Cadenas de Markov

Después de mucho estudio sobre el clima, hemos visto que si un día está soleado, en el 70% de los casos el día siguiente continua soleado y en el 30% se pone nublado. En términos de probabilidad, lo que nos sirve entonces para predecir el clima, vemos que la probabilidad de que continúe soleado el día siguiente es .7 y la probabilidad de que al día siguiente esté nublado es .3. También nos fijamos en que si un día está nublado, la probabilidad de que esté soleado el día siguiente es .6 y la probabilidad de que se ponga nublado es .4.

Pregunta

Hoy está nublado, ¿cuál es la probabilidad de que mañana continúe nublado? ¿cuál es la probabilidad de que está nublado pasado mañana?

Podemos ilustrar esta situación por medio de un diagrama de árbol:

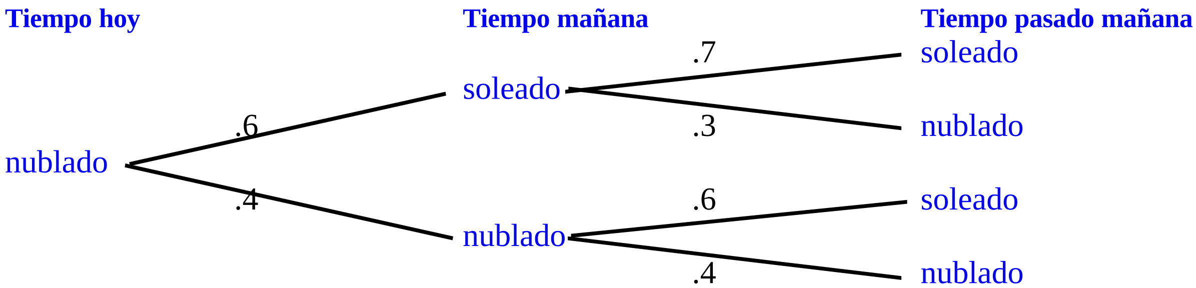

Figura 1 Posibles estados del tiempo a partir de que hoy está nublado

Con la ayuda de la Figura 1 podemos predecir qué ocurrirá mañana si sabemos que hoy está nublado. Vemos que la probabilidad de que mañana continúe nublado es .4, es decir, si hiciéramos esta predicción muchas veces estaríamos en lo correcto cerca del 40% de las veces. Para conocer la probabilidad de esté nublado pasado mañana buscamos en las hojas del árbol correspondientes al Tiempo pasado mañana los lugares donde dice nublado. Hay dos hojas donde esto ocurre. Ahora lo que queda es determinar cómo desde el principio, desde la raíz del árbol, podemos llegar allí.

Si hoy está nublado, para que pasado mañana esté nublado, podríamos tener un día de mañana soleado o nublado. Así tenemos las siguientes secuencias en orden de (hoy, mañana, pasado mañana): (nublado, soleado, nublado) o (nublado, nublado, nublado) donde pasado mañana es nublado. Estas secuencias son mutuamente excluyentes, corresponden a caminos distintos en el árbol, así tenemos que:

P(pasado mañana nublado | hoy nublado)
= P((nublado, soleado, nublado) o (nublado, nublado, nublado))
= P(nublado, soleado, nublado) + P (nublado, nublado, nublado) = $(.6 \times .3) + (.4 \times .4) = .34$.

Este resultado se obtuvo multiplicando las probabilidades condicionales a lo largo de los caminos desde hoy nublado hasta pasado mañana nublado. No es necesario que seamos tan específicos en términos de hoy, mañana o pasado mañana, podemos darnos cuenta que lo realmente importante es el número de días que pasa entre una predicción y otra. El problema que tratamos es equivalente al problema en que si en el día 0 está nublado, ¿cuál es la probabilidad un día después también esté nublado?, ¿dos días después?, ¿100 días después?...

Pregunta
Hoy está nublado, ¿cuál es la probabilidad de que esté nublado tres días después? Representa esta situación con un árbol.

