Probabilidad

Clase 13 Curso Propedéutico
2017/07/04

ITAM

Propiedades de Esperanza y Varianza

Esperanza $\mathbb{E}(X) =$ "Promedio ponderado de posibles valores de X"

$$= \begin{cases} \sum_x x\,\mathbb{P}(X=x) \\ \int_{\mathbb{R}} x f_X(x)\,dx \end{cases} = \int_{\mathbb{R}} x\,dF_X(x)$$

Propiedades: X, Y v.a $a \in \mathbb{R}$

(1) $\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$

y $\mathbb{E}[aX] = a\,\mathbb{E}(X)$

Dem: $\mathbb{E}(X+Y) = \sum_z z\,\mathbb{P}(X+Y=z)$

Marginalización

$$= \sum_z \sum_y z\,\mathbb{P}(X+Y=z, Y=y)$$

conjunta

$x = z-k$ $\quad z = x+k$

$$= \sum_z z \sum_y \mathbb{P}(X=z-k, Y=k) = \sum_x \sum_y (x+y)\,\mathbb{P}(X=x, Y=y)$$

$$= \sum_x \sum_y x\, \mathbb{P}(X=x, Y=y) + \sum_x \sum_y y\, \mathbb{P}(X=x, Y=y)$$

$$= \sum_x x \left(\underbrace{\sum_y \mathbb{P}(X=x, Y=y)}\right) + \sum_y y \left(\sum_x \mathbb{P}(X=x, Y=y) \right)$$

$$\overset{\text{Marg.}}{=} \sum_x x\, \mathbb{P}(X=x) + \sum_y y\, \mathbb{P}(Y=y) = \mathbb{E}(X) + \mathbb{E}(Y)$$

·) $\mathbb{E}(aX) = \sum_k k\, \mathbb{P}(aX=k) = \sum_x a x\, \mathbb{P}(X=x) = a\, \mathbb{E}(X)$.

Más general

② $$\mathbb{E}[h(X)] = \begin{cases} \int h(x) f_X(x)\, dx \\ \sum_x h(x)\, \mathbb{P}(X=x) \end{cases}$$

$h: \mathbb{R} \to \mathbb{R}$

Ejemplos

$h(X,Y) = X+Y$

$h(X) = aX$

De hecho, en vectores aleatorios

$$\mathbb{E}[h(X_1, \dots, X_n)] = \begin{cases} \int \cdots \int h(x_1, \dots, x_n) f_{X_1, \dots, X_n}(x_1, \dots, x_n)\, dx_1 \cdots dx_n \\ \sum_{x_1} \cdots \sum_{x_n} h(x_1, \dots, x_n)\, \mathbb{P}(X_1 = x_1, \dots, X_n = x_n) \end{cases}$$

$h: \mathbb{R}^n \to \mathbb{R}$

... entonces la esperanza preserva combinaciones lineales $\mathbb{E}(\sum \alpha X) = \sum \alpha \mathbb{E}(X)$

¿Qué pasa con la varianza?

③ $\cdot \mathrm{Var}(aX) = a^2 \mathrm{Var}(X)$

Dem

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

$\cdot)$ $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y)$

Dem:

$(a+b)^2 = a^2 + b^2 + 2ab$

$$\mathrm{Var}(X+Y) = \mathbb{E}\left((X+Y - \mathbb{E}(X+Y))^2\right)$$
$$= \mathbb{E}\left(\left([X - \mathbb{E}(X)] + [Y - \mathbb{E}(Y)]\right)^2\right)$$
$$= \mathbb{E}\left((X - \mathbb{E}(X))^2 + (Y - \mathbb{E}(Y))^2 + 2(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\right)$$
$$= \underbrace{\mathbb{E}\left((X - \mathbb{E}(X))^2\right)}_{\mathrm{Var}(X)} + \underbrace{\mathbb{E}\left((Y - \mathbb{E}(Y))^2\right)}_{\mathrm{Var}(Y)} + 2\,\underbrace{\mathbb{E}\left((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\right)}_{\mathrm{Cov}(X,Y)}$$

(7) Sea X, Y v.a $\Sigma = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{xy} & \sigma_{yy} \end{pmatrix}$

$\sigma_{xx} = Var(X) = Cov(X, X)$
$\sigma_{xy} = Cov(X, Y)$
$\sigma_{yy} = Var(Y) = Cov(Y, Y)$

Σ se llama matriz de covarianzas

$$\begin{aligned} Var(aX + bY) &= a^2 Var(X) + b^2 Var(Y) + 2ab\, Var(X, Y) \\ &= a^2 \sigma_{xx} + b^2 \sigma_{yy} + 2ab\, \sigma_{xy} \\ &= [a, b] \begin{bmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{xy} & \sigma_{yy} \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} \\ &= w^T \Sigma w \quad \text{con } w = \begin{pmatrix} a \\ b \end{pmatrix} \end{aligned}$$

De manera más general

Si $X_1, \ldots, X_n$ son v.a.s y $\Sigma = [Cov(X_i, X_j)]_{i,j=1}^{n}$

y $a = (a_1, \ldots, a_n)$ vector de coeficientes, entonces

$$\boxed{Var\left(\sum a_i X_i\right) = a^T \Sigma a}$$

Ejemplo aplicación

- Finanzas μ vector de rendimientos esperados de activos.

  Σ matriz de covarianzas

Problema de Optimización de portafolios

$$\max \quad w^T\mu \quad \leftarrow \text{rendimiento de un portafolio}$$

$$\text{s.a.} \quad w^T\Sigma w = c \quad \leftarrow \text{sujeto a un nivel de riesgo constante}$$

- Reducción de dimensionalidad [Ver con más detalle después]

Buscar el eigenvector de eigenvalor más grande

Dadas $X_1,\dots,X_n$ cuál será la combinación lineal que preserve mayor información (en términos de varianza)

$$\max \ \operatorname{Var}\left(\sum a_i X_i\right) \quad \text{s.a.} \ \|a\|=1 \quad \sim \quad \max \ a^T\Sigma a \quad \text{s.a.} \ a^Ta-1=0$$

el eigenvalor o la varianza

Solución es a eigenvector de Σ. (por lagrangeanos)

i.e. $\boxed{\Sigma a=\lambda a}$ y $\operatorname{Var}\left(\sum_i a_i X_i\right) = a^T \overbrace{\Sigma a}^{\lambda a} = \lambda(a^Ta) = \lambda$

Procesos $\boxed{\text{Estocásticos}}$ = Aleatorio

Estocásticos es un sinónimo de aleatorio.

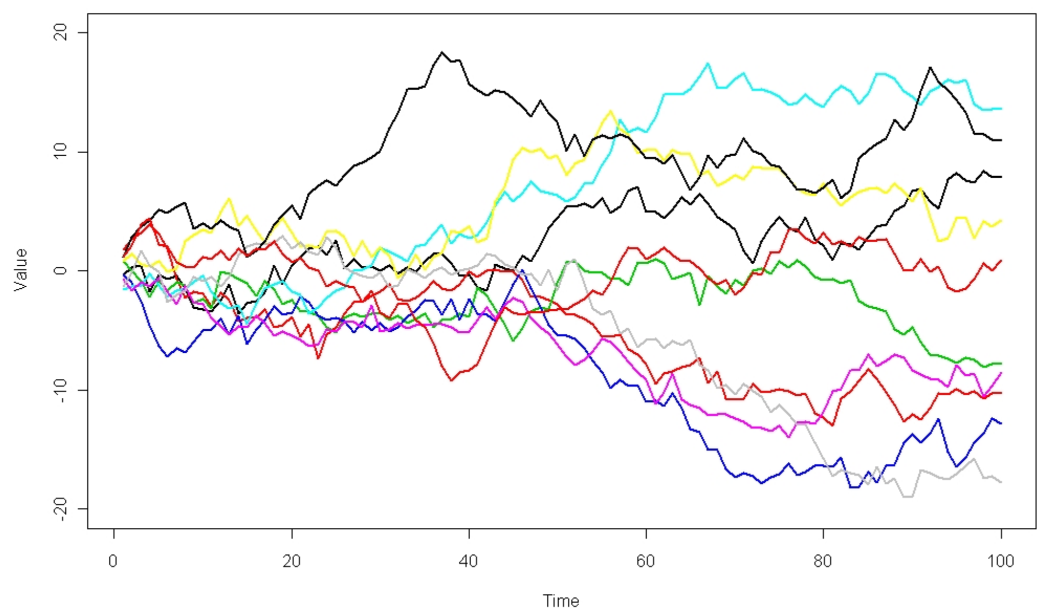
En vez de tener una sola variable aleatoria X, queremos estudiar el comportamiento de una *secuencia* de variables aleatorias $(X_t)_{t\geq 0} = X_1, X_2, \ldots$.

