

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, ¿Qué podemos decir acerca de el próximo casillero (X_4) que ocupara el jugador? A pesar de que sabemos el historial completo de tiradas del jugador, la única información relevante para predecir su posición en el futuro es la ubicación más 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 resultado 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 posición más actual del jugador X_3 , su futura posición 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 estocástico 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 visitó la posición i en n pasos.

Una cadena de Markov es **homogénea 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 numerable, 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 Estocástica. Una matriz estocástica 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 Cálculos básicos

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 comenzó 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 $n \geq 0$,

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

Observacion. 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 Relación Chapman–Kolmogorov

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 Distribución 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 distribución marginal de X_n depende en el n -ésimo paso con su matriz de transición P^n , así como de la distribución inicial α . Para obtener la función 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 -ésima columna de P^n . Es decir, es el j -ésimo componente del vector-matriz resultante del producto αP^n

Ejemplo p.56

1.2.4 Presente, futuro y pasado más reciente

La propiedad de Markov dice que el pasado y el futuro son independientes dado el presente. También es verdad que el pasado y el futuro son independientes, dado el pasado más 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 Distribución conjunta

La distribución marginal de una cadena de Markov está determinada por la distribución inicial α y la matriz de transición 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 distribución inicial y la matriz de transición dan una completa descripción probabilística 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

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

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

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

Distribución límite.

Sea X_0, X_1, \dots una cadena de Markov con una matriz de transición P . Una distribución límite de la cadena de Markov es una distribución 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 α

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

- (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 límite, si una cadena de Markov tiene una distribución límite, entonces esta distribución es única. Un método rapido y sucio de verificar de que una distribución límite existe es aplicar grandes potencias a la matriz inicial hasta que de una de ellas se obtenga una matriz límite con filas iguales.

Ejemplo p.77-78

2.2 Proporción del tiempo en cada estado

La distribución límite nos presenta la probabilidad a plazos largos de que una cadena de Markov se encuentre en cada estado. También puede ser interpretada como 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 qué pasa si asignamos la matriz límite 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 alcanzó su distribución estacionaria. Si una cadena de Markov tiene una distribución límite 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 límites y aquellas que tienen una única distribución estacionaria que no son distribuciones límites.

2.4 Matrices de transición regulares

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

Ejemplo p.83

Si una matriz de transición de una cadena de Markov es regular, entonces la cadena tiene una distribución límite, la cual es la única distribución estacionaria de la cadena.

Teorema del límite para cadenas de Markov regulares.

Una cadena de Markov cuya matriz de transición P es regular tiene una distribución límite, la cual es única, positiva y distribución estacionaria de la cadena. Esto es, existe un vector de probabilidad único $\pi > 0$, 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 estocástica positiva Π tal que

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

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

$$\pi P = \pi$$

Ejemplo p.84

Una forma de decir si una matriz estocástica 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 aparecerán en los mismos lugares para todas las potencias mayores, y la matriz no es regular.

Ejemplo p.85

2.5 Buscando la distribución estacionaria.

Supongamos que π es una distribución estacionaria para una cadena de Markov con una matriz de transición 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 incógnitas. Ya que las filas de P suman 1, el sistema $k \times k$ contendrá una ecuación 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 condición $\pi_1 + \pi_2 = 1$, la única solución es

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

Ejemplo p.86

Una técnica útil para encontrar distribuciones estacionarias, la cual reduce el número de ecuaciones a resolver en uno. Esta técnica 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 único vector de probabilidad $\pi = cx$ puede ser obtenido a partir de una decisión apropiada de c de tal forma que las filas de cx sumen 1. En particular, sea $c = \frac{1}{\sum_j x_j}$, el recíproco de la suma de los componentes de x

El sistema lineal $\sum_i \pi_i P_{ij} = \pi_j$ sin la restricción $\sum_i \pi_i = 1$, tiene una ecuación redundante. Nuestro método de solución consiste en (i) eliminar la ecuación 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 método reduce el problema a resolver un sistema lineal $(k-1) \times (k-1)$.

