

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, \dots 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, \dots , cuyos espacios muestrales son S con la siguiente propiedad: Dado $n \geq 0$

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

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 \mathbf{P} cuya ij -ésima entrada corresponde a $P_{ij} = P(X_1 = j | X_0 = i)$. A esta matriz se la llama **matriz de transición** 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 transición tiene $k \times k$. Si el espacio de estados es infinito contable, la matriz de transición es infinita.

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

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

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 , $\sum_j P_{ij} = 1$.

Ejemplos p.42-52

1.2 Calculos basicos

Una poderosa característica de las Cadenas de Markov es la habilidad de usar matrices para computar probabilidades. Para usar los métodos de matriz, consideramos a las distribuciones de probabilidades como vectores. Un vector de probabilidad es un vector de números no-negativos cuya suma es igual a 1. Generalmente se denotan con letras griegas en negrita, como por ejemplo α , λ , y π .

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

Para el cálculo matricial identificaremos distribución de probabilidades discretas como vectores fila.

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

1.2.1 Probabilidad de transición en n pasos

Para los estados i y j , con $n \geq 1$, $P(X_n = j | X_0 = i)$ 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 -ésima entrada es $P(X_n = j | X_0 = i)$ es la matriz de transición en n pasos de la cadena de Markov. Por supuesto, para $n=1$ es la usual matriz de transición P . Para $n \geq 1$, uno de los resultados centrales del cálculo de cadenas de Markov es que la matriz de transición en n pasos es precisamente P^n , la n -ésima potencia de P .

D/ p.53

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

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

Observación. Note que $P_{ij}^n = (P^n)_{ij}$. No hay que confundirse con $(P_{ij})^n$. También note que P^0 es la matriz identidad. Esto es

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

Ejemplos p.54-55

1.2.2 Chapman–Kolmogorov Relationship

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

$$\forall_{i,j} P_{ij}^{n+m} = \sum_k P_{ik}^m P_{kj}^n$$

D. p.55

La interpretación probabilística es que realizando una transición desde i hacia j en $m+n$ pasos es equivalente a transicionar desde i a algún estado k en m pasos y después transicionar desde este último estado a j en los restantes n pasos. Esto es conocido como **la relación Chapman-Kolmogorov**

1.2.3 Distribucion de X_n

En general, una cadena de Markov X_0, X_1, \dots , no es una secuencia de variables aleatorias igualmente distribuidas. Para $n \geq 1$, 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 ,

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

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, \dots una cadena de Markov. Luego, para toda $m < n$,

$$\begin{aligned} P(X_{n+1} = j | X_0 = i_0, \dots, X_{n-m-1} = i_{n-m-1}, X_{n-m} = i) \\ = P(X_{n+1} = j | X_{n-m} = i) = P(X_{m+1} = j | X_0 = i) = P_{ij}^{m+1} \end{aligned}$$

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

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 probabilidad conjunta de cualquier subconjunto finito de X_0, X_1, \dots . En ese sentido, la distribucion inicial y la matriz de transicion dan una completa descripcion probabilistica de una cadena de Markov

Para ilustrar consideremos una probabilidad conjunta arbitraria como:

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

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 distribución inicial α , la intuición nos dice que

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

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

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

La probabilidad conjunta es obtenida apartir de la distribución inicial α y la matriz de transición P . Para completitud, aquí dejamos la fórmula general

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

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

Ejemplo p.59

1.2.6 Comportamiento a largo plazo, la evidencia numérica

En cualquier proceso estocástico o determinista, el comportamiento a largo plazo del sistema es, muchas veces, de interés. Evidencia numérica sugiere que la potencia de una matriz converge a un límite. Lo que es más, las filas de esta matriz llevada al límite son iguales. El hecho de que las filas de una cierta matriz P^n , $n \rightarrow \infty$ son iguales significa que la probabilidad de una potencia k en particular (para un k en el cual ya se puede observar esta propiedad), no depende del estado actual. Después de cierto tiempo, el estado inicial se diluye y no afecta a la distribución de la potencia de la matriz.

Ejemplos. p59-65 y Simulaciones p.65-68

Chapter 2

Cadenas de Markov a largo plazo

2.1 Limitacion de la distribucion

En muchos casos, una cadena de Markov presenta un comportamiento limitante a largos plazos. La cadena se decanta a una distribucion equilibrada, la cual es independiente de su estado inicial.

Limitacion de la distribucion.