El proceso de este ejemplo sólo puede adquirir uno de dos estados posibles s_1 = nublado y s_2 = soleado. La probabilidad con que se va de un estado a otro depende del estado en que estamos en el presente. Dejemos que X_1 represente el estado del clima del día número 1, X_2 el estado del clima del día número 2 y así sucesivamente. En general, para n = 1, 2, ... sea X_n el estado del clima en el enésimo día.

La sucesión de observaciones X_1, X_2, ... se llama un **proceso estocástico** o **proceso aleatorio**. La primera observación X_1 se conoce como el **estado inicial** del proceso y para n = 2, 3, ... , X_n es el **estado del proceso** en el tiempo n. En un proceso de este tipo los valores de las observaciones no pueden predecirse con precisión de antemano. Sin embargo puede especificarse una probabilidad de observar determinado valor, tal como en nuestro ejemplo.

En un proceso estocástico el estado varía en una forma aleatoria. Para describir el modelo de probabilidad es necesario especificar una probabilidad para cada uno de los posibles valores del estado inicial. También es necesario especificar para cada estado subsiguiente X_{n+1} todas las probabilidades condicionales de la forma siguiente: $P(X_{n+1} = s_{n+1} \mid X_1 = s_1, X_2 = s_2, ..., X_n = s_n)$. Esto quiere decir que para todos los tiempos n, el modelo de probabilidad debe especificar la probabilidad condicional de que el proceso esté en el estado s_{n+1} en el tiempo n+1, dado que en los tiempos 1, 2, ..., n el proceso estuvo en los estados s_1, s_2, ..., s_n.

Muchos procesos reales, como el del ejemplo, se pueden modelar examinando únicamente la historia más reciente, es decir, examinando su último estado, sin considerar todos los estados anteriores. Una **cadena de Markov** (general) es un proceso de esta naturaleza: en el momento n el estado actual del proceso y todos los estados anteriores son conocidos, entonces las probabilidades de todos los estados futuros X_j (j > n) dependen únicamente del estado actual X_n y no de los anteriores, X_1, X_2, ... X_{n-1}. Esto se puede ver en el diagrama de árbol, si sabemos cuál es estado del clima hoy, no tenemos que saber cuál fue el de ayer, antier o antes.

Formalmente, una cadena de Markov es un proceso estocástico tal que para n = 1, 2, ... y para cualquier sucesión posible de estados s_1, s_2, ..., s_{n+1}, tenemos
$P(X_{n+1} = s_{n+1} \mid X_1 = s_1, X_2 = s_2, ..., X_n = s_n) = P(X_{n+1} = s_{n+1} \mid X_n = s_n)$.

Tal como en el ejemplo del clima, si usamos la regla de multiplicación repetidas veces vemos que las probabilidades en una cadena de Markov deben cumplir:

$$P(X_1 = s_1, X_2 = s_2, ..., X_n = s_n)$$
$$= P (X_1 = s_1) P(X_2 = s_2 \mid X_1 = s_1) P(X_3 = s_3 \mid X_2 = s_2) ... P(X_n = s_n \mid X_{n-1} = s_{n-1}).$$

Pregunta
Demuestra este último resultado.

En general, consideramos una cadena de Markov que en cualquier momento estará en alguno de un número finito k de estados s_1, s_2, ..., s_k . Un proceso aleatorio de esta naturaleza se llama una **cadena de Markov finita**.

Pregunta

¿Es el ejemplo del estado del clima una cadena de Markov finita? Explica. Identifica los estados.

La probabilidad condicional $P(X_{n+1} = s_j \mid X_n = s_i)$, de que la cadena estará en el estado s_j en el tiempo n + 1 si está en el estado s_i en el tiempo n se conoce como una **probabilidad de transición**. Si para una cadena de Markov esta probabilidad de transición tiene el mismo valor para todo los tiempos n (n = 1, 2, 3, ...) decimos que la cadena tiene **probabilidades de transición estacionarias**. Es decir una cadena de Markov tiene probabilidades de transición estacionarias si para cualquier estados s_i y s_j existe una probabilidad de transición p_{ij} tal que $P(X_{n+1} = s_j \mid X_n = s_i) = p_{ij}$ para n =1, 2, 3,...