Si la cadena original tenía una única distribución estacionaria, entonces el sistema lineal reducido tendrá una única solución, 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 única distribución estacionaria es

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

Ilustración y ejemplos p.87-90

2.6 Clases de comunicación

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

Digamos que el estado j es **accesible** desde el estado i si $P_{ij}^n > 0$, para algún $n \geq 0$. 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 **comunicación** es una relación 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

Observaciones. La propiedad $a)$ se debe a que $P_{ii}^0 = P(X_0 = i | X_0 = i) = 1$. La propiedad $b)$ se debe a que la definición de comunicación es simétrica. Para la propiedad $c)$, supongamos que i se comunica con j y que j se comunica con k . Entonces, existen $n \geq 0$ y $m \geq 0$ tal que $P_{ij}^n > 0$ y $P_{jk}^m > 0$. Por lo tanto

$$P_{ik}^{n+m} = \sum_t P_{it}^n P_{tk}^m \geq P_{ij}^n P_{jk}^m > 0$$

Lo que nos da como resultado que k es accesible desde i e i es accesible desde k

Dado que la comunicación es una relación de equivalencia el espacio de estados puede ser particionado en clases de equivalencias, llamadas **clases de comunicación**. 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 ningún estado fuera de la misma. Un grafo de transición modificado es una herramienta útil para encontrar las clases de comunicacion de una cadena de Markov. Los vértices del grtafo son los estados de la cadena. Una arista direccionada es dibujada entre i y j si $P_{ij} > 0$.

Ejemplo p.95

Irreducibilidad. Cuando una cadena de Markov tiene una sola clase de comunicación, esta es **irreducible**.

Ejemplo p.95

2.7 Recurrencia y transitoriedad

Ejemplo ilustrador p.95-96

Dada una cadena de Markov X_0, X_1, \dots , y sea $T_j = \min\{n > 0 : X_n = j\}$ la primera vez que la cadena se situa en el estado j . Si $X_n \neq j$, para toda $n > 0$, el conjunto $T_j = \infty$. Sea

$$f_j = P(T_j < \infty | X_0 = j)$$

la probabilidad de que la cadena que comienza en j eventualmente retorne a j

Estados de recurrencia y de transicion. Un estado j es llamado recurrente si la cadena de Markov iniciada en j revisita j . Esto es, $f_j = 1$. Un estado k es llamado de transicion si hay una probabilidad positiva de que la cadena de Markov iniciada en k nunca retorne a k . Esto es, $f_k < 1$.

Si un estado es eventualmente revisitado o no está estrechamente relacionado con cuan seguido ese estado es visitado. Para la cadena que inicia en i , sea

$$I_n = \begin{cases} 1 & \text{si } X_n = j, \\ 0 & \text{si } X_n \neq j \end{cases},$$

para $n \geq 0$. Luego, $\sum_{n=0}^{\infty} I_n$ es el número de visitas a j . El número esperado de visitas a j es

$$E\left(\sum_{n=0}^{\infty} I_n\right) = \sum_{n=0}^{\infty} E(I_n) = \sum_{n=0}^{\infty} P(X_n = j | X_0 = i) = \sum_{n=0}^{\infty} P_{ij}^n$$

donde la suma infinita puede divergir a ∞ . Para la cadena iniciada en i , el número esperado de visitas a j es la ij -ésima entrada de la matriz $\sum_{n=0}^{\infty} P^n$.

Supongamos que j es recurrente. La cadena que inicia en j volvera eventualmente a visitar j con probabilidad 1. Una vez que alcance 1, la cadena comienza denuevo y se comporta como una nueva version de la cadena que comienza en j . Decimos que la cadena de Markov se **regenera** asi misma. (Este comportamiento intuitivo se conoce como la propiedad fuerte de Markov.) Desde j , la cadena re-visitara j denuevo, con probabilidad 1, y asi sucesivamente. A esto le sigue que j sera revisitada una cantidad infinita de veces y

