

Procesos estocásticos

Enrique Miranda

Universidad of Oviedo

Máster Universitario en Análisis de Datos
para la Inteligencia de Negocios

Contenidos del curso

1. Introducción.
2. Procesos a tiempo discreto:
 - ▶ Cadenas de Markov a tiempo discreto.
 - ▶ Martingalas a tiempo discreto.
3. Procesos a tiempo continuo:
 - ▶ Procesos de Poisson.
 - ▶ Procesos de renovación.
 - ▶ Cadenas de Markov a tiempo continuo.
 - ▶ Procesos de nacimiento y muerte.
 - ▶ El movimiento browniano.

Sesión 1. Repaso de probabilidad

1. Resumen de teoría de probabilidad.

- ▶ Espacio probabilístico.
- ▶ Probabilidad condicionada.
- ▶ Independencia.

2. Variables aleatorias.

- ▶ Distribución de probabilidad inducida.
- ▶ Esperanza de una variable.
- ▶ Distribuciones discretas.
 - ▶ Binomial.
 - ▶ Poisson.
- ▶ Distribuciones continuas.
 - ▶ Normal.
 - ▶ Exponencial.

Introducción

Distinguimos entre experimentos **deterministas**, cuyo resultado puede predecirse y es siempre el mismo cuando el experimento se realiza en las mismas condiciones, y experimentos **aleatorios**, que son los que cumplen las siguientes condiciones:

- ▶ El resultado no puede predecirse antes de su realización.
- ▶ Los resultados posibles están bien definidos de antemano.

La probabilidad es una herramienta matemática con la que medimos la variabilidad del resultado de un experimento aleatorio.

Espacio muestral

Dado un experimento aleatorio, cada uno de los posibles resultados del mismo se denomina **suceso elemental**. El espacio de todos los sucesos elementales se denomina **espacio muestral**.

En un experimento aleatorio estamos interesados en calcular la probabilidad de subconjuntos del espacio muestral, a los que llamaremos **sucesos**. Entre ellos destacamos:

- ▶ El **suceso seguro** Ω , que contiene a todos los sucesos elementales.
- ▶ El **suceso imposible** $\emptyset$, que no contiene a ninguno.

El conjunto de sucesos sobre los que vamos a calcular la probabilidad se denotará $\mathcal{A}$.

Operaciones entre sucesos

- ▶ **Unión** de sucesos: $A \cup B := \{w \in A \text{ o } w \in B\}$.
- ▶ **Intersección** de sucesos: $A \cap B := \{w \in A \text{ y } w \in B\}$.
- ▶ **Diferencia** de sucesos: $A \setminus B := \{w \in A \text{ y } w \notin B\}$.
- ▶ **Complementario** de un suceso: $\bar{A} := \{w \notin A\}$.
- ▶ **Inclusión** de sucesos: $A \subset B$ si $w \in A \Rightarrow w \in B$.
- ▶ Sucesos **incompatibles** o **disjuntos**: A, B tales que $A \cap B = \emptyset$.

σ -álgebra de sucesos

Sea $\mathcal{A}$ una clase no vacía formada por subconjuntos del espacio muestral Ω . Diremos que es una σ -álgebra cuando satisface las siguientes propiedades:

- ▶ $A \in \mathcal{A} \Rightarrow \bar{A} \in \mathcal{A}$.
- ▶ $(A_n)_n \subseteq \mathcal{A} \Rightarrow \bigcup_n A_n \in \mathcal{A}$.

En nuestra definición de probabilidad $\mathcal{A}$ tendrá estructura de σ -álgebra, y se dirá σ -álgebra de sucesos del experimento aleatorio.

Definición axiomática de la probabilidad

La siguiente definición fue establecida por Kolmogorov en 1933.

Sea $\mathcal{A}$ la σ -álgebra de sucesos asociada a un experimento aleatorio. Una **probabilidad** es una función $P : \mathcal{A} \rightarrow \mathbb{R}$ que verifica los siguientes axiomas:

- ▶ $P(A) \geq 0$ para todo $A \in \mathcal{A}$.
- ▶ $P(\Omega) = 1$.
- ▶ Dados $(A_n)_n$ disjuntos dos a dos, se cumple

$$P(\cup_n A_n) = \sum_n P(A_n).$$

$(\Omega, \mathcal{A}, P)$ se dice **espacio de probabilidad**.