Pregunta

¿Son las probabilidades del ejemplo del estado del clima estacionarias? Construye ejemplos de situaciones que se puedan representar por procesos estocásticos, por cadenas de Markov generales, por cadenas de Markov finitas, por cadenas de Markov finitas con probabilidades estacionarias.

Las probabilidades del ejemplo pueden presentarse en forma de matriz:

$$\begin{array}{cc} & \begin{array}{cc} soleado(s_1) & nublado(s_2) \end{array} \\ \begin{array}{c} soleado(s_1) \\ nublado(s_2) \end{array} & \begin{pmatrix} .7 & .3 \\ .6 & .4 \end{pmatrix} \end{array}$$

Esta matriz que incluye las probabilidades de pasar de un estado a otro en un paso se llama la **matriz de transición**. Los elementos en cada una de las filas suman uno. En la primera fila representamos $P(X_n = s_1 \mid X_{n-1} = s_1) = P(\text{ soleado} \mid \text{soleado}) = .7$ y $P(X_n = s_2 \mid X_{n\ 1} = s_1) = P(\text{ nublado} \mid \text{soleado}) = .3$.

Pregunta

¿Qué probabilidades condicionales representa la segunda fila?

En general, considera una cadena de Markov con k estados posibles $s_1, s_2, ..., s_k$ y probabilidades estacionarias. Para i = 1, 2, 3, ..., k y j = 1, 2, 3, ..., k denotaremos por p_{ij} la probabilidad condicional de que el proceso estará en el estado s_j en un determinado momento s_i está en el estado si en el momento inmediatamente anterior. Entonces la matriz de transición de la cadena de Markov se define como una matriz de dimensiones $k \times k$, que llamamos **P** con elementos p_{ij}:

$$\mathbf{P} = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1k} \\ p_{21} & p_{22} & \cdots & p_{2k} \\ \vdots & \vdots & & \vdots \\ p_{k1} & p_{k2} & \cdots & p_{kk} \end{pmatrix}.$$

El elemento en la fila i, columna j $p_{ij} = P(X_n = s_j \mid X_{n-1} = s_i)$, representa la probabilidad de transición de un paso. Estos elementos son probabilidades, vemos que $p_{ij} \geq 0$ para todo i, j y que además la suma de estos valores en cada fila es igual a 1: $\sum_{j=1}^{k} p_{ij} = 1$. Esto nos dice que si estamos en el estado i, entonces la suma de las probabilidades de ir a un estado $s_1, s_2, ..., s_k$ es uno.

El usar esta representación en forma de matriz nos facilita el cómputo de las probabilidades de transición en más de un paso. Regresemos al ejemplo del estado del clima y añadamos algo de notación. Digamos que el estado s_1 es igual a n, indicando que el día está nublado y s_2 es igual a s indicando que está soleado. Sea X_n el estado del clima en el enésimo día en que se observe. Así X_n será igual a n si el día está nublado y será igual a s si el día está soleado.

La pregunta que contestamos antes, la probabilidad de que pasado mañana esté nublado si sabemos que hoy está nublado corresponde entonces a encontrar $P(X_3 = n \mid X_1 = n)$. Para calcular esta probabilidad examinamos los posibles estados del día de mañana y las formas de cómo llegar a $X_3 = n$, vemos esto en el siguiente árbol:

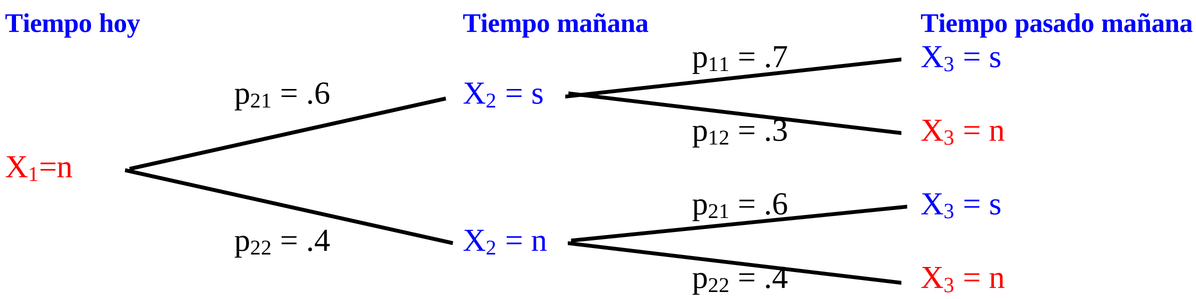

Figura 2 Notación para los posibles estados del tiempo

Así la probabilidad que nos interesa se puede calcular como hicimos antes, usando la fórmula de probabilidad total:

$P(X_3 = n \mid X_1 = n) \quad = (.6 \times .3) + (.4 \times .4)$ en términos de nuestra notación esto es igual a

$= (p_{21} \times p_{12}) + (p_{22} \times p_{22})$

$= P(X_2 = s \mid X_1 = n)\, P(X_3 = n \mid X_2 = s) + P(X_2 = n \mid X_1 = n)\, P(X_3 = n \mid X_2 = n).$

Es decir, hemos descompuesto el evento de que pasado mañana esté nublado si sabemos que hoy lo está en términos de todas los estados que se pueden observar mañana. Entonces, para cada posible estado del clima de mañana examinamos cómo podemos tener el día de pasado mañana nublado. Ahora veamos la expresión resultante. Si examinamos la expresión $(p_{21} \times p_{12}) + (p_{22} \times p_{22})$, vemos que ésta corresponde al elemento en la segunda fila y segunda columna de la matriz que resulta al multiplicar $\mathbf{P} \times \mathbf{P}$. Es decir

$$\mathbf{P} \times \mathbf{P} = \begin{pmatrix} .7 & .3 \\ .6 & .4 \end{pmatrix} \begin{pmatrix} .7 & .3 \\ .6 & .4 \end{pmatrix} = \begin{pmatrix} .7 \times .7 + .3 \times .6 & .7 \times .3 + .3 \times .4 \\ .6 \times .7 + .4 \times .6 & .6 \times .3 + .4 \times .4 \end{pmatrix} = \begin{pmatrix} .67 & .33 \\ .66 & .34 \end{pmatrix}.$$

Esta última matriz es la matriz de transición del proceso en dos pasos. Nos da las probabilidades de llegar en dos pasos a cualquier estado si partimos de un estado particular. Por ejemplo, de ahí podemos leer que $P(X_3 = n \mid X_1 = s) = .33$ y $P(X_3 = s \mid X_1 = n) = .66$. Como el proceso es estacionario podemos inclusive determinar que $P(X_5 = n \mid X_3 = n) = .34$, ya que lo único que afecta el resultado es el número de días en el futuro en que queremos hacer la predicción.

Podemos extender este argumento a cualquier número de días en el futuro en que queremos hacer la predicción, vemos que $\mathbf{P} \times \mathbf{P} = \mathbf{P}^{\mathbf{2}}$ corresponde a las probabilidades de transición en dos pasos, entonces $\mathbf{P}^{\mathbf{3}}$, $\mathbf{P}^{\mathbf{4}}$, ..., $\mathbf{P}^{\mathbf{m}}$, ... corresponden a las probabilidades de transición en 3, 4, ..., m pasos respectivamente. De hecho, la matriz $\mathbf{P}^{m}$ se conoce como la **matriz de transición en m pasos** de la cadena de Markov.

Estas matrices pueden encontrarse fácilmente usando una calculadora con esta capacidad o un programa de computadoras tal como Excel.

