cuando se busca un compromiso entre estas dos métricas. Su rango oscila entre 0 – 100%, donde cuánto más alto sea el valor mejor será la calidad del modelo.

Su fórmula sigue como:

$$F1 \text{ Score} = 2 * \frac{(\text{Precision} * \text{Recall})}{\text{Precision} + \text{Recall}}$$

En un conjunto binario, el cálculo es simple, sustituyendo cada uno de los valores de precisión y sensibilidad por aquellos vistos anteriormente. Sin embargo, cuando tenemos múltiples clases, existen diversas técnicas para reducirlo.

Un ejemplo sería el Macro F1 Score, donde se emplea una técnica one-vs-all para reducir una matriz de confusión mayor a 2x2 a una matriz 2x2 para cada una de las clases, poder calcular su F1 Score y calcular su media.

Esta técnica nos permite reducir, por ejemplo, la matriz inferior izquierda en la matriz inferior derecha:

$$\begin{array}{c}
 \text{Realidad} \\
 \begin{array}{c} 1 \quad 2 \quad 3 \quad 4 \\
 \text{Predicción} \quad \begin{bmatrix} 52 & 3 & 7 & 2 \\ 2 & 28 & 2 & 0 \\ 5 & 2 & 25 & 12 \\ 1 & 1 & 9 & 40 \end{bmatrix} \end{array}
 \end{array} \rightarrow \begin{array}{c}
 \text{Realidad} \\
 \begin{array}{c} + \quad - \\
 \text{Predicción} \quad + \quad \begin{bmatrix} 28 & 2 + 2 + 0 \\ 3 + 2 + 1 & 153 \text{ (Resto)} \end{bmatrix}
 \end{array}
 \end{array}$$

entendiendo que el resultado 1 sería 2 y el resultado 2 sería la suma del resto (1,3 y 4), quedando el Macro F1 Score como:

$$\text{Macro F1 Score} = \frac{\sum_{i=1}^n \text{F1 Score}_i}{n}$$

Sin embargo, esto solo es útil cuando todas las clases están balanceadas y tienen la misma importancia por lo que una siguiente aproximación, la Micro-averaged F1 Score intenta resolver este problema mediante el uso de las TP, FP y FN netas, que disolviendo la matriz de confusión quedaría:

$$\begin{aligned}
 \text{Micro F1 Score} &= \frac{\text{Net TP}}{\text{Net TP} + \frac{1}{2}(\text{Net FP} + \text{Net FN})} \\
 &= \frac{\sum_i M_{ii}}{\sum_{i=1}^n M_{ii} + \frac{1}{2} \left[\left(\sum_{i=1}^n \sum_{j=1, i \neq j}^n M_{ij} \right) + \left(\sum_{i=1}^n \sum_{j=1, i \neq j}^n M_{ji} \right) \right]}
 \end{aligned}$$

y que para un dataset binario sería igual a la Precisión.

$$\text{Micro F1 Score} = \frac{TP + TN}{TP + FP + FN + TN} = \text{Precision}$$

Otro método para obtener un F1 Score más idóneo cuando no tenemos clases balanceadas sería también el Sample-weighted F1 Score, donde se añadiría un peso relacionado con las muestras por clase y se aplicaría este a los F1 Score individuales.

$$\text{Weighted F1 Score} = \sum_{i=1}^n w_i * \text{F1 Score}_i$$

donde $w_i = \frac{\text{Num of sample in class } i}{\text{Total number of samples}}$

Fβ Score

Otra métrica es el Fβ Score, una generalización del F1 Score que nos permite ajustar el equilibrio entre precisión y sensibilidad.

Su fórmula quedaría como:

$$F\beta = (1 + \beta^2) * \frac{Precision * Sensibilidad}{\beta^2 * Precision + Sensibilidad}$$

Esto nos permitiría, por ejemplo, priorizar, la precisión o la sensibilidad. Imaginando, un ejemplo, podríamos tener la detección del COVID-19, donde los falsos negativos podían contribuir negativa y exponencialmente a la expansión del virus, y, en lugar de un F1 Score, podríamos, emplear, por ejemplo, un F2 Score que nos ayudase a minimizar los falsos negativos manteniendo la precisión lo más alta posible.