Sea X_0, X_1, \dots una cadena de Markov con una matriz de transicion P . Una limitacion para la distribucion de la cadena de Markov es una distribucion de probabilidad λ con la propiedad que

$$\forall_{i,j} \lim_{n \rightarrow \infty} P(X_n = j) = \lambda_j$$

En otras palabras:

- (i) Para cualquier distribucion inicial y para todo j

$$\lim_{n \rightarrow \infty} P(X_n = j) = \lambda_j$$

- (ii) Para cualquier distribucion inicial λ

$$\lim_{n \rightarrow \infty} \alpha P^n = \lambda$$

- (iii) Sea Λ la matriz estocastica cuyas filas son iguales a λ .

$$\lim_{n \rightarrow \infty} P^n = \Lambda$$

Interpretaremos λ_j como la probabilidad a largo plazo de que la cadena se encuentre en el estado j . Por la unicidad del limite, si una cadena de Markov tiene una limitacion de la distribucion, entonces esta distribucion es unica. Un metodo rapido y sucio de verificar de que una limitacion a la distribucion existe es aplicar grandes potencias a la matriz inicial hasta que de una de ellas se obtenga una matriz limitada con filas iguales.

Ejemplo p.77-78

2.2 Proporción del tiempo en cada estado

La limitación de la distribución nos da la probabilidad de que la probabilidad a plazos largos de una cadena de Markov se encuentre en cada estado. También puede ser interpretada como la probabilidad de que la proporción del tiempo, a plazos largos, de la cadena visitando cada estado.

Planteo y ejemplos p.78-80

2.3 Distribución estacionaria

Es interesante considerar que pasa si asignamos la matriz limitada de una cadena de Markov a ser la distribución inicial de la cadena. **Planteo p. 80-81**

Distribución estacionaria.

Sea X_0, X_1, \dots una cadena de Markov con una matriz de transición P . Una distribución estacionaria es una distribución de probabilidades π , que satisface

$$\pi = \pi P.$$

Esto es

$$\forall_j \pi_j = \sum_i \pi_i P_{ij}$$

Si asumimos que una distribución estacionaria π , es la distribución inicial, entonces todas las X_n tienen la misma distribución:

$$\pi P^n = (\pi P) P^{n-1} = \pi P^{n-1} \dots = \pi P = \pi$$

Si la distribución inicial es una distribución estacionaria, entonces X_0, X_1, \dots es una secuencia de variables aleatorias idénticamente distribuidas. Esto no implica que las variables aleatorias son independientes. Al contrario, la estructura dependiente entre las variables aleatorias sucesivas en la cadena de Markov está gobernada por la matriz de transición independientemente de la distribución inicial.

Nos referimos a la cadena de Markov estacionaria o a la cadena de Markov en estado estacionario cuando la cadena alcanza su distribución estacionaria. Si una cadena de Markov tiene una limitación en la distribución entonces esa distribución es una distribución estacionaria.

Hay que tener en cuenta que hay cadenas de Markov con más de una distribución estacionaria, también aquellas que no tienen distribuciones limitantes y aquellas que tienen una única distribución estacionaria que no son distribuciones limitantes.

2.4 Matrices de transicion regulares

Una matriz de transicion P es llamada regular si alguna potencia de P es positiva. Esto es, $P^n > 0$, para algun $n \geq 1$

Ejemplo p.83

Si una matriz de transicion de una cadena de Markov es regular, entonces la cadena tiene una distribucion limitante, la cual es la unica distribucion estacionaria de la cadena.

Teorema del limite para cadenas de Markov regulares.

Una cadena de Markov cuya matriz de transicion P es regular tiene una distribucion limitante, la cual es unica, positiva y distribucion estacionaria de la cadena. Esto es, existe un vector de probabilidad unico $\pi > 0$, tal que

$$\lim_{n \rightarrow \infty} P_{ij}^n = \pi_j$$

para todo i, j donde

$$\sum_i \pi_i P_{ij} = \pi_j$$

Equivalentemente, existe una matriz estocastica positiva Π tal que

$$\lim_{n \rightarrow \infty} P^n = \Pi$$

Donde Π tiene todas las filas iguales a π y π es el unico vector de probabilidad que satisface

$$\pi P = \pi$$

Ejemplo p.84

Una forma de decir si una matriz estocastica no es regular es que si para alguna potencia n , todos los 0s en P^n aparecen en los mismo lugares que todos los 0s en P^{n+1} , luego estos apareceran en los mismos lugares para todas las potencias mayores, y la matriz no es regular.

