Probabilidad y estadística

Unidad 7: Procesos estocásticos - Cadenas de Markov

Katherine Sullivan

FCEIA - UNR

Índice

¿Qué es una cadena de Markov?

- Una cadena $\rightarrow$ un proceso estocástico con espacio de estados, S, discreto (a veces -y en este caso-, conjunto de instantes de observación, T, discreto)
- de Markov $\rightarrow$ cuyas variables aleatorias cumplen la propiedad markoviana

$$\forall i_0, i_1, \dots, i_{n-1}, i, j \text{ se cumple}$$

$$P\left(\underbrace{X_{n+1}=j}_{\textbf{futuro}} / \underbrace{X_0=i_0, X_1=i_1, \dots, X_{n-1}=i_{n-1}}_{\textbf{pasado}}, \underbrace{X_n=i}_{\textbf{presente}}\right) = P\left(X_{n+1}=j / X_n=i\right)$$

Diferentes enfoques

Podríamos estudiar las CM viéndolas como:

- una familia de variables aleatorias, o
- un sistema de transición con estados y transiciones probabilistas entre estados

Usaremos este último

¿Por qué pensar un enfoque con estados?

Pensar las cadenas de Markov con un enfoