Probabilidad

Universidad Internacional de Valencia

Máster Universitario en Inteligencia Artificial

02MIAR | Matemáticas:

Matemáticas para la Inteligencia Artificial

Profesor:

David Zorío Ventura

De:
Planeta Formación y Universidades

Variables aleatorias

Definición

Una **variable aleatoria** *(V.A.) X sobre un espacio muestral Ω es una función $\Omega \to \mathbb{R}$. En ese caso, es posible inducir una probabilidad asociada a esa variable aleatoria sobre $X(\Omega) \subseteq \mathbb{R}$. Dado un suceso $A \subseteq \mathbb{R}$, definimos*

$$P_X(A) \equiv P(X \in A) := P\big(X^{-1}(A)\big) = P\big(\{\omega \in \Omega \mid X(\omega) \in A\}\big).$$

Ejemplos

1. *Sea Ω el conjunto formado por la población mundial $X : \Omega \to \mathbb{R}$ la variable aleatoria que a una persona $\omega \in \Omega$ le asigna su año de nacimiento, $X(\omega) \in \mathbb{R}$. Entonces si $A = \{1984, 1985, 1986\}$:*

$$X^{-1}(A) = \{\textit{personas nacidas entre 1984 y 1986}\}$$

$$P_X(A) = \textit{proporción de personas nacidas entre 1984 y 1986.}$$

Ejemplos

2. *Sea Ω el espacio muestral asociado al lanzamiento de tres monedas:*

$$\Omega = \{ZZZ, CZZ, ZCZ, ZZC, CCZ, CZC, ZCC, CCC\}.$$

- *Sea $X : \Omega \to \mathbb{R}$ la V.A. que cuenta el número de caras obtenidas:*

- $X(ZZZ) = 0.$
- $X(CZZ) = 1.$
- $X(ZCZ) = 1.$
- $X(ZZC) = 1.$
- $X(CCZ) = 2.$
- $X(CZC) = 2.$
- $X(ZCC) = 2.$
- $X(CCC) = 3.$
- $X^{-1}(0) = \{ZZZ\}.$
- $X^{-1}(1) = \{CZZ, ZCZ, ZZC\}.$
- $X^{-1}(2) = \{CCZ, CZC, ZCC\}.$
- $X^{-1}(3) = \{CCC\}.$
- $P(X = 0) = \frac{1}{8}.$
- $P(X = 1) = \frac{3}{8}.$
- $P(X = 2) = \frac{3}{8}.$
- $P(X = 3) = \frac{1}{8}.$

- $P(1 < X \leq 3.5) = P\left(X^{-1}\left(]1, 3.5]\right)\right) = P\left(\{\omega \in \Omega \mid X(\omega) \in]-1, 3.5]\}\right) =$
  $P\left(\{CCZ, CZC, ZCC, CCC\}\right) = \dfrac{1}{2}$

Variables aleatorias discretas

Definición

Sea X una V.A. Si el conjunto $X(\Omega)$ es finito o numerable (indexable por números naturales), entonces diremos que X es una **variable aleatoria discreta**.

En ese caso, definimos la **función de densidad** *o función de masa de probabilidad asociada como $f_X : \mathbb{R} \to [0,1]$ dada por*

$$f_X(x) = P(X = x) = P_X(\{x\}) = P(X^{-1}(\{x\})) = P(\{\omega \in \Omega \mid X(\omega) = x\}), \quad x \in \mathbb{R}.$$

Definimos la **función de distribución** *asociada como $F_X : \mathbb{R} \to [0,1]$ dada por*

$$F_X(x) = P(X \leq x) = \sum_{t \in \Omega, X(t) \leq x} f_X(X(t)), \quad x \in \mathbb{R}.$$