$$\sum_{n=0}^{\infty} P_{jj}^n = \infty.$$

Por otro lado, supongamos que j es de transicion. Comenzando en j , la probabilidad de eventualmente alcanzar j denuevo es f_j , y la probabilidad de nunca alcanzar j es $1-f_j$. Si la cadena alcanza j , el evento que volvera eventualmente a visitar j es independiente de la historia pasada. A esto le sigue que la secuencia de visitas sucesivas a j se comporta como una secuencia de tiradas de moneda independientes e igualmente distribuidas donde cara ocurre si se llega eventualmente a j y cruz ocurre si j nunca es alcanzada denuevo. El numero de veces que j es alcanzada es el numero de tiradas de moneda hasta que cruz ocurra, el cual tiene una distribucion geometrica con parametro $1-f_j$. Asi, el numero esperado de visitas a j es $\frac{1}{1-f_j}$ y

$$\sum_{n=0}^{\infty} P_{jj}^n = \frac{1}{1-f_j} < \infty.$$

En particular un estado de transicion sera visitado solo un numero finito de veces.

Recurrencia y transitoriedad

(i) El estado j es recurrente si y solo si

$$\sum_{n=0}^{\infty} P_{jj}^n = \infty.$$

(ii) El estado j es de transicion si y solo si

$$\sum_{n=0}^{\infty} P_{jj}^n < \infty.$$

Supongamos que j es recurrente y accesible desde i . Como la cadena empezo en i hay una probabilidad positiva de alcanzar j . Y desde j , el numero esperado de visitas a j es infinito. A esto le sigue que el numero esperado de visitas a j para la cadena iniciada en i es tambien infinito y asi

$$\sum_{n=0}^{\infty} P_{ij}^n = \infty.$$

Supongamos que j es de transicion y accesible desde i . Por un argumento similar el numero de visitas a j para la cadena iniciada en i es finita y asi

$$\sum_{n=0}^{\infty} P_{ij}^n < \infty.$$

desde el cual le sigue que

$$\lim_{n \rightarrow \infty} P_{ij}^n = 0.$$

La probabilidad a largo plazo de que una cadena de Markov eventualmente se pose sobre un estado de transicion es 0.

Recurrencia y transitoriedad son propiedades de clase. Los estados de una clase de comunicacion son todos recurrentes o todos de transicion.

D/p.99

Corolario. Para una cadena de Markov finita e irreducible, todos los estados son recurrentes.

D/p.99

Ejemplo p.100-101

2.8 Descomposicion Canonica.

Un conjunto de estados C se dice cerrado si ningun estado fuera de C es accesible desde cualquier estado en C . Si C es cerrado, entonces $P_{ij=0}$ para todo $i \in C$ y $j \notin C$

Clase de comunicacion cerrada. Una clase de comunicacion es cerrada si consiste de todos estados recurrentes. Una clase de comunicacion finita es cerrada solo si consiste de estados recurrentes.

D/101

El espacio S de una cadena de Markov finita puede ser particionada en estados de transicion y de recurrencia de la siguiente manera, $S = T \cup R_1 \cup R_2 \cup \dots \cup R_m$, donde T es el conjunto de todos los estados de transicion y los R_i son clases de comunicacion cerrada de estados de recurrencia. Este proceso es llamado descomposicion canonica. La computacion de muchas cantidades asociadas con cadenas de Markov puede ser simplificada por esta descomposicion. Dada una descomposicion canonica, el espacio de estados puede ser reordenado de manera tal que la matriz de transicion de Markov tiene la siguiente forma de matriz en bloques

$$P = \begin{matrix} & \begin{matrix} T & R_1 & \dots & R_m \end{matrix} \\ \begin{matrix} T \\ R_1 \\ \vdots \\ R_m \end{matrix} & \begin{pmatrix} * & * & \dots & * \\ 0 & P_1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & P_m \end{pmatrix} \end{matrix}$$

Cada submatriz P_1, \dots, P_m es una matriz estocastica cuadrada que corresponde a una clase de comunicacion cerrada. Por si sola, cada una de esas matrices es una cadena de Markov irreducible con un espacio de estados reducido.

