Universidad Nacional de Rosario

Facultad de Ciencias Exactas, Ingeniería y Agrimensura

Probabilidad y Estadística

Unidad 7

Autor del resumen:

Charles Chaplin

Julio 2020

Chapter 1

Cadenas de Markov : Primeros pasos

1.1 Introduccion

Consideren un juego con un tablero que se base en 10 casillas cuadradas (numeradas 1-10) dispuestas en circulo. (Un Monopoly en miniatura.) Un jugador empieza en la casilla 1. A cada turno, el jugador tira un dado y se mueve alrededor del tablero las cantidad casillas que aparezca en la cara del dado. El jugador se sigue moviendo alrededor de las casillas de acuerdo a la tirada de dado (Esta garantizado que este no es un juego muy exitante...)

Ahora bien, sea X_k el numero de la casilla en el cual el jugador se encuentra luego de k movimientos, con $X_0 = 1$. Supongamos que el jugador obtiene las siguientes tiradas 2,1 y 4. Las primeras cuatro posiciones son:

$$(X_0, X_1, X_2, X_3) = (1, 3, 4, 8).$$

Dada esta informacion, que podemos decir acerca de el proximo casillero (X_4) que ocupara el jugador? A pesar de que sabemos el historial completo de tiradas del jugador, la unica informacion relevante para predecir su posicion en el futuro es la ubicacion mas reciente (en este caso X_3). Ya que $X_3 = 8$ entonces necesariamente $X_4 \in A = \{9, 10, 1, 2, 3, 4\}$ y cada resultable posible tiene la misma probabilidad. Formalmente

$$\forall_{j \in A}\ P(X_4 = j | X_0 = 1, X_1 = 3, X_2 = 4, X_3 = 8) = P(X_4 = j | X_3 = 8) = \frac{1}{6}.$$

Dada la posicion mas actual del jugador X_3, su futura posicion X_4 es independiente del pasado de la historia X_0, X_1, X_2.

La secuencia de posiciones del jugador $X_0, X_1, X_2, ...$ es un proceso estocastico llamado **Cadena de Markov**. Este juego ilustra una muy importante propiedad de la cadena de Markov: **El futuro, dado el presente, es independiente del pasado**

Cadena de Markov. Sea S un conjunto discreto. Una cadena de Markov es una secuencia de variables aleatorias $X_0, X_1, ...$, cuyos espacios muestrales son S con la siguiente propiedad: Dado $n \geq 0$

$$\forall_{i,j \in S}\ P(X_{n+1} = j | X_0 = x_0, ..., X_{n-1} = x_{n-1}, X_n = i) = P(X_{n+1} = j | X_n = i)$$

El conjunto S es el **Espacio de Estados** de la cadena de Markov. Generalmente nos referimos a $X_n = i$ como que la cadena visito la posicion i en n tiempos.

Una cadena de Markov es **homogenea en el tiempo** si las probabilidades mencionadas anteriormente no dependen de n. Esto es

$$\forall_{n\geq 0}\ P(X_{n+1}=j|X_n=i)=P(X_1=j|X_0=i)$$

Dado que las probabilidades anteriores dependen solamente de i y j, pueden ser almacenadas en una matriz **P** cuya ij-esima entrada corresponde a $P_{ij}=P(X_1=j|X_0=i)$. A esta matriz se la llama **matriz de transicion** o **matriz de Markov**, que contiene las probabilidades de moverse de un estado actual a cualquier otro (lo que se conoce como "probabilidad de un paso"). Si el espacio de estados tiene k elementos, entonces la matriz de transicion tiene $k\times k$. Si el espacio de estados es infinito contable, la matriz de transicion es infinita.

Las entradas de todas las matrices de transicion de Markov son no-negativas y cada fila suma 1,

$$\forall_{fila\ i}\ \sum_j P_{ij}=\sum_j P(X_1=j|X_0=i)=\sum_j \frac{P(X_1=j,X_0=i)}{P(X_0=i)}=\frac{P(X_0=i)}{P(X_0=i)}=1$$

Ejemplo p.42 Dobrow

Matriz Estocastica Una matriz estocastica es una matriz cuadrada, que satisface:

1. $\forall_{ij}\ P_{ij}\geq 0$
2. Para cada fila i, $\sum_j Pij=1$.

Ejemplos p.42-52

1.2 Calculos basicos

Una poderosa caracteristica de las Cadenas de Markov es la habilidad de usar matrices para computar probabilidades. Para usar los metodos de matriz, consideramos a las distribuciones de probabilidades como vectores. Un vector de probabilidad es un vector de numeros no-negativos cuya suma es igual a 1. Generalmente se denotan con letras griegas en negrita, como por ejemplo $\alpha,\ \lambda, y\ \pi$.

Supongamos que X es una variable aleatoria discreta con P(X=j) = α_j, para j = 1,2,... .Luego α=($\alpha_1,\alpha_2,...$) es un vector de probabilidad. Decimos que la distribucion de X es α.

Para el calculo matricial identificaremos distribucion de probabilidades discretas como vectores fila.

Para una cadena de Markov $X_0,X_1,...$, la distribucion de X_0 es llamada **distribucion inicial de la cadena de Markov**. Si α es la distribucion inicial, entonces P(X_0=j)=α_j, para toda j.

1.2.1 Probabilidad de transición en n pasos

Para los estados i y j, con n≥1, $P(X_n = j|X_0 = i)$ es la probabilidad de que la cadena que comenzo en i llegue a j en n pasos. La probabilidad de transición en n pasos puede ser acomodada en una matriz. La matriz cuya ij-esima entrada es $P(X_n = j|X_0 = i)$ es la matriz de transicion en n pasos de la cadena de Markov. Por supuesto, para n=1 es la usual matriz de transicion P. Para $n \geq 1$, uno de los resultados centrales del calculo de cadenas de Markov es que la matriz de transicion en n pasos es precisamente P^n, la enesima potencia de P.

D/ p.53

Matriz de transicion en n pasos. Sea $X_0, X_1, ...$, una cadena de Markov con una matriz de transicion **P**. La matriz P^n es la matriz de transicion en n pasos de la cadena. Para $n \geq 0$,

$$\forall_{ij}\ P^n_{ij} = P(X_n = j|X_0 = i)$$

Observacion. Note que $P^n_{ij} = (P^n)_{ij}$. No hay que confundirse con $(P_i j)^n$. Tambien note que P^0 es la matriz identidad. Esto es

$$P^0_{ij} = P(X_0 = j|X_0 = i) = \begin{cases} 1 & si \quad i = j \\ 0 & si \quad i \neq j \end{cases}$$

Ejemplos p.54-55

1.2.2 Chapman–Kolmogorov Relationship

Para m,n ≥0, la identidad matricial $P^{n+m} = P^m P^n$ nos da como resultado que

$$\forall_{i,j}\ P^{n+m}_{ij} = \sum_k P^m_{ik} P^n_{kj}$$

D. p.55

La interpretacion probabilistica es que realizando una transicion desde i hacia j en m+n pasos es equivalente a transicionar desde i a algun estado k en m pasos y despues transicionar desde este ultimo estado a j en los restantes n pasos. Esto es conocido como **la relacion Chapman-Kolmogorov**

1.2.3 Distribucion de X_n

En general, una cadena de Markov $X_0, X_1, ...,$ no es una secuencia de variables aleatorias igualmente distribuidas. Para n ≥ 1, la distribucion marginal de X_n depende en el n-esimo paso con su matriz de transicion P^n, asi como de la distribucion inicial α. Para obtener la funcion de probabilidad de X_n, dado que ocurre X_0,

$$\forall_j \ P(X_n = j) = \sum_i P(X_n = j|X_0 = i)P(X_0 = i) = \sum_i P^n_{ij}\alpha_i.$$

La suma anterior puede ser interpretada como el producto punto del vector de probabilidad inicial α con la j-esima columna de p^n. Es decir, es el j-esimo componente del vector-matriz resultante del producto αP^n

Ejemplo p.56

1.2.4 Presente, futuro y pasado mas reciente

La propiedad de Markov dice que el pasado y el futuro son independientes dado el presente. Tambien es verdad que el pasado y el futuro son independientes, dado el pasado mas reciente

Propiedad de Markov

Sea $X_0, X_1, ...$ una cadena de Markov. Luego, para toda $m < n$,

$$P(X_{n+1} = j|X_0 = i_0, ..., X_{n-m-1} = i_{n-m-1}, X_{n-m} = i)$$

$$= P(X_{n+1} = j|X_{n-m} = i) = P(X_{m+1} = j|X_0 = i) = P^{m+1}_{ij}$$

Para toda $i, j, i_0, ..., i_{m-n-1}, n \geq 0$

D. p.57

1.2.5 Distribucion conjunta

La distribucion marginal de una cadena de Markov esta determinada por la distribucion inicial α y la matriz de transicion P. De hecho, α y P determinan todas las distribuciones conjuntas de una cadena de Markov, esto es, la probabiludad conjunta de cualquier subconjunto finito de $X_0, X_1, ...$ En ese sentido, la distribucion inicail y la matriz de transicion dan una completa descripcion probabilistica de una cadena de Markov

Para ilustrar consideremos una probabilidad conjunta arbitraria como:

$$P(X_5 = i, X_6 = j, X_9 = k, X_{17} = l), \ \textit{para algunos estados } i, j, k, l.$$

Para el evento previamente mencionado, la cadena se mueve a i en 5 pasos, luego a j en un paso, luego a k en tres pasos y luego a l en 8 pasos. Con una distribucion inicial α, la intuicion nos dice que

$$P(X_5 = i, X_6 = j, X_9 = k, X_{17} = l) = (\alpha P^5)_i P_{ij} P^3_{jk} P^8_{kl}.$$

Y de hecho, la probabilidad condicional, la propiedad de Markov y la homogeniedad del tiempo nos dan

$$\begin{aligned}
& P(X_5 = i, X_6 = j, X_9 = k, X_17 = l) \\
= \; & P(X_{17} = l | X_5 = i, X_6 = j, X_9 = k) P(X_9 = k | X_5 = i, X_6 = j) P(X_6 = j | X_5 = i) P(X_5 = i) \\
= \; & P(X_{17} = l | X_9 = k) P(X_9 = k | X_6 = j) P(X_6 = j | X_5 = i) P(X_5 = i) \\
= \; & P(X_8 = l | X_0 = k) P(X_3 = k | X_0 = j) P(X_1 = j | X_0 = i) P(X_5 = i) \\
= \; & P^8_{kl} P^3_{jk} P_{ij} (\alpha P^5)_i
\end{aligned}$$

La probabilidad conjunta es obtenida apartir de la distribucion inicial α y la matriz de transicion P. Para completitud, aqui dejamos la formula general

Probabilidad conjunta Sea $X_0, X_1, ...,$ una cadena de Markov con matriz de transicion P y distribucion inicial α. Para todo $0 \leq n_1 \leq n_2 \leq ... \leq n_{k-1} \leq n_k$ y estados $i_1, i_2, ..., i_{k=1}, i_k$,

$$P(X_{n1} = i_i, X_{n2} = i_2, ..., X_k = i_k) = (\alpha P^{n_1})_{i_1} (P^{n_2 - n_1})_{i_1 i_2} ... (P^{n_k - n_{k-1}})_{i_{k-1} i_k}$$