Condición de normalización: $\sum_{t \in \Omega} f_X(X(t)) = 1.$

Ejemplos

1. *Sea* $\Omega = \{uno, dos, tres, cuatro, cinco, seis\}$ *el conjunto de los posibles resultados al lanzar un dado y* $X : \Omega \to \mathbb{R}$ *la V.A. que le asigna su valor:* $X(uno) = 1$, $X(dos) = 2$, $X(tres) = 3$, $X(cuatro) = 4$, $X(cinco) = 5$ *y* $X(seis) = 6$. *Entonces*

$$X(\Omega) = \{1, 2, 3, 4, 5, 6\}.$$

$$f_X(4) = P(X = 4) = P(\{cuatro\}) = \frac{1}{6}.$$

Por otra parte,

$$F_X(4) = P(X \leq 4) = P(\{uno, dos, tres, cuatro\}) = \frac{2}{3}.$$

Ejemplos

2. *Consideramos el experimento aleatorio consistente en lanzar una moneda hasta obtener como resultado cara. El espacio muestral es por tanto*

$$\Omega = \{cara, (cruz, cara), (cruz, cruz, cara), (cruz, cruz, cruz, cara), \ldots\}.$$

- *Sea X la V.A. aleatoria consistente en contar el número de lanzamientos hasta obtener cara. Entonces:*

- $X(cara) = 1.$
- $X\big((cruz, cara)\big) = 2.$
- $X\big((cruz, cruz, cara)\big) = 3.$
- $X\big((cruz, cruz, cruz, cara)\big) = 4.$

  $\vdots$
- *Por tanto,* $X(\Omega) = \{1, 2, 3, 4, \ldots\} = \mathbb{N}.$

- $f_X(1) = \frac{1}{2}.$
- $f_X(2) = \frac{1}{4}.$
- $f_X(3) = \frac{1}{8}.$
- $f_X(4) = \frac{1}{16}.$

  $\vdots$
- $f_X(n) = \frac{1}{2^n}, \forall n \in \mathbb{N}.$

- $$\sum_{t\in\Omega} f_X(X(t)) = \sum_{n\in\mathbb{N}} f_X(n) = \sum_{n\in\mathbb{N}} P(X = n) = \sum_{n=1}^{\infty} \frac{1}{2^n} = 1.$$

Variables aleatorias continuas

Definición

Sea X una V.A. asociada a un espacio muestral Ω. Diremos que X es una V.A. **continua** *si $X(\Omega)$ es continuo (por ejemplo, un intervalo o todos los reales).*

Definimos la **función de distribución** *asociada a X como $F_X : \mathbb{R} \to [0, 1]$ dada por*

$$F_X(x) = P(X \leq x).$$

Asimismo, diremos que X admite una **función de densidad** *si $\exists f_X : \mathbb{R} \to [0, +\infty[$ no negativa e integrable tal que para cada $A \subseteq \mathbb{R}$*

$$P(X \in A) = \int_A f_X(t)\mathrm{d}t.$$

Condición de normalización: $\displaystyle\int_{-\infty}^{+\infty} f_X(t)\mathrm{d}t = 1.$

Variables aleatorias continuas

- Si X es una V.A. continua que admite una función de densidad, entonces podemos escribir

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)\mathrm{d}t.$$

- La función de densidad debe satisfacer la condición de normalización:

$$\int_{-\infty}^{+\infty} f_X(t)\mathrm{d}t = P(X \in \mathbb{R}) = 1.$$

- De la existencia de una función de densidad se deduce que

$$P(X < x) = \lim_{t \to x^-} F_X(t) = F_X(x) = P(X \leq x).$$

- En particular, se deduce que en ese caso

$$P(X = x) = 0, \ \forall x \in \mathbb{R}.$$

Ejemplo

Se elige un número al azar dentro del intervalo real $[0, 1]$ y se considera la V.A. aleatoria asociada a dicho experimento. Entonces la función de distribución asociada a X es:

$$F_X(x) = P(X \leq x) = \begin{cases} 0 & \textit{si } x < 0, \\ x & \textit{si } 0 \leq x \leq 1, \\ 1 & \textit{si } x > 1. \end{cases}$$

Por tanto, la función de densidad asociada a X es:

$$f_X(x) = \begin{cases} 0 & \textit{si } x < 0, \\ 1 & \textit{si } 0 \leq x \leq 1, \\ 0 & \textit{si } x > 1. \end{cases}$$

Esperanza

Definición

Sea X una V.A. asociada a un espacio muestral Ω, y con función de densidad $f_X(x)$. Definimos la **esperanza** *de X como*

$$E[X] = \begin{cases} \sum_{x \in X(\Omega)} x f_X(x) & \text{si } X \text{ es una V.A. discreta,} \\ \int_{\mathbb{R}} x f_X(x) \mathrm{d}x & \text{si } X \text{ es una V.A. continua.} \end{cases}$$

Propiedades

Sean X, Y variables aleatorias, $a \in \mathbb{R}$. Entonces:

1. $E[X + Y] = E[X] + E[Y]$.
2. $E[aX] = aE[X]$.
3. *Si X, Y son independientes, entonces* $E[XY] = E[X]E[Y]$.

Ejemplos

1. *El precio del número de una rifa es de 1 euro. Si ésta tiene un total de 1000 números y el premio tiene un valor de 500 euros, ¿cuál es el promedio de beneficio esperado?*

   Llamamos X a la V.A. que cuantifica el beneficio obtenido al participar en la rifa:

   $$\text{Perder la rifa: } X = -1 \rightarrow P(X = -1) = \tfrac{999}{1000}.$$
   $$\text{Ganar la rifa: } X = 499 \rightarrow P(X = 499) = \tfrac{1}{1000}.$$

   Por tanto:

   $$E[X] = (-1) \cdot P(X = -1) + 499 \cdot P(X = 499) = -\frac{999}{1000} + \frac{499}{1000} = -0.5.$$

Ejemplos

2. *Consideramos el intervalo* $[0, 1]$ *y seleccionamos un punto al azar y nos planteamos cuál es el valor sobre el que oscilará el promedio. Vimos que*

$$f_X(x) = \begin{cases} 0 & si\ x < 0, \\ 1 & si\ 0 \leq x \leq 1, \\ 0 & si\ x > 1. \end{cases}$$

Por tanto:

$$E[X] = \int_{-\infty}^{+\infty} x f_X(x) \mathrm{d}x = \int_0^1 x \mathrm{d}x = \left[\frac{x^2}{2}\right]_0^1 = \frac{1}{2}.$$

Varianza

Universidad Internacional de Valencia

Definición

Sea X una V.A. y $g : \mathbb{R} \to \mathbb{R}$. Definimos

$$E[g(X)] = \begin{cases} \sum_{x \in X(\Omega)} g(x) f_X(x) & \textit{si } X \textit{ es una V.A. discreta,} \\ \int_{\mathbb{R}} g(x) f_X(x) \mathrm{d}x & \textit{si } X \textit{ es una V.A. continua.} \end{cases}$$

Definición

Sea X una V.A. aleatoria. Definimos la **varianza** *de X como:*

$$\mathrm{Var}(X) = E\left[(X - E[X])^2\right] = E\left[X^2\right] - E[X]^2.$$

Propiedades

1. $\mathrm{Var}(aX) = a^2 \mathrm{Var}(X)$, $\forall a \in \mathbb{R}$.
2. $\mathrm{Var}(X + a) = \mathrm{Var}(X)$, $\forall a \in \mathbb{R}$.
3. $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$ *si X, Y son independientes.*

Ejemplos

1. *En el ejemplo de la rifa, se tiene*

$$E\left[X^2\right] = \sum_{x\in\{-1,499\}} x^2 f_X(x) = (-1)^2 \cdot P(X=-1) + 499^2 \cdot P(X=499)$$

$$= \frac{999}{1000} + \frac{249001}{1000} = 250.$$

Por tanto, $\mathrm{Var}(X) = E[X^2] - E[X]^2 = 250 - 0.5^2 = 249.75.$

2. *Selección al azar de un número en el intervalo* $[0,1]$*:*

$$E\left[X^2\right] = \int_{-\infty}^{+\infty} x^2 f_x(x)\mathrm{d}x = \int_0^1 x^2 \mathrm{d}x = \left[\frac{x^3}{3}\right]_0^1 = \frac{1}{3}.$$

Luego $\mathrm{Var}(X) = E[X^2] - E[X]^2 = \dfrac{1}{3} - \left(\dfrac{1}{2}\right)^2 = \dfrac{1}{12}.$

Distribución de Bernoulli

Una V.A. X sigue la **distribución de Bernoulli** de parámetro $p \in [0, 1]$, y lo denotaremos por $X \sim \text{Be}(p)$, si $X \in \{0, 1\}$, verificando

$$P(X = 1) = p, \quad P(X = 0) = 1 - p.$$

Podemos calcular la esperanza y varianza de X como sigue:

$$E[X] = p, \quad \text{Var}(X) = p(1 - p).$$

Ejemplo

Consideramos el experimento consistente en devolver éxito ($X = 1$) si al lanzar un dado obtenemos un 5 y fracaso ($X = 0$) en otro caso. Entonces $X \sim \text{Be}(\frac{1}{6})$.

Distribución binomial

Si $X = X_1 + \cdots + X_n$, con $X_i \sim \mathrm{Be}(p)$ independientes entre sí, es decir, si X es una V.A. aleatoria que cuenta el número de éxitos tras realizar n pruebas Bernoulli de parámetro p independientes, diremos que X sigue una **distribución binomial** de parámetros n y p, que denotaremos por $X \sim \mathrm{B}(n, p)$. Se cumple

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad \text{donde } \binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Por otra parte,

$$E[X] = np, \quad \mathrm{Var}(X) = np(1-p).$$

Distribución binomial

Universidad Internacional de Valencia

Ejemplo

Lanzamos un dado 10 veces y queremos calcular cuál es la probabilidad de obtener tres veces un cinco. Si X es la V.A. que cuenta el número de veces que se obtiene un cinco tras lanzar el dado 10 veces, entonces $X \sim \mathrm{B}(10, \frac{1}{6})$, por lo que $n = 10$, $p = \frac{1}{6}$ y $k = 3$, con lo cual se tiene:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} = \frac{10!}{3! \cdot 7!} \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^7 = \frac{390625}{2519424}.$$

Distribución de Poisson

Supongamos que realizamos un número de pruebas Bernoulli muy alto n, con una probabilidad p_n cada vez más pequeña, de forma que

$$\lim_{n\to\infty} np_n = \lambda.$$

Si $X_n \sim \mathrm{B}(n, p_n)$, entonces se puede demostrar que

$$\lim_{n\to\infty} P(X_n = k) = \frac{e^{-\lambda}\lambda^k}{k!}.$$

Si X es una variable de este tipo, diremos que X sigue una **distribución de Poisson** de parámetro λ, que denotaremos por $X \sim \mathrm{Po}(\lambda)$. Se cumple

$$E[X] = \lambda, \quad \mathrm{Var}(X) = \lambda.$$

Universidad Internacional de Valencia

Ejemplo

La probabilidad media de tener un accidente de tráfico al coger el coche en una determinada región es de un 0.01 %. *Si una persona coge el coche una media total de* 20000 *veces a lo largo de su vida, ¿cuál es la probabilidad de que sufra algún accidente de tráfico en algún momento de su vida?*

Sea X la V.A. aleatoria que contabiliza el número de accidentes de tráfico en esas circunstancias. Como $n = 20000$ y $p = 0.0001$, tenemos $\lambda = np = 2$, y podemos decir que $X \approx \mathrm{Po}(2)$. Por tanto, se nos pide

$$P(X > 0) = 1 - P(X = 0) = 1 - \frac{e^{-\lambda}\lambda^0}{0!} = 1 - e^{-2} \approx 0.8647.$$

- **Distribución uniforme:**

  Si X es una V.A. continua que toma valores uniformemente en el intervalo $[a, b]$, entonces diremos que X sigue una **distribución uniforme**, y lo denotaremos por $X \sim \text{Unif}\big([a, b]\big)$.

Ejercicio

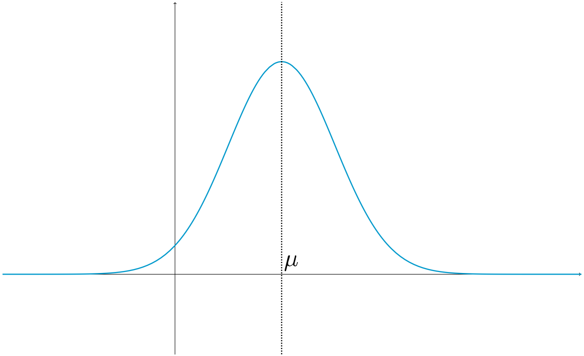
Si $X \sim \text{Unif}\big([a, b]\big)$, *obtener razonadamente* F_X, f_x, $E[X]$ *y* $\text{Var}(X)$.

- **Distribución normal:**

  Una V.A. X sigue una **distribución normal** de **media** μ y **desviación típica** σ, denotado por $X \sim N(\mu, \sigma)$, si se cumple

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Distribución normal



Se cumple $E[X] = \mu$ y $\mathrm{Var}(X) = \sigma^2$.

Teorema

Si $X \sim N(\mu, \sigma)$, entonces $\dfrac{X-\mu}{\sigma} \sim N(0,1)$. Usualmente esta última distribución recibe el nombre de ***normal tipificada****.*

Universidad Internacional de Valencia

Ejemplo

La altura (en centímetros) de una determinada población sigue una distribución normal de media 175 y desviación típica 10. ¿Cuál es la probabilidad de que una persona elegida al azar mida más de 185 cm?

Sea X la V.A. asociada, luego $X \sim N(175, 10)$. Si definimos $Z = \frac{X-175}{10} \sim N(0,1)$, se tiene

$$\begin{aligned} P(X > 185) &= P\left(\frac{X-175}{10} > \frac{185-175}{10}\right) \\ &= P(Z > 1) = 1 - P(Z \leq 1) \approx 1 - 0.8413 = 0.1587. \end{aligned}$$

Comandos utilizados:

```
from scipy.stats import norm
1-norm.cdf(1)
```

Teorema del límite central

Teorema (Teorema del límite central)

Sea $\{X_i\}_{i=1}^{\infty}$ *una sucesión de variables aleatorias idénticamente distribuidas (en adelante, i.i.d.), con* $E[X_i] = \mu$ *y* $\mathrm{Var}(X_i) = \sigma^2 \; \forall i \in \mathbb{N}$*. Denotamos para cada* $n \in \mathbb{N}$

$$S_n = \sum_{i=1}^{n} X_i \text{ y definimos } Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}.$$

Entonces se cumple

$$\lim_{n\to\infty} P(Z_n \leq t) = P(Z \leq t),$$

donde $Z \sim N(0,1)$.

En particular, $\overline{X}_n \approx N\left(\mu, \frac{\sigma^2}{n}\right)$ para valores de n grandes.

Teorema del límite central

Ejemplo

¿Cuál es la probabilidad de que al lanzar una moneda 400 *veces, salga cara en menos de* 210 *ocasiones?*

Sea $X_i \sim \mathrm{Be}(0.5)$ *la V.A. que vale* 1 *si se obtiene cara en el lanzamiento i-ésimo y* 0 *en otro caso, por lo que* $E[X_i] = 0.5$ *y* $\mathrm{Var}[X_i] = 0.5^2 = 0.25$*. Si definimos* $S_{400} = \sum_{i=1}^{400} X_i$*, entonces se nos pide*

$$\begin{aligned} P(S_{400} < 210) &= P(S_{400} \leq 209) \\ &= P\left(\frac{S_{400} - 400 \cdot 0.5}{\sqrt{400} \cdot \sqrt{0.25}} \leq \frac{209 - 400 \cdot 0.5}{\sqrt{400} \cdot \sqrt{0.25}}\right) \\ &\approx P(Z < 0.9) \approx 0.8159, \end{aligned}$$

donde $Z \sim N(0, 1)$.

Universidad Internacional de Valencia

Teorema (Ley de grandes números)

Sea $\{X_i\}_{i=1}^{\infty}$ *una sucesión de variables aleatorias independientes cumpliendo* $E[X_i] = \mu$ *y* $\mathrm{Var}(X_i) = \sigma^2 \ \forall i \in \mathbb{N}$. *Si denotamos para cada* $n \in \mathbb{N}$

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

entonces

$$\lim_{n\to\infty} P(|\overline{X}_n - \mu| < \varepsilon) = 1, \quad \forall \varepsilon > 0.$$

¡Muchas gracias!

Universidad Internacional de Valencia

Contacto:

david.zorio@campusviu.es