Ejemplo 102

La descomposicion canonica es util para describir el comportamiento a largo plazo de una cadena de Markov. La forma de bloque de la matriz facilita obtener la potencia de las matrices. Para $n \geq 1$.

$$P^n = \begin{matrix} & \begin{matrix} T & R_1 & \dots & R_m \end{matrix} \\ \begin{matrix} T \\ R_1 \\ \vdots \\ R_m \end{matrix} & \begin{pmatrix} * & * & \dots & * \\ 0 & P_1^n & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & P_m^n \end{pmatrix} \end{matrix}$$

Tomando limites obtenemos

$$\lim_{n \rightarrow \infty} P^n = \begin{matrix} & \begin{matrix} T & R_1 & \dots & R_m \end{matrix} \\ \begin{matrix} T \\ R_1 \\ \vdots \\ R_m \end{matrix} & \begin{pmatrix} 0 & * & \dots & * \\ 0 & P_1^n & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & P_m^n \end{pmatrix} \end{matrix}$$

Notese que las entradas de las columnas correspondientes a estados de transicion estan todas en 0, como consecuencia de lo planteado anteriormente.

Las clases de comunicacion recurrentes y cerradas R_1, \dots, R_m se comportan como mini irreducibles cadenas Markov donde todos los estados se comunican entre si. Las propiedades asintoticas de las submatrices P_1, \dots, P_m nos guian a considerar las propiedades de cadenas de Markov irreducibles.

2.9 Cadenas de Markov irreducibles

Teorema del limite para cadenas de Markov irreducibles.

Suponga que X_0, X_1, \dots es una cadena de Markov irreducible. Para cada estado j , sea $\pi_j = E(T_j | X_0 = j)$ sea el tiempo esperado para retornar a j . Entonces π_j es finito, y existe una unica y positiva distribucion estacionaria π tal que

$$\forall_j \pi_j = \frac{1}{\mu_j}$$

Lo que es mas, para todos los estados i ,

$$\pi_j = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=0}^{n-1} P_{ij}^m.$$

El teorema anterior caracteriza la distribucion estacionaria π para cadenas de Markov finitas e irreducibles. Relaciona la probabilidad estacionaria π_j a el numero esperado de pasos entre visitar j . Recordemos que $T_j = \min\{n > 0 : X_n = j\}$ es el tiempo minimo que tarda en llegar al estado j .

Observaciones

- 1 El hecho de que $\pi_j = \frac{1}{\mu_j}$ es intuitivo. Si hay una visita a j cada μ_j pasos, entonces la proporcion de visitas a j es $\frac{1}{\mu_j}$.
- 2 El teorema no afirma que π es una distribucion limite. La convergencia en la ultima ecuacion es una forma debil de convergencia respecto a $\pi_j = \lim_{n \rightarrow \infty} P_{ij}^n$. Discutiremos mas adelante que se debe hacer una suposicion adicional a π para ser una distribucion limite.
- 3 Para cadenas de Markov finitas e irreducibles, todos los estados son recurrentes y el tiempo esperado $E(T_j | X_0 = j)$ es finito, para todo j . Sin embargo, para una cadena de Markov infinita si j es un estado recurrente, a pesar de que la cadena eventualmente revisitara j con probabilidad 1, el numero esperado de pasos entre estas visitas no necesariamente son finitos. El teorema puede extenderse a cadenas de Markov infinitas e irreducibles para la cual el tiempo esperado de retorno $E(T_j | X_0 = j)$ es finito, para todo j . Un estado recurrente j es llamado recurrente positivo si $E(T_j | X_0 = j) < \infty$, y nulo recurrente si $E(T_j | X_0 = j) = \infty$. Asi, el teorema se sostiene para cadenas de Markov irreducibles para la cual todos sus estados son positivos recurrentes.

En muchas aplicaciones, el tiempo esperado entre visitas a un estado en particular es de importancia.

Ejemplos. 104-105

Analisis del primer paso.

El tiempo esperado de retorno $E(T_j | X_0 = j)$ es obtenido apartir del reciproco de la probabilidad estacionaria π_j . Otro acercamiento es condicionar en el primero paso de la cadena y usar la ley de la expectativa total. Este es llamado **analisis del primer paso**.

Ejemplos. 105-106

2.10 Periodicidad

Cadenas de Markov finitas e irreducibles tienen una única distribución estacionaria positiva. A pesar de que es posible que no tengan distribución límite, ellas tienen al menos un comportamiento límite en el sentido de que para todos los estados i y j , el promedio parcial $\frac{1}{n} \sum_{m=0}^{n-1} P_{ij}^m$ converge.

Ejemplo ilustrador 107

Recordar que el divisor común más grande (gcd) de un conjunto de enteros positivos es el entero más grande que divide a todos los números del conjunto de manera tal que el resto sea 0.

Periodo. Para una cadena de Markov con una matriz de transición P , el periodo del estado i , denotado como $d(i)$, es el divisor común más grande del conjunto de posibles tiempos de retornos a i . Esto es,

$$d(i) = \gcd\{n > 0 : P_{ii}^n > 0\}.$$

Si $d(i) = 1$, el estado i es llamado **a-periodico**. Si el conjunto de tiempos de retorno al nodo i es vacío, el conjunto $d(i) = \infty$.

La periodicidad es una propiedad de clase. Los estados en una clase de comunicación tienen el mismo periodo.

D/ p.108. Ejemplos p.108-109

Cadenas de Markov periódicas y aperiódicas. Una cadena de Markov es llamada periódica si es irreducible y todos los estados tienen un periodo mayor a 1. Una cadena de Markov es aperiódica si es irreducible y todos los estados tienen periodo igual a 1.

Notar que cualquier estado i con la propiedad que $P_{ii} > 0$ es necesariamente aperiódico. Así, una condición suficiente para una cadena de Markov irreducible para ser aperiódica es que $P_{ii} > 0$ para algún i . Esto es, al menos una entrada diagonal de la matriz de transición es distinta a 0.

2.11 Cadenas de Markov ergódicas.

Una cadena de Markov es llamada ergódica si es irreducible, aperiódica y todos los estados tienen esperanza de retornar en el tiempo finitas. Lo siguiente es siempre verdadero para cadenas finitas. Así, una cadena de Markov finita es ergódica si es irreducible y aperiódica. Precisamente la clase de cadenas de Markov ergódicas tienen distribución límite positiva.

Teorema fundamental del límite para cadenas de Markov ergódicas

Sea X_0, X_1, \dots , una cadena de Markov ergódica. Existe una única y positiva distribución estacionaria π , la cual es la distribución límite de la cadena. Esto es,

$$\pi_j = \lim_{n \rightarrow \infty} P_{ij}^n, \text{ para toda } i, j$$

Este teorema afirma el mismo resultado del límite para cadenas de Markov con matrices de transición regulares. La demostración está dada en la sección 3.10, donde también se demuestra que una cadena de Markov es ergódica si y solo si su matriz de transición es regular.

Ejemplos p.110-114

2.12 Reversibilidad del tiempo

Algunas cadenas de Markov exhiben un comportamiento direccional en su evolucion. La propiedad de reversibilidad del tiempo puede ser explicada intuitivamente como sigue. Si pudieses ver un video de la cadena de Markov moviendose hacia adelante en el tiempo y luego correr el video hacia atras, no podrias decir la diferencia entre estos dos.

Reversibilidad del tiempo. Una cadena de Markov irreducible con matriz de transicion P y distribucion estacionaria π es reversible o reversible en el tiempo si

$$\pi_i P_{ij} = \pi_j P_{ji}, \text{ para toda } i, j.$$

Las ecuaciones anteriores son llamadas ecuaciones de equilibrio detalladas. Ellas dicen que para una cadena en estado estacionario,

$$P(X_0 = i, X_1 = j) = P(X_0 = j, X_1 = i), \text{ para todos } i, j.$$

Esto es, la frecuencia de transiciones de i a j es igual a la frecuencia de transiciones de j a i . Mas generalmente, si una cadena de Markov estacionaria es reversible entonces

$$P(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = P(X_0 = i_n, X_1 = i_{n-1}, \dots, X_n = i_0), \text{ para toda } i_0, i_1, \dots, i_n.$$

Si la distribucion estacionaria de una cadena de Markov es uniforme, entonces por la equacion descripta anteriormente la cadena es reversible si la matriz de transicion es simetrica.

Ejemplos p.115-119

2.13 Cadenas absorbentes

Serpientes y Escaleras. Ejemplo ilustrador.

Estado absorbente, cadena absorbente. El estado i es un estado absorbente si $P_{ii} = 1$. Una cadena de Markov es llamada cadena absorbente si tiene al menos un estado absorbente.