5. Distribuciones Estadísticasⁱⁱⁱ

¿Qué es una distribución de probabilidad?

Una distribución probabilística una descripción matemática de las posibles ocurrencias de un evento o variable aleatoria y las probabilidades asociadas con cada una de estas ocurrencias. En otras palabras, representa cómo se distribuyen las probabilidades entre los diferentes valores que una variable aleatoria puede tomar y permiten modelar fenómenos matemáticamente.

Según si la variable aleatoria puede tomar valores discretos (como el número de caras en una serie de lanzamientos de una moneda) o valores continuos (como la altura de las personas en una población), puede ser categorizadas en discretas o continuas.

¿Qué tipos de distribuciones de probabilidad podemos encontrar?

A continuación, veremos algunos tipos de distribuciones junto con ejemplos de fenómenos que se pueden modelar con ellas:

A. Distribución Binomial

La distribución binomial es una distribución de probabilidad discreta que modela el número de éxitos (eventos de éxito) en un número fijo de ensayos independientes idénticos, donde cada ensayo tiene dos resultados posibles: éxito o fracaso. A continuación, se presentan algunos conceptos clave y ejemplos:

Conceptos Clave:

- **Parámetros n y p:** La distribución binomial tiene dos parámetros: n y p . n representa el número de ensayos o intentos independientes, y p representa la probabilidad de éxito en un solo ensayo.
- **Función de Probabilidad:** La función de probabilidad de la distribución binomial se define de la siguiente manera:

$$P(X = x) = \binom{n}{x} * p^x * (1 - p)^{n-x}$$

donde:

- $\binom{n}{x}$ representa el coeficiente binomial, que mide el número de formas de elegir x éxitos de n ensayos.
- p^x representa la probabilidad de x éxitos.
- $(1 - p)^{n-x}$ representa la probabilidad de $n - x$ fracasos.
- **Valor Esperado y Varianza:** El valor esperado ($E[X]$) de una variable aleatoria binomial con parámetros n y p es igual a np , y la varianza ($Var[X]$) es igual a $np(1 - p)$.

Ejemplos de Aplicación:

1. **Lanzamiento de una Moneda Justa Repetidamente:** Si lanzamos una moneda justa repetidamente (cara = éxito, cruz = fracaso) y queremos saber la probabilidad de obtener 3 caras en 5 lanzamientos, podemos modelar esto como una distribución binomial con $n=5$ y $p=0.5$.
2. **Prueba de Éxito en Producción:** En una línea de producción, si la probabilidad de que un artículo sea defectuoso es del 0.1 y deseamos calcular la probabilidad de que 4 de los 10 artículos producidos sean defectuosos, podemos usar una distribución binomial con $n=10$ y $p=0.1$.
3. **Probabilidad de Encestar Tiros Libres:** Si un jugador de baloncesto tiene una tasa de éxito del 70% en tiros libres y quiere calcular la probabilidad de encestar exactamente 5 de 7 tiros libres, esto se modela con una distribución binomial con $n=7$ y $p=0.7$.

Cálculos y Propiedades:

- Para calcular la probabilidad de obtener exactamente x éxitos en n ensayos, utilizamos la función de probabilidad de la distribución binomial.
- El valor esperado de una variable binomial se calcula como $E[X] = np$, donde n es el número de ensayos y p es la probabilidad de éxito.
- La varianza de una variable binomial se calcula como $Var[X] = np(1 - p)$.
- La distribución binomial se utiliza comúnmente en estadísticas para modelar eventos que involucren un número fijo de ensayos independientes con dos resultados posibles.