Cada variable aleatoria representa un momento en el tiempo. Las variables no tienes porque ser independientes.

Ejemplos:

P. Browniano

- El precio de una acción.
- La posición de una partícula de polen suspendida en el aire.
- El número de clientes que hay en una tienda P. Poisson
- El estado del clima



Independencia

$P(A|B)=P(A)$

Independencia condicional a C

$P(A|B,C)=P(A|C)$

Cadenas de Markov

- El estado próximo depende sólo del estado actual.
- Más precisamente, las *probabilidades de transición* están determinadas por el estado actual.

$$P(X_{t+1}=x_{t+1} \mid X_0=x_0,\dots,X_t=x_t) = P(X_{t+1}=x_{t+1} \mid X_t=x_t)$$

El futuro es independiente del pasado condicional al presente

X_{t+1} $X_0,\dots,X_{t-1}$ X_t

- **Motivación inicial:** un modelo para estudiar el lenguaje.

Si cada X_t es una palabra

a) Predecir la próxima palabra de un texto

¿Es Markoviano?

b) ¿Si cada X_t son las últimas 5 palabras?

A. A. Марков (1886).

Inspiración Física

estado de un sistema dinámico

$$s_{t+1} = f(s_t)$$

Markov

componente determinista

componente aleatorio

$$s_{t+1} = f(s_t, \varepsilon_t)$$

Nota: La propiedad de Markov es un supuesto aparentemente fuerte pero muchas cosas se pueden "modelar" en la práctica Markovianamente:

.) Incluyendo información adicional relevante

.) Haciendo trampa e incluyendo "lags".

- Para conocer las probabilidades del futuro solo necesitas conocer el presente!

Matemáticamente…

- Si S son todos los posibles "estados" de la cadena (los valores que pueden tomar las X_t) **la propiedad de Markov dice:**

(homogénea)