Consideremos una cadena de Markov absorbente con k estados de los cuales t estados son de transicion y $k-t$ son absorbentes. Estos estados pueden ser reordenados, como en la descomposicion canonica, con su matriz de transicion escrita bloques de la siguiente manera,

$$P = \left(\begin{array}{c|c} Q & R \\ \hline 0 & I \end{array} \right)$$

Donde Q es una matrix $t \times t$, R es una matriz $t \times (k-t)$, 0 es una matriz $(k-t) \times t$ de 0s y I es una matriz identidad $(k-t) \times (k-t)$.

Computar potencias de P es facilitado por la forma en bloque de la matriz. Tenemos

$$P^2 = \left(\begin{array}{c|c} Q & R \\ \hline 0 & I \end{array} \right) \left(\begin{array}{c|c} Q & R \\ \hline 0 & I \end{array} \right) = \left(\begin{array}{c|c} Q^2 & (I+Q)R \\ \hline 0 & I \end{array} \right)$$

En general para $n \geq 1$ tenemos

$$P^n = \left(\begin{array}{c|c} Q^n & (I+Q+\dots+Q^{n-1})R \\ \hline 0 & I \end{array} \right)$$

Lema. Sea A una matriz cuadrada con la propiedad que $A^n \rightarrow 0$, cuando $n \rightarrow \infty$. Entonces,

$$\sum_{n=0}^{\infty} A^n = (I - A)^{-1}.$$

D/ p.122-123.

Ahora podremos calcular $\lim_{n \rightarrow \infty} P^n$,

$$\begin{aligned} \lim_{n \rightarrow \infty} P^n &= \\ &= \lim_{n \rightarrow \infty} \left(\begin{array}{c|c} Q^n & (I+Q+\dots+Q^{n-1})R \\ \hline 0 & I \end{array} \right) \\ &= \left(\begin{array}{c|c} \lim_{n \rightarrow \infty} Q^n & \lim_{n \rightarrow \infty} (I + Q + \dots + Q^{n-1})R \\ \hline 0 & I \end{array} \right) \\ &= \left(\begin{array}{c|c} 0 & (I-Q)^{-1}R \\ \hline 0 & I \end{array} \right) \end{aligned}$$

Consideremos la interpretacion de $(I - Q)^{-1}R$. Esta matriz es indexada por filas transitorias y columnas absorbentes. La ij -esima entrada es la probabilidad a largo plazo que una cadena iniciada en un estado de transicion i sea absorbida en el estado j . Si la cadena de Markov tiene solo un estado absorbente, esta submatriz sera una $(k - 1)$ - vector columna de 1s

Ejemplos p.123-125

Numero esperado de visitas a estados de transicion. Consideremos una cadena de Markov absorbente con t estados de transicion. Sea F una matriz $t \times t$ indexada por estados de transicion, donde F_{ij} es el numero esperado de visitas a j dado que la cadena comienza en i . Entonces,

$$F = (I - Q)^{-1}$$

La matriz F es llamada matriz fundamental. **D/125-126**

2.14 Tiempo esperado de absorcion.

Para una cadena de Markov absorbente iniciada en un estado de transicion i , sea a_i el tiempo de absorcion esperado, el numero de pasos esperados para alcanzar algun estado de absorcion. El numero de transiciones desde i hasta un estado de absorcion es simplemente la suma del numero de transiciones desde i hacia cada uno de los estados de transicion hasta que eventualmente llegue a un estado de absorcion. El numero esperado de pasos desde i hacia un estado de transicion j es F_{ij} . A esto le sigue que,

$$a_i = \sum_{k \in T} F_{ik}$$

En forma de vector, $a = F1$ donde 1 es el vector columna de todos 1 s. Esto es, el tiempo de absorcion esperado son la suma de las filas de la matriz fundamental.

Cadenas de Markov absorbentes. Para una cadena de Markov con todos los estados de transicion o absorbentes, sea $F = (I - Q)^{-1}$,

1. (Probabilidad de absorcion) la probabilidad de que desde un estado de transicion i la cadena sea absorbida en el estado j es $(FR)_{ij}$
2. (Tiempo de absorcion) El numero de pasos esperados desde un estado de transicion i hasta que la cadena sea absorbida en algun estado absorbente es $(F1)_i$.