Ejemplo en Python:

Supongamos que lanzamos una moneda 10 veces. ¿Cuál es la probabilidad de obtener exactamente 6 caras?

```
python:  
from scipy.stats import binom  
  
n = 10  
p = 0.5  
k = 6  
probabilidad = binom.pmf(k, n, p)  
print(probabilidad)
```

B. Distribución normal o Gaussiana

La distribución normal, también conocida como gaussiana, es una de las distribuciones de probabilidad más importantes y ampliamente utilizadas en estadísticas y probabilidad. Se caracteriza por su forma de campana y es aplicable en una amplia variedad de situaciones debido a su propiedad de ser simétrica y tener una media y una desviación estándar bien definidas. A continuación, se presentan algunos conceptos clave y ejemplos de la distribución normal:

Conceptos Clave:

- **Parámetros μ (mu) y σ (sigma):** La distribución normal tiene dos parámetros: la media (μ) y la desviación estándar (σ). La media determina el centro de la distribución y la desviación estándar controla la dispersión de los datos.
- **Función de Densidad de Probabilidad (PDF):** La función de densidad de probabilidad de la distribución normal se define como:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

donde x es el valor que deseamos evaluar.

- **Valor Esperado y Varianza:** El valor esperado ($E[X]$) de una variable aleatoria normal es igual a la media (μ), y la varianza ($\text{Var}[X]$) es igual al cuadrado de la desviación estándar (σ^2).

Ejemplos de Aplicación:

1. **Altura de Individuos:** La altura de las personas tiende a seguir una distribución normal. Si la media de altura en una población es de 170 cm y la desviación estándar es de 10 cm, podemos usar una distribución normal para modelar la altura de las personas.
2. **Puntajes en un Examen:** Los puntajes en un examen pueden seguir una distribución normal. Si la media de puntajes es 75 y la desviación estándar es 10, podemos usar una distribución normal para estimar la probabilidad de obtener ciertos puntajes.
3. **Errores de Medición:** Los errores de medición en experimentos científicos a menudo se distribuyen normalmente. La media representa el valor verdadero y la desviación estándar mide la variabilidad de las mediciones.

Cálculos y Propiedades:

- El valor esperado de una variable normal es igual a la media ($E[X] = \mu$).
- La varianza de una variable normal es igual al cuadrado de la desviación estándar ($\text{Var}[X] = \sigma^2$).
- La distribución normal es simétrica alrededor de su media, y su forma de campana indica que la mayoría de los valores se concentran cerca de la media.
- La regla empírica de la distribución normal establece que aproximadamente el 68% de los datos caen dentro de una desviación estándar de la media, el 95% caen dentro de dos desviaciones estándar y el 99.7% caen dentro de tres desviaciones estándar.
- La distribución normal es ampliamente utilizada en inferencia estadística, pruebas de hipótesis, análisis de regresión y muchas otras aplicaciones debido a sus propiedades matemáticas y su aparición frecuente en la naturaleza.

Ejemplo en Python:

Para visualizar una distribución normal con $\mu = 0$ y $\sigma = 1$:

```
python:
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import norm

x = np.linspace(-5, 5, 1000)
y = norm.pdf(x, 0, 1)
plt.plot(x, y)
plt.title('Distribución Normal Estándar')
plt.xlabel('x')
plt.ylabel('f(x)')
plt.show()
```

C. Distribución Poisson

La distribución de Poisson es una distribución de probabilidad discreta que modela la ocurrencia de eventos raros en un intervalo fijo de tiempo o espacio. Se utiliza comúnmente para contar el

número de eventos que ocurren en un intervalo específico cuando la tasa promedio de ocurrencia es conocida. A continuación, algunos conceptos clave y ejemplos:

Conceptos Clave:

- **Parámetro λ (lambda):** En la distribución de Poisson, el parámetro λ (lambda) representa la tasa promedio de ocurrencia de eventos en un intervalo fijo. Es una medida de la frecuencia esperada de eventos en ese intervalo.
- **Función de Probabilidad:** La función de probabilidad de la distribución de Poisson se define como:

$$P(X = x) = \frac{e^{-\lambda} * \lambda^x}{x!}$$

Donde x es el número de eventos que deseamos contar.

- **Valor Esperado y Varianza:** El valor esperado ($E[X]$) de una variable aleatoria Poisson con parámetro λ es igual a λ , y la varianza ($Var[X]$) también es igual a λ .