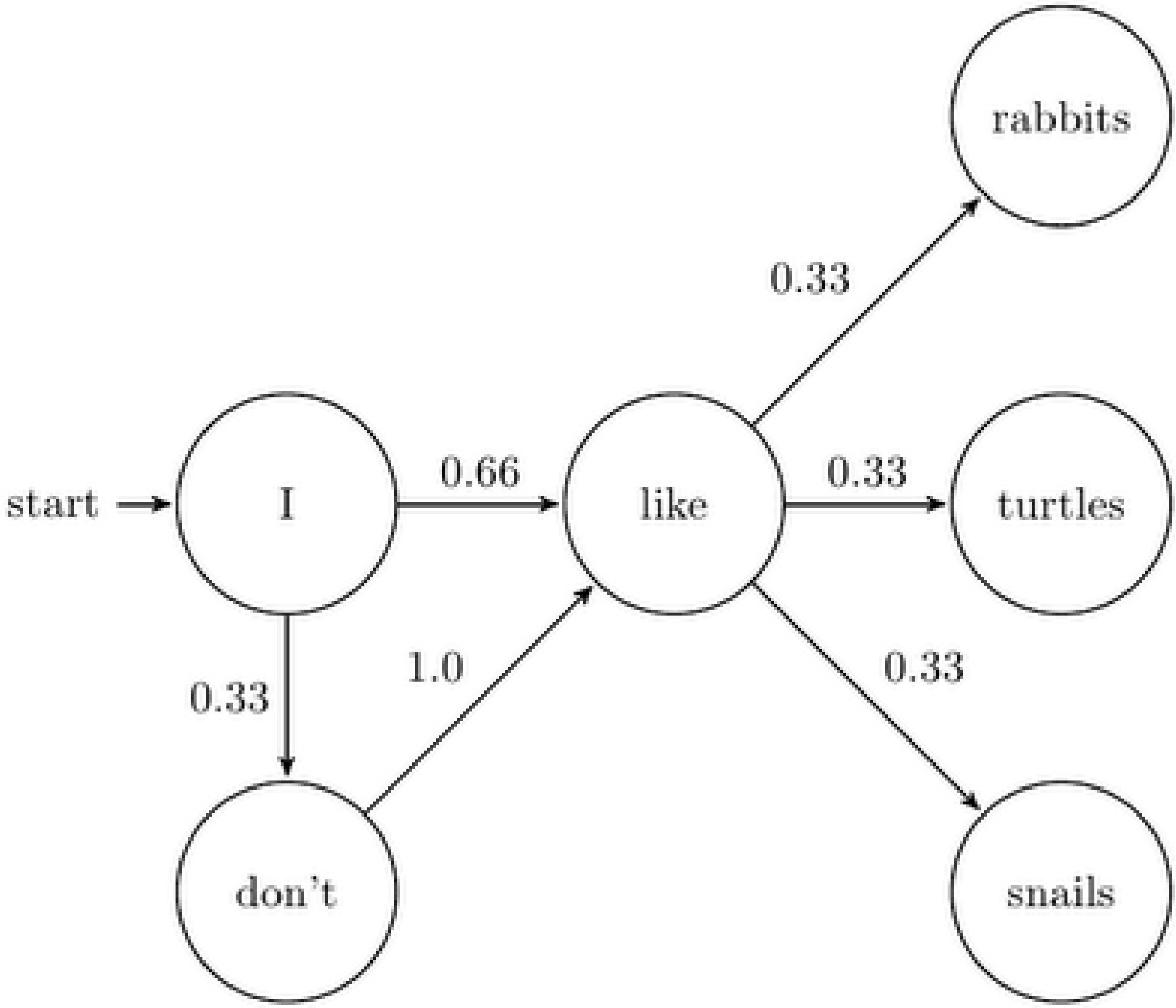
Que es una cadena de Markov

- Es tres cosas:

1. Un conjunto de estados S
2. Una distribución inicial para X_0
3. Un "Kernel" de transición. Unas probabilidades de transición por cada estado actual.

S finito…. El kernel puede representarse como matriz $P = (P_{ij})$

$$p_{ij} = \mathbb{P}(X_{t+1} = j \mid X_t = i)$$


start
I
0.66
like
0.33
don't
1.0
0.33
rabbits
0.33
turtles
0.33
snails

Distribución de las variables X_t

- ¿Cómo son las distribuciones de X_t condicionando a la distribución inicial de X_0?

Aplicación de las cadenas de Markov

- Hay dos aplicaciones principales

1. Aprender cómo es P para poder predecir
2. Ya sabes como es P pero cómo se comportará un sistema en el largo plazo.

A continuación vamos a hablar del comportamiento largo plazo de las cadenas de Markov y vamos a dar uno de los ejemplos de aplicación más famosos… (Google PageRank).

Distribución límite/estacionaria