Propiedades de una probabilidad

- ▶ $A \subseteq B \Rightarrow P(A) \leq P(B)$.
- ▶ $P(\bar{A}) = 1 - P(A)$.
- ▶ $P(A \setminus B) = P(A) - P(A \cap B)$.
- ▶ $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
- ▶ $A_1 \subseteq A_2 \subseteq \dots \Rightarrow P(\cup_n A_n) = \lim_n P(A_n)$.

Ejemplo

En una universidad el 50% de los alumnos habla inglés, el 20% francés y el 5% los dos idiomas ¿Cuál es la probabilidad de encontrar alumnos que hablen alguno de los dos idiomas?

Solución: Sea A el suceso hablar inglés: $P(A) = 0.5$. Sea B el suceso hablar francés: $P(B) = 0.2$ y sea $A \cap B$ el suceso hablar francés e inglés: $P(A \cap B) = 0.05$. Así,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.5 + 0.2 - 0.05 = 0.65.$$

Definición de probabilidad condicionada

Sea $B \subset \Omega$ un suceso aleatorio de probabilidad no nula, $P(B) > 0$. Para cualquier otro suceso A se define la **probabilidad condicionada de A dado B** y se denota $P(A|B)$ al valor

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

La probabilidad de A condicionada por B representa la probabilidad de que ocurra A, cuando sabemos que ha ocurrido B. $P(\cdot|B)$ es una medida de probabilidad, con las propiedades que señalamos anteriormente.

Ejemplo

Es importante no confundir $P(A|B)$ con $P(B|A)$, lo que se conoce como la **falacia de la acusación**: es habitual que en los juicios la acusación presente $P(\text{Inocencia}|\text{Hechos})$ como algo muy pequeño cuando lo que se conoce es $P(\text{Hechos}|\text{Inocencia})$.

También es importante no confundir $P(A|B)$ con $P(A \cap B)$: mientras que ésta es necesariamente menor que $P(A)$, $P(A|B)$ puede ser menor, igual o mayor que $P(A)$.

Regla del producto

Como consecuencia de la definición anterior, se cumple

$$P(A \cap B) = P(A)P(B|A) = P(B)P(A|B).$$

Una extensión de esta idea para una cantidad finita de sucesos es la llamada **regla del producto**: dados sucesos $A_1, \dots, A_n$ para los que $P(A_1 \cap \dots \cap A_{n-1}) > 0$, se cumple

$$P(A_1 \cap \dots \cap A_n) = P(A_1)P(A_2|A_1) \dots P(A_n|A_1 \cap \dots \cap A_{n-1}).$$

Independencia de sucesos

Decimos que dos sucesos A y B son **independientes** cuando la ocurrencia de uno de ellos no aporta información acerca de la posible ocurrencia del otro.

Matemáticamente, A y B son independientes cuando se cumple cualquiera de las siguientes condiciones equivalentes:

- ▶ $P(A \cap B) = P(A)P(B)$;
- ▶ $P(B) > 0$ y $P(A|B) = P(A)$;
- ▶ $P(A) > 0$ y $P(B|A) = P(B)$.

Independencia de varios sucesos

El concepto anterior se extiende para cualquier cantidad finita de sucesos. Se dice que $A_1, \dots, A_n$ son independientes cuando para cualesquiera $i_1, \dots, i_k$ en $\{1, \dots, n\}$, se cumple

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \dots P(A_{i_k}).$$

Nótese que este concepto no es equivalente a la independencia dos a dos entre los sucesos $A_1, \dots, A_n$.

Concepto de variable aleatoria

Consideremos un espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Una variable aleatoria es una aplicación X definida en Ω , que a cada individuo $\omega \in \Omega$ le asigna un valor $X(\omega)$ de interés:

- ▶ Consumo de un cierto producto.
- ▶ Ganancia obtenida en un juego de azar.
- ▶ Número de llegadas a una cola en un día determinado.
- ▶ ...

En particular, una función constante también es una variable aleatoria.

Ejemplo

Lanzamos una moneda dos veces y observamos el resultado obtenido. Puede que no estemos interesado en la probabilidad de un suceso cualquiera (cx, por ejemplo), sino en la probabilidad de conseguir un número determinado de caras (0, 1, 2 caras). Entonces, tendríamos que definir una aplicación:

$$X : \Omega \rightarrow \mathbb{R}$$

$$\omega \rightarrow X(\omega) = \text{número de caras} .$$

Así, $X((c, c)) = 2$, $X((c, x)) = X((x, c)) = 1$ y $X((x, x)) = 0$.

Se cumple entonces

$$P(X = 1) = P(\{(c, x), (x, c)\}) = P((c, x)) + P((x, c)) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.$$

Parámetros de una variable aleatoria

Dada una variable aleatoria X , su probabilidad P_X contiene toda la información probabilística acerca de X . Sin embargo, podemos querer resumir esta información a través de algunos parámetros.

Así, podemos considerar la esperanza, varianza, etc.

Estos parámetros son análogos a los de estadística descriptiva, pero utilizando, en lugar de la frecuencia relativa de los distintos valores, su probabilidad.

Tipos de variables aleatorias

- ▶ Variables **discretas**: son aquéllas que pueden tomar una cantidad finita o infinita numerable de valores distintos. Su probabilidad P_X viene caracterizada por la función f_X dada por $f_X(x) = P_X(\{x\})$.
- ▶ Variables **continuas**: son aquéllas que pueden tomar una cantidad infinita no numerable de valores distintos, y para las que $P(X = x) = 0$ para todo x .

Función de masa de probabilidad

Cuando $\{X(w) : w \in \Omega\} = \{x_1, x_2, \dots\}$ es un conjunto finito o infinito numerable, decimos que la variable aleatoria X es discreta. Si teníamos sobre el espacio muestral Ω la probabilidad P , entonces la probabilidad inducida por X será

$$f_X(x) = \begin{cases} P(X = x_i) & \text{si } x = x_i \text{ para } i = 1, 2, \dots \\ 0 & \text{en el resto.} \end{cases}$$

Esta función se denomina **función de masa de probabilidad**.

Cálculo de probabilidades de sucesos

A partir de la función de masa de probabilidad podemos calcular la probabilidad de que X tome valores en cualquier subconjunto A del conjunto de resultados:

$$P(X \in A) = \sum_{x \in A} P(X = x) = \sum_{x \in A} f_X(x).$$

En particular, $\sum_{x \in \Omega} f_X(x) = 1$.

Esperanza y varianza de una variable discreta

Se define la **esperanza** de una variable discreta con valores en un conjunto numerable $\{x_n\}_n$ como

$$E(X) = \sum_n x_n f_X(x_n),$$

La noción de esperanza puede aplicarse a cualquier transformación $g(X)$ de una variable X :

$$E(g(X)) = \sum_n g(x_n) f_X(x_n).$$

Dadas dos variables X, Y y constantes $a, b \in \mathbb{R}$, se cumple que $E(aX + bY) = aE(X) + bE(Y)$.

Por otro lado, la **varianza** de X viene dada por

$$\text{Var}(X) = \sum_n (x_n - E(X))^2 f_X(x_n) = E(X^2) - E(X)^2.$$

Vectores aleatorios discretos

Un **vector aleatorio discreto** (X_1, X_2) representa el comportamiento simultáneo de 2 variables aleatorias discretas X_1, X_2 sobre un mismo espacio. Viene caracterizada por la función de masa de probabilidad **conjunta**

$$f_{(X_1, X_2)}(x_1, x_2) = P(X_1 = x_1, X_2 = x_2).$$

Dados conjuntos A_1, A_2 , se cumple

$$P(X_1 \in A_1, X_2 \in A_2) = \sum_{x_1 \in A_1} \sum_{x_2 \in A_2} P(X_1 = x_1, X_2 = x_2).$$

Análogamente se puede modelizar la observación simultánea de k variables aleatorias.

Independencia

Se dice que las variables X_1, X_2 son **independientes** cuando

$$f_{(X_1, X_2)}(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2) \quad \forall x_1, x_2.$$

Como consecuencia, si X_1 y X_2 son independientes, también lo son los sucesos $(X_1 \in A_1)$ y $(X_2 \in A_2)$ para cualesquiera subconjuntos A_1, A_2 de $\mathbb{R}$.

Procesos de Bernoulli

Si consideramos repeticiones (finitas o infinitas) de un ensayo con sólo dos resultados posibles, tal experimento será un **proceso de Bernoulli** cuando:

- ▶ Las repeticiones son **independientes**: la probabilidad de éxito en la repetición n -ésima no viene influida por los resultados anteriores.
- ▶ La probabilidad de éxito en cada repetición es la misma (p).

Distribución Binomial

En un proceso de Bernoulli de n repeticiones, nos interesa estudiar el número de éxitos, así si definimos la variable aleatoria:

X = número de éxitos obtenidos en esas n pruebas,

decimos entonces que X sigue una **distribución binomial** de parámetros n y p , la denotamos $X \sim B(n, p)$.

Características

La función de masa de probabilidad de una binomial es

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

donde $x = 0, 1, \dots, n$ y $\binom{n}{x} = \frac{n!}{x!(n-x)!} = \frac{n(n-1)\dots(n-x+1)}{x(k-1)\dots 1}$.

La media y la varianza de una $B(n, p)$ son respectivamente

$$\begin{aligned} E[X] &= np \\ \text{Var}[X] &= np(1 - p). \end{aligned}$$

Distribución de Poisson

La **distribución de Poisson** se usa habitualmente en experimentos en los cuales estamos interesados en contar el número de ocurrencias de un suceso para un intervalo fijo de tiempo, cuando se cumple que *la probabilidad de ocurrencia de un suceso en un intervalo pequeño de tiempo es proporcional a la longitud del intervalo*.

Ejemplos:

- ▶ Número de reclamaciones semanales a un servicio.
- ▶ Número de saltos en el precio de mercado de un producto en un intervalo de tiempo.
- ▶ Personas que llegan a un mostrador en una hora.

Características

La distribución de Poisson tiene un único parámetro λ , también llamado **parámetro de intensidad**. La función de masa de una $P(\lambda)$ es:

$$P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!} \quad x = 0, 1, \dots$$

Su media y su varianza son

$$\begin{aligned} E[X] &= \lambda \\ \text{Var}[X] &= \lambda. \end{aligned}$$

λ se interpreta como tasa de llegadas por unidad de tiempo.

Variables aleatorias continuas

Las variables aleatorias continuas aparecen cuando se estudian características numéricas que pueden tomar una cantidad infinita no numerables de valores (todos los de la recta real o los de un intervalo de la misma).

En ellas, la probabilidad se acumula suavemente a lo largo del campo de variación y no sólo en unos cuantos puntos.

Función de densidad. Definición

Una **función de densidad** es una aplicación $f : \mathbb{R} \rightarrow \mathbb{R}$ que cumple las siguientes dos propiedades:

- ▶ $f(x) \geq 0$ para todo $x \in \mathbb{R}$.
- ▶ $\int_{-\infty}^{\infty} f(x)dx = 1$.

Podemos ver a la función de densidad como una generalización de los histogramas para variables aleatorias continuas.

Función de densidad. Propiedades

- ▶ Dado un conjunto medible A , $P(X \in A) = \int_A f(x)dx$.
- ▶ $f(x)$ NO es la probabilidad de que X tome el valor x :
 - ▶ En una variable continua la probabilidad de tomar cualquier valor concreto es 0.
 - ▶ $f(x)$ puede ser mayor que 1.
- ▶ $f(x)$ NO tiene por qué ser una función continua, pero lo será salvo en una cantidad finita de puntos.

Esperanza y varianza de una variable continua

Dada una variable aleatoria continua X con función de densidad f_X , se definen:

$$\begin{aligned}E(X) &= \int_{-\infty}^{\infty} xf_X(x)dx \\ \text{Var}(X) &= \int_{-\infty}^{\infty} (x - E(X))^2 f_X(x)dx \\ &= \int_{-\infty}^{\infty} x^2 f_X(x)dx - E(X)^2.\end{aligned}$$

De nuevo se cumple $E(aX + bY) = aE(X) + bE(Y)$ para todo par de variables aleatorias X, Y y todo $a, b \in \mathbb{R}$.

Vectores aleatorios continuos

Un vector aleatorio continuo (X_1, X_2) representa el comportamiento simultáneo de 2 variables aleatorias continuas X_1, X_2 . Viene caracterizado por la **función de densidad conjunta** $f_{(X_1, X_2)}(x_1, x_2)$, para $x_1, x_2 \in \mathbb{R}$.

- ▶ $f_{(X_1, X_2)}(x_1, x_2) \geq 0 \quad \forall x_1, x_2$.
- ▶ $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{(X_1, X_2)}(x_1, x_2) dx_1 dx_2 = 1$.

Análogamente se modelizaría la observación simultánea de k variables aleatorias continuas.

Independencia

Se dice que las variables X_1, X_2 son **independientes** cuando

$$f_{(X_1, X_2)}(x_1, x_2) = f_{X_1}(x_1) \cdot f_{X_2}(x_2) \quad \forall x_1, x_2.$$

Como consecuencia, si X_1 y X_2 son independientes, también lo son los sucesos $(X_1 \in A_1)$ y $(X_2 \in A_2)$ para cualesquiera subconjuntos A_1, A_2 de $\mathbb{R}$.

Conjuntos de variables independientes

Dadas n variables aleatorias $X_1, \dots, X_n$, se dice que son **independientes** cuando su función de distribución conjunta es igual al producto de las marginales, o equivalentemente si para cualesquiera subconjuntos $A_1, \dots, A_n$ de $\mathbb{R}$,

$$P(X_1 \in A_1, \dots, X_n \in A_n) = \prod_{i=1}^n P(X_i \in A_i).$$

Esta condición implica (pero NO es equivalente a) que todo par de variables X_i, X_j son independientes.

Como consecuencia, se dice que las variables de una sucesión $\{X_n\}_n$ son independientes cuando las variables de cualquier subconjunto finito lo son.

Distribución exponencial

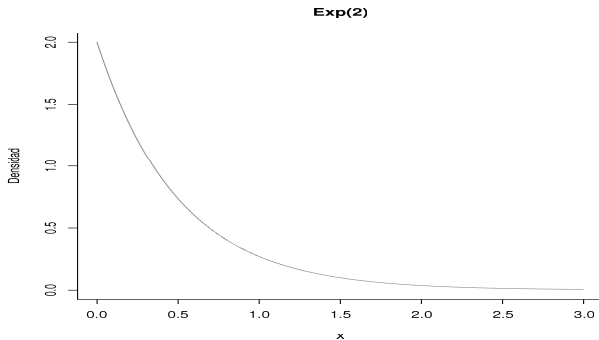
Se dice que $X \sim \exp(\alpha)$ cuando su función de densidad es

$$f(x) = \begin{cases} \alpha e^{-\alpha x} & \text{si } x > 0 \\ 0 & \text{en otro caso,} \end{cases}$$

donde $\alpha > 0$.

Ejemplo

A continuación representamos la función de densidad de la distribución $\text{exp}(2)$:



Cuanto mayor sea el valor del parámetro α , más rápidamente decae la función de densidad.

Algunos aspectos de la distribución exponencial

- ▶ Los principales momentos de una distribución $\exp(\alpha)$ son

$$E(X) = \frac{1}{\alpha}, \text{Var}(X) = \frac{1}{\alpha^2}.$$

- ▶ La principal propiedad de esta distribución es la **falta de memoria**: se cumple

$$P(X \geq t + h | X \geq t) = P(X \geq h) \quad \forall t, h > 0.$$

Relación con la distribución de Poisson

Si el número de llegadas a un servicio en un intervalo de tiempo t sigue una distribución de Poisson $\mathcal{P}(\lambda)$, entonces el tiempo entre dos llegadas consecutivas (tomando como unidad t) sigue una distribución exponencial $\exp(\lambda)$.

Si queremos usar otras unidades de tiempo, basta con tener en cuenta que si las llegadas en t siguen una Poisson $\mathcal{P}(\lambda)$, entonces las llegadas en kt siguen una distribución $\mathcal{P}(k\lambda)$.

La correspondencia anterior es la base de los llamados **procesos de Poisson**, que veremos más adelante.

La distribución normal

Decimos que una variable aleatoria X tiene una **distribución normal** $X \sim N(\mu, \sigma)$ cuando su función de densidad es

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

para $x \in \mathbb{R}$.

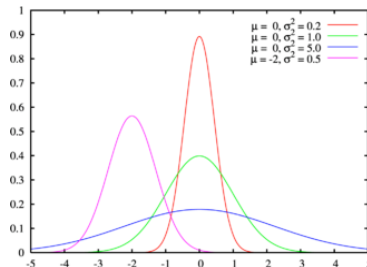
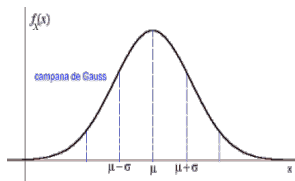
► $E(X) = \mu, \text{Var}(X) = \sigma^2.$

La distribución normal **estándar** se corresponde con $\mu = 0, \sigma = 1.$

Ejemplo

La función de densidad para distintos valores de μ, σ viene dada por:

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



El Teorema Central del Límite

Sea $X_1, \dots, X_n$ una sucesión de variables aleatorias independientes e idénticamente distribuidas, con $E(X_i) = \mu$ y $Var(X_i) = \sigma^2 < +\infty$. Entonces,

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

es una variable aleatoria cuya función de densidad se aproxima a la distribución $N(0, 1)$ cuando n es grande.

Suele considerarse aplicable la aproximación cuando $n > 30$.