Ejemplos de Aplicación:

1. **Llamadas a una Línea de Atención al Cliente:** Supongamos que en promedio se reciben 5 llamadas por minuto en una línea de atención al cliente. Podemos modelar el número de llamadas que llegan en un minuto específico utilizando una distribución de Poisson con $\lambda = 5$.
2. **Errores de Típo en un Documento:** Si en promedio hay 2 errores de tipeo por página en un documento, podemos usar una distribución de Poisson con $\lambda = 2$ para modelar el número de errores en una página dada.
3. **Llegada de Correos Electrónicos:** Si en promedio se reciben 15 correos electrónicos por hora en una bandeja de entrada, podemos aplicar una distribución de Poisson con $\lambda = 15$ para estimar cuántos correos electrónicos llegarán en una hora específica.

Cálculos y Propiedades:

- El valor esperado de una variable Poisson se calcula como $E[X] = \lambda$.
- La varianza de una variable Poisson también es igual a $Var[X] = \lambda$.
- La distribución de Poisson es especialmente útil para modelar eventos raros donde la ocurrencia de un evento no afecta la ocurrencia de otros eventos en el mismo intervalo (propiedad de independencia).
- A medida que el intervalo se vuelve más grande, la distribución de Poisson se aproxima a la distribución normal.
- La distribución de Poisson es adecuada para eventos que ocurren a una tasa constante y baja, como accidentes de tráfico en una intersección o llamadas a un centro de servicio al cliente.

Ejemplo en Python:

¿Cuál es la probabilidad de que 5 autos pasen por un punto específico en 10 minutos si en promedio pasan 7 autos en ese intervalo?

```
python:
from scipy.stats import poisson

lambda_val = 7
k = 5
probabilidad = poisson.pmf(k, lambda_val)
print(probabilidad)
```

D. Distribución uniforme

La distribución uniforme es un tipo de distribución de probabilidad que se caracteriza por asignar la misma probabilidad a todos los valores posibles de una variable aleatoria dentro de un intervalo específico. Hay dos tipos de distribuciones uniformes: la distribución uniforme continua y la distribución uniforme discreta.

Distribución Uniforme Continua

Conceptos Clave:

- **Parámetros a y b:** En la distribución uniforme continua, los parámetros a y b definen el intervalo en el cual la variable aleatoria puede tomar valores. La probabilidad es constante en este intervalo y es igual a $\frac{1}{b-a}$.
- **Función de Densidad de Probabilidad (PDF):** La función de densidad de probabilidad de una variable uniforme continua se define como:

$$\begin{cases} \frac{1}{b-a} & \text{si } a \leq x \leq b \\ 0 & \text{en otro caso} \end{cases}$$

- **Valor Esperado y Varianza:** El valor esperado ($E[X]$) de una variable uniforme continua con parámetros a y b es igual a $\frac{a+b}{2}$, y la varianza ($\text{Var}[X]$) es igual a $\frac{(b-a)^2}{12}$.

Ejemplos de Aplicación:

1. **Selección Aleatoria en un Intervalo:** Si seleccionamos un número aleatorio entre 0 y 1 de manera uniforme, estamos modelando una distribución uniforme continua en el intervalo $[0, 1]$. La probabilidad de seleccionar cualquier valor dentro de este intervalo es la misma.
2. **Tiempo de Espera Aleatorio:** Supongamos que el tiempo que toma que un autobús llegue a una parada está uniformemente distribuido entre 5 y 15 minutos. Esto se puede modelar como una distribución uniforme continua en el intervalo $[5, 15]$.

Distribución Uniforme Discreta

Conceptos Clave:

- **Parámetros a y b:** En la distribución uniforme discreta, los parámetros a y b definen un conjunto finito de valores enteros que la variable aleatoria puede tomar. Cada valor en este conjunto tiene la misma probabilidad de ocurrencia.
- **Función de Probabilidad:** La función de probabilidad de una variable uniforme discreta asigna la misma probabilidad a cada valor en el conjunto. La probabilidad de cada valor es $\frac{1}{b-a+1}$.
- **Valor Esperado y Varianza:** El valor esperado ($E[X]$) de una variable uniforme discreta con parámetros a y b es igual a $\frac{a+b}{2}$, y la varianza ($\text{Var}[X]$) es igual a $\frac{(b-a+1)^2-1}{12}$.