Ejemplo p.85

2.5 Buscando la distribucion estacionaria.

Supongamos que π es una distribucion estacionaria para una cadena de Markov con una matriz de transicion P . Entonces,

$$\sum_i \pi_i P_{ij} = \pi_j, \text{ para todos los estados } j$$

Lo que nos da un sistema de ecuaciones lineales. Si P es una matriz $k \times k$, el sistema tiene k ecuaciones y k incognitas. Ya que las filas de P suman 1, el sistema $k \times k$ contendra una ecuacion redundante.

Para la cadena de dos estados general, con

$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix} \quad (2.1)$$

Las ecuaciones son

$$(1-p)\pi_1 + q\pi_2 = \pi_1$$

$$p\pi_1 + (1-q)\pi_2 = \pi_2$$

Las ecuaciones son redundantes y conducen a $\pi_1 p = \pi_2 q$. Si p y q son distintos a cero, entonces junto a la condicion $\pi_1 + \pi_2 = 1$, la unica solucion es

$$\pi = \left(\frac{q}{p+q}, \frac{p}{p+q} \right)$$

Ejemplo p.86

Una tecnica util para encontrar distribuciones estacionarias, la cual reduce el numero de ecuaciones a resolver en uno. Esta tecnica usa el hecho de que si x es un vector, no necesariamente un vector de probabilidad, que satisfaga $xP=x$, entonces $(cx)P=cx$ para todas las constantes c . Si uno puede encontrar un vector x no negativo, que satisfaga $xP=x$, entonces un unico vector de probabilidad $\pi = cx$ puede ser obtenido a partir de una decision apropiada de c de tal forma que las filas de cx sumen 1. En particular, sea $c = \frac{1}{\sum_j x_j}$, el reciproco de la suma de los componentes de x

El sistema lineal $\sum_i \pi_i P_{ij} = \pi_j$ sin la restriccion $\sum_i \pi_i = 1$, tiene una ecuacion redundante. Nuestro metodo de solucion consiste en (i) eliminar la ecuacion redundante y (ii) resolver el sistema resultante para $x = (1, x_2, x_3, \dots)$, donde la primer componente (o cualquier otra) es reemplazada por 1. Para una cadena de Markov con k estados, este metodo reduce el problema a resolver un sistema lineal $(k-1) \times (k-1)$.

Si la cadena original tenia una unica distribucion estacionaria, entonces el sistema lineal reducido tendra una unica solucion, pero una la cual no es necesariamente un vector de probabilidad. Para hacerla vector de probabilidad cuyos componentes sumen 1, hay que dividir por la suma de los componentes. En otras palabras, la unica distribucion estacionaria es

$$\pi = \frac{1}{1 + x_2 + x_3 + \dots + x_k} (1, x_2, \dots, x_k)$$

Ilustracion y ejemplos p.87-90

2.6 Puedes encontrar el camino a el estado a?

El comportamiento a largo plazo de una cadena de Markov esta relacionada a cuan frecuentemente los estados son visitados. Aqui observaremos de cerca a la relacion entre estados y cuan alcanzables, o accesibles, grupos de estados son entre si.

Digamos que el estado j es **accesible** desde el estado i si $P_{ij}^n > 0$, para algun $n \geq 0$. Esto es, hay una probabilidad positiva de alcanzar j desde i en una cantidad finita de pasos. Los estados i y j se **comunican** si i es accesible desde j y j es accesible desde i . La **comunicacion** es una relacion de equivalencia, lo que significa que satisface las siguientes tres propiedades:

- a) **Reflexiva.** Todos los estados se comunican con si mismos
- b) **Simetrica.** Si i se comunica con j entonces j se comunica con i
- c) **Transitiva.** Si i se comunica con j y j se comunica con k entonces i se comunica con k

Dado que la comunicacion es una relacion de equivalencia el espacio de estados puede ser particionado en clases de equivalencias, llamadas **clases de comunicacion**. Esto es, el espacio puede ser dividido entre subconjuntos disjuntos, en los cuales todos los estados pertenecientes a una clase estan intercomunicados pero no se comunican con ningun estado fuera de la misma. Un grafo de transicion modificado es una herramienta util para encontrar las clases de comunicacion de una cadena de Markov. Los vertices del grafo son los estados de la cadena. Una arista direcccionada es dibujada entre i y j si $P_{ij} > 0$.