Ejemplos 127-128

2.15 Tiempo esperado para cadenas irreducibles

Para una cadena de Markov irreducible, la primera vez que la cadena llega a un estado puede ser analizado como tiempos de absorcion para una cadena adecuadamente modificada. En particular, asumamos que P es la matriz de transicion de una cadena de Markov irreducible. Para encontrar el tiempo esperado hasta que el estado i sea alcanzado por primera vez, consideremos una nueva cadena en la cual i es un estado de absorcion. La matriz de transicion \tilde{P} para la nueva cadena es obtenida haciendo 0 la i -esima fila de la matriz P y la i -esima columna de P . El tiempo que la cadena original P llega a i es igual a el tiempo de que la cadena modificada \tilde{P} es absorbida en i .

Ejemplos 129-... (sigue propiedad fuerte de Markov y demas cosas complementarias.)

2.16 Procesos de Poisson

Introduccion

Mensajes de textos arrivando a tu celular a tiempos irregulares durante el dia. Accidentes que ocurren en una avenida en tiempos y lugares aparentemente aleatoriamente distribuidos. Bebes nacidos en momentos fortuitos en un centro de maternidad. Todos estos fenomenos son modelados por el proceso de Poisson, un proceso estocastico usado para modelar la ocurrencia, o llegada, de eventos en un intervalo continuo. Tipicamente, el intervalo representa tiempo. Un proceso de Poisson es un tipo especial de proceso de conteo. Dado una corriente de eventos que arriivan en tiempos aleatorios empezando en $t=0$, sea N_t el numero de llegadas que ocurren para el tiempo t , esto es, el numero de eventos en $[0,t]$. Por ejemplo, N_t podria ser el numero de mensajes de textos recibidos hasta el tiempo t .

Para cada $t \geq 0$, N_t es una variable aleatoria. La coleccion de variables aleatorias $(N_t)_{t \geq 0}$, es un proceso estocastico en tiempo continuo de valor entero, llamado proceso de conteo. Ya que N_t cuenta eventos en $[0,t]$, a medida que t aumenta, los numeros de eventos N_t tambien lo hacen.

Proceso de conteo. Un proceso de conteo $(N_t)_{t \geq 0}$ es una coleccion de variables aleatorias enteras y no negativas tal que si $0 \leq s \leq t$, entonces $N_s \leq N_t$.

A diferencia de una cadena de Markov, la cual es una secuencia de variables aleatorias, un proceso de conteo conforma una coleccion incontable, ya que esta indexada sobre un intervalo de tiempo continuo.

El trayecto de un proceso de conteo es una funcion escalonada continua por derecha. Si $0 \leq s < t$, entonces $N_t - N_s$ es el numero de eventos en el intervalo $(s,t]$.

Grafico ilustrativo p.224

Hay varias maneras para caracterizar el proceso de Poisson. Una puede centrarse en (i) el numero de eventos que ocurren en intervalos fijos, (ii) cuando un evento ocurre, y el tiempo entre esos eventos, o (iii) el comportamiento probabilistico de eventos individuales en intervalos infinitesimales. Lo cual nos guia a tres definiciones de un proceso de Poisson, donde cada una ofrece puntos de vistas iluminadores para el modelo estocastico.

Proceso de Poisson, definicion 1.

Un proceso de Poisson con parametro λ es un proceso de conteo $(N_t)_{t \geq 0}$ con las siguientes propiedades:

1. $N_0 = 0$
2. Para todo $t > 0$, N_t tiene una distribucion de Poisson con parametro λt .
3. **Incrementos estacionarios.** Para toda $s, t > 0$, $N_{t+s} - N_s$ tiene la misma distribucion que N_t . Esto es,

$$P(N_{t+s} - N_s = k) = P(N_t = k) = \frac{e^{-\lambda t} (\lambda t)^k}{k!}, \text{ para } k = 0, 1, 2, \dots$$

4. **Incrementos independientes.** Para $0 \leq q < r \leq s < t$, $N_t - N_s$ y $N_r - N_q$ son variables aleatorias independientes.