Ejemplos de Aplicación:

1. **Lanzamiento de un Dado:** Si lanzamos un dado justo de seis caras, cada número del 1 al 6 tiene la misma probabilidad de $1/6$ de aparecer en un solo lanzamiento. Esto sigue una distribución uniforme discreta en el conjunto $\{1, 2, 3, 4, 5, 6\}$.
2. **Selección Aleatoria de Elementos:** Si seleccionamos un elemento aleatorio de una baraja de cartas estándar, cada carta tiene la misma probabilidad de $1/52$ de ser seleccionada. Esto sigue una distribución uniforme discreta en el conjunto de cartas.

Cálculos y Propiedades:

- Para calcular el valor esperado de una variable uniforme continua, utilizamos $E[X] = (a + b)/2$ y para una variable uniforme discreta, también usamos $E[X] = (a + b)/2$.
- La varianza de una variable uniforme continua se calcula como $Var[X] = (b - a)^2/12$, y para una variable uniforme discreta, se calcula como $Var[X] = [(b - a + 1)^2 - 1]/12$.
- La distribución uniforme es útil en situaciones donde no se tiene información previa que sugiera una preferencia por valores particulares y se asume igual probabilidad para todos los valores dentro de un rango.

Ejemplo en Python:

Generar números aleatorios con una distribución uniforme y visualizarlos.

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

# Generar números aleatorios
valores = np.random.uniform(-10.0, 10.0, 10000)

# Visualizar histograma
plt.hist(valores, 50)
plt.title('Distribución Uniforme')
plt.show()
```

E. Distribución Exponencial

La distribución exponencial es una distribución de probabilidad continua que se utiliza comúnmente para modelar el tiempo entre eventos sucesivos en un proceso de Poisson. Es especialmente útil para describir eventos que ocurren de manera aleatoria e independiente en el tiempo. A continuación, se presentan algunos conceptos clave y ejemplos de la distribución exponencial:

Conceptos Clave:

- **Parámetro λ (lambda):** El parámetro λ (lambda) en la distribución exponencial representa la tasa de ocurrencia promedio de eventos por unidad de tiempo. Es el inverso de la media ($\lambda = 1/\mu$), donde μ es la media.
- **Función de Densidad de Probabilidad (PDF):** La función de densidad de probabilidad de la distribución exponencial se define como:

$$\begin{cases} \lambda e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$$

Donde x es el valor que deseamos evaluar.

- **Valor Esperado y Varianza:** El valor esperado ($E[X]$) de una variable aleatoria exponencial con parámetro λ es igual a $1/\lambda$, y la varianza ($Var[X]$) es igual a $1/\lambda^2$.

Ejemplos de Aplicación:

1. **Tiempo entre Llegadas de Clientes:** Si en promedio llega un cliente cada 10 minutos a un restaurante, podemos modelar el tiempo entre llegadas como una distribución exponencial con $\lambda = 1/10$.

2. **Tiempo de Espera en una Cola:** La distribución exponencial se utiliza para modelar el tiempo que una persona debe esperar en una cola antes de ser atendida.
3. **Duración de Llamadas Telefónicas:** Si las llamadas telefónicas en una línea de servicio al cliente tienen una tasa promedio de 4 por hora, el tiempo entre llamadas se modela con una distribución exponencial con $\lambda=4$.

Cálculos y Propiedades:

- El valor esperado de una variable exponencial es igual a $1/\lambda$ ($E[X]=1/\lambda$).
- La varianza de una variable exponencial es igual a $1/\lambda^2$ ($Var[X] = 1/\lambda^2$).
- La distribución exponencial tiene una cola larga a la derecha, lo que significa que eventos raros, pero extremadamente largos pueden ocurrir.
- La función de densidad de probabilidad de la distribución exponencial es decreciente, lo que indica que a medida que pasa el tiempo, es menos probable que ocurran eventos.
- La distribución exponencial es ampliamente utilizada en teoría de colas, fiabilidad, y en la modelación de eventos aleatorios en una variedad de campos.

Ejemplo en Python:

Generaremos números aleatorios para una distribución exponencial y crearemos su gráfico correspondiente.