Hasta ahora no sabemos las probabilidades de encontrar el clima en un estado particular (soleado o nublado) irrespectivamente del estado del tiempo de hoy, es decir no sabemos $P(X_n = s)$, por ejemplo. Si aplicamos la definición de probabilidad condicional y la fórmula de probabilidad total, sabemos que

$$P(X_2 = s) = P(X_2 = s \text{ y } X_1 = s) + P(X_2 = s \text{ y } X_1 = n)$$

$$= P(X_2 = s \mid X_1 = s)P(X_1 = s) + P(X_2 = s \mid X_1 = n)P(X_1 = n).$$

Así que para calcular $P(X_2 = s)$ debemos conocer el valor de $P(X_1 = s)$ y de $P(X_1 = n)$ además de las probabilidades condicionales. Estas últimas están dadas en la matriz de transición, así que son conocidas. Las probabilidades $P(X_1 = s) = w_1$ y $P(X_1 = n) = w_2$ se conocen como las **probabilidades iniciales** del sistema. Su valor se puede estimar o suponer a través del historial de días soleados o nublados con que se cuente hasta ese momento. Podemos escribir $P(X_2 = s) = w_1p_{11} + w_2p_{21}$, recordando que el estado número 1 corresponde a soleado y el estado número 2 corresponde a nublado. Debemos darnos cuenta que $w_1 + w_2 = 1$, pues estas probabilidades incluyen todos los posibles estados en que puede estar el sistema.

Pregunta

Encuentra $P(X_2 = s)$ si $w_1 = .7$ y $w_2 = .3$.

Podemos extender la notación de matrices a las probabilidades iniciales, esta vez usando un vector de probabilidad. Considera una cadena de Markov con k estados posibles $s_1, s_2, ..., s_k$ y probabilidades estacionarias. Supongamos que al inicio, la cadena puede estar en cualquiera de los k estados con la probabilidad de que esté en el estado s_i con probabilidad $w_i \geq 0$, es decir $P(X_1 = s_i) = w_i$ para $i = 1, 2, ..., k$ y que además $w_1 + w_2 + ...+ w_k = 1$. Estas probabilidades describen un vector de probabilidad $\mathbf{w} = (w_1, w_2, ..., w_k)$, en este caso llamado **vector de probabilidad inicial**.

Usando este vector de probabilidad inicial podemos calcular por ejemplo,

$$P(X_2 = s_j) = \sum_{i=1}^{k} P(X_1 = s_i \; y \; X_2 = s_j)$$

$$= \sum_{i=1}^{k} P(X_1 = s_i) P(X_2 = s_j \mid X_1 = s_i) = \sum_{i=1}^{k} w_i p_{ij}.$$

Esta última sumatoria corresponde al componente j del vector **wP**, así las probabilidades de que X_2 esté en cualquiera de los k estados están dadas por el vector **vP**. Podemos generalizar este resultado aún más. Supongamos que en el tiempo n, la probabilidad de que el sistema esté en el estado s_i es $P(X_n = s_i) = v_i$, entonces $P(X_{n+1} = s_j) = \sum_{i=1}^{k} v_i p_{ij}.$ para j = 1, 2, ..., k. Dicho de otra forma, si las probabilidades de los estados en el tiempo n están especificadas por el vector de probabilidad **v**, entonces las probabilidades en el tiempo n+1 están especificadas por el vector de probabilidad **vP**. De aquí podemos ver que si el vector de probabilidad inicial para una cadena de Markov con probabilidades de transición estacionarias es **w**, entonces las probabilidades de los varios estados en el tiempo n+1 están especificados por el vector de probabilidad $\mathbf{vP^n}$.

Pregunta

Demuestra esta última aseveración. Encuentra el vector de probabilidad correspondiente a X_3 si $\boldsymbol{w} = (.7, .3)$.

Ejercicios y Problemas

1. Caminata aleatoria con barreras que reflejan.
2. Caminata aleatoria con barreras absorbentes.
3. Difusión de gases
4. Modelo de colas
5. Cajas con canicas