La propiedad del incremento estacionario dice que la distribución del número de llegadas a un intervalo depende solamente de la longitud del mismo.

La propiedad del incremento independiente dice que el número de llegadas en intervalos de tiempo disjuntos son variables aleatorias independientes.

Ya que N_t tiene una distribución de Poisson, $E(N_t) = \lambda t$. Esto es, esperamos alrededor de λt llegadas en t unidades de tiempo. Por esto, la proporción de llegadas es $E(N_t)/t = \lambda$.

Ejemplos p.225-226

Proceso de Poisson trasladado.

Sea $(N_t)_{t \geq 0}$ un proceso de Poisson con parámetro λ . Para un tiempo fijo $s > 0$, consideren el proceso trasladado $(N_{t+s} - N_s)_{t \geq 0}$. El proceso trasladado es probabilísticamente equivalente al del proceso original.

El proceso de poisson trasladado es un proceso de Poisson.

Sea $(N_t)_{t \geq 0}$ un proceso de Poisson con parámetro λ . Para $s > 0$, sea

$$N'_t = N_{t+s}, \text{ para } t \geq 0$$

Luego, $(N'_t)_{t \geq 0}$ es un proceso de Poisson con parámetro λ .

Tenemos que $(N'_t)_{t \geq 0}$ es un proceso de conteo con $N'_0 = N_s - N_s = 0$. Con aumentos estacionarios, N'_t tiene la misma distribución que N_t . Y el nuevo proceso hereda incrementos estacionarios e independientes del original. A esto le sigue que si $N_s = k$, la distribución N_{t+s} es igual a la distribución de N_t .

Ejemplo p.227

2.16.1 La llegada, los tiempos entre-llegadas.

Para un proceso de Poisson con parámetro λ , sea X el tiempo de la primer llegada. Entonces, $X > t$ si y solo si no hay llegadas en $[0, t]$. Además,

$$P(X > t) = P(N_t = 0) = e^{-\lambda t}, \text{ para } t > 0.$$

Por lo tanto, X tiene una distribución exponencial con parámetro λ . La distribución exponencial juega un rol central en el proceso de Poisson. Lo que es verdadero para el tiempo de la primer llegada lo es también para el tiempo entre la primer y segunda llegada, y para todos los tiempos entre-llegadas. Un proceso de Poisson es un proceso de conteo para el cual los tiempos entre-llegadas son independientes e idénticamente distribuidas variables aleatorias exponenciales.

Proceso de Poisson, definicion 2.

Sean X_1, X_2, \dots , una secuencia de variables aleatorias exponenciales independientes e idénticamente distribuidas con parámetro λ . Para $t > 0$, sea

$$N_t = \max\{n : X_1 + \dots + X_n \leq t\}$$

, con $N_0 = 0$. Then, $(N_t)_{t \geq 0}$ define un proceso de Poisson con parámetro λ . Sea

$$S_n = X_1 + \dots + X_n, \text{ para } n = 1, 2, \dots$$

Llamaremos a S_1, S_2, \dots los tiempos de llegada del proceso, donde S_k es el tiempo de la k-esima llegada. Lo que es mas,

$$X_k = S_k - S_{k-1}, \text{ para } k = 1, 2, \dots$$

es el tiempo entre-llegada entre la (k-1)-esima y k-esima llegada, donde $S_0 = 0$.

Grafico ilustrador p.228-229

Simular

Un beneficio de la definicion 2 es que nos guia a un metodo directo para construir y simular un proceso de Poisson:

1. Let $S_0 = 0$.
2. Generar variables aleatorias exponenciales independientes e idénticamente distribuidas X_1, X_2, \dots
3. Sea $S_n = X_1 + \dots + x_n$, para $n = 1, 2, \dots$
4. Para cada $k = 0, 1, \dots$, sea $N_t = k$, para $S_k \leq t < S_{k+1}$

No tiene memoria.

La distribucion exponencial es la unica distribucion continua que es falta de memoria. Una variable aleatoria X tiene la propiedad faltarle memoria si para toda, $s, t > 0$,

$$P(X > s + t | X > s) = P(X > t)$$

Ilustracion, ejemplos y el minimo de variables aleatorias exponenciales independientes p.229-232