```
python:
lambda_val = 0.1
valores = np.random.exponential(lambda_val, 10000)
plt.hist(valores, 50)
plt.title('Distribución Exponencial')
plt.show()
```

6. Medidas de Información

Entropía

La entropía es una medida fundamental en la teoría de la información y se utiliza para cuantificar la cantidad promedio de sorpresa o incertidumbre en una distribución de probabilidad. Cuanto más incierta sea una distribución, mayor será su entropía.

La fórmula general para calcular la entropía $H(X)$ de una variable aleatoria discreta X con distribución de probabilidad $P(X)$ es:

$$H = - \sum_{i=1}^n P(c_i) \log_2 P(c_i)$$

donde c_i son los posibles valores de X y $P(c_i)$ es la probabilidad de que X tome el valor c_i .

La entropía alcanza su valor máximo cuando todos los resultados son igualmente probables, lo que significa que no podemos hacer ninguna suposición sobre lo que ocurrirá. Por el contrario, la entropía es mínima cuando un resultado es seguro y los demás son imposibles.

Entropía Cruzada

La entropía cruzada, también conocida como divergencia cruzada o cross-entropy en inglés, es una medida que se utiliza para cuantificar la diferencia entre dos distribuciones de probabilidad. Se usa comúnmente en el contexto de la clasificación y el aprendizaje automático para evaluar cuán bien una distribución de probabilidad (generalmente una predicción) se ajusta a otra (generalmente la distribución real de los datos).

La fórmula general para calcular la entropía cruzada $H(P,Q)$ entre dos distribuciones de probabilidad P y Q es:

$$H = - \sum_{i=1}^n P(c_i) \log_2 Q(c_i)$$

donde c_i son los posibles valores, $P(c_i)$ es la probabilidad real y $Q(c_i)$ es la probabilidad predicha.

La entropía cruzada es mayor cuando las dos distribuciones difieren significativamente. En el contexto del aprendizaje automático, minimizar la entropía cruzada es un objetivo común al entrenar modelos de clasificación y regresión, ya que indica que las predicciones del modelo se acercan a las distribuciones reales de los datos.

Z-Score

El Z-Score es una medida estadística que se utiliza para evaluar y comparar datos en términos de desviaciones estándar por encima o por debajo de la media. Se utiliza comúnmente para estandarizar y normalizar datos facilitando la comparación e identificación de valores atípicos en un conjunto de datos.

El cálculo se realiza mediante la siguiente forma:

$$z = \frac{x - \mu}{\sigma}$$

donde

- z es el Z-Score
- x es el valor que se está evaluando
- μ es la medida del conjunto de datos
- σ es la desviación estándar del conjunto de datos

Si Z es positivo indica que el valor está por encima de la media, y si es negativo, que está por debajo de la media. Además, cuanto mayor sea su valor absoluto más lejos estará el valor de la media en términos de desviaciones estándar.

7. Inferencia Estadística y Tests^{iv}

Comprendiendo test de hipótesis y p-values

El proceso de inferencia estadística permite a los científicos y profesionales de datos hacer afirmaciones sobre poblaciones basadas en muestras.

Una herramienta fundamental en este proceso es la prueba de hipótesis, que evalúa dos afirmaciones mutuamente exclusivas sobre una población.

- Hipótesis Nula (H_0): Afirmación inicial o de "status quo" sobre un parámetro poblacional.
- Hipótesis Alternativa (H_a): Afirmación que queremos probar contra la hipótesis nula.

A modo de ejemplo, supongamos que deseamos probar si una moneda está cargada. Aquí, nuestra H_0 podría ser que la moneda esta equilibrada (probabilidad de cara = 0.5), mientras que H_a podría ser que la probabilidad de cara es diferente 0.5.

Una métrica crucial en la prueba de hipótesis es el valor-P que representa la probabilidad de observar un resultado tan extremo como el obtenido, o más, dado que la hipótesis nula es cierta. Si el valor-P es pequeño rechazamos la hipótesis nula en favor de la alternativa.

Supongamos que lanzamos una moneda 100 veces y obtenemos 60 caras. ¿Está cargada la moneda?

Matemáticamente, nuestro test sería:

$$H_0: p = 0.5$$
$$H_a: p \neq 0.5$$

que desarrollado en Python usando la distribución binomial quedaría

```
python:
from scipy.stats import binom_test

# Resultados observados
num_caras = 60
num_total = 100

# Realizar test de hipótesis
p_value = binomtest(num_caras, num_total, 0.5)
print("p-value:", p_value)
```

Aplicación: Implementación de A/B Testing en Python

Supongamos que “TechAI Ltd.” desea probar un nuevo diseño de landing page. La métrica clave que desean mejorar es la tasa de clics en el botón “Registrarse” para lo cual van a hacer un A/B testing donde evalúen el diseño nuevo (Diseño B) contra el diseño actual (Diseño A).

En este contexto, antes de indicar los pasos, introduzcamos un concepto estadístico como estadística t. Este concepto nos permite comparar las tasas de conversión de ambas variantes y determinar si las diferencias observadas son significativas.

Si p_A y p_B son las tasas de conversión $p_X = \frac{\text{Numero de conversiones (o usuarios que realicen X)}}{\text{Visitantes totales del sitio}}$, $p = \frac{\text{Numero de conversiones A y B}}{\text{Visitantes totales del sitio para A y B}}$ la tasa de conversión global y n_A y n_B son los números de visitantes en cada grupo, la estadística t se calcula como:

$$t = \frac{p_A - p_B}{\sqrt{p(1-p) \left(\frac{1}{n_A} + \frac{1}{n_B} \right)}}$$

Una vez calculada la estadística t, se compara con un valor crítico de la distribución t (basado en un nivel de significancia predefinido y los grados de libertad). Si la magnitud de t supera el valor crítico, se rechaza la hipótesis nula (que supone que no hay diferencia entre las variantes A y B en cuanto a tasas de conversión) y se puede concluir que hay una diferencia significativa en las tasas de conversión entre las dos variantes.

Una vez dicho esto, procedamos con el ejercicio:

1) Hipótesis

- a. H_0 : No hay diferencia entre el diseño A y B en términos de tasa de clics.
- b. H_a : El diseño B tiene una tasa de clics significativamente diferente (puede ser mayor o menor) que el diseño A

2) Recopilación de Datos

En el primer día del experimento, 1000 visitantes fueron expuestos al diseño A y 150 hicieron clic en "Registrarse". Por otro lado, 1000 visitantes vieron el diseño B y 180 hicieron clic en "Registrarse".

3) Cálculo y Test Estadístico

Si implementamos en python, tendremos:

```
python:
import scipy.stats as stats

# Datos
n_A, clics_A = 1000, 150
n_B, clics_B = 1000, 180

# Tasas de clics
p_A, p_B = clics_A / n_A, clics_B / n_B

# Tasa de clics global
p = (clics_A + clics_B) / (n_A + n_B)

# Estadística t
numerator = p_A - p_B
denominator = (p*(1-p)*(1/n_A + 1/n_B))**0.5
t_stat = numerator / denominator

# Valor-P
p_value = stats.t.sf(abs(t_stat), df=min(n_A, n_B)-1)*2

print("Estadística t:", t_stat)
print("Valor-P:", p_value)
```

4) Interpretación

Si el valor-P es inferior a un umbral, como 0.05, rechazamos H_0 y concluimos que el diseño B tiene un impacto significativo en la tasa de clics. En el caso de que el valor-P sea superior al umbral, no podemos afirmar con seguridad que el diseño B sea mejor.

ⁱ Wasserman, L. (2013). All of statistics: a concise course in statistical inference. Springer Science & Business Media.

ⁱⁱ DeGroot, M. H., & Schervish, M. J. (2012). Probability and statistics. Pearson Education.

ⁱⁱⁱ DeGroot, M. H., & Schervish, M. J. (2012). Probability and statistics. Pearson Education.

^{iv} Casella, G., & Berger, R. L. (2002). Statistical inference. Pacific Grove, CA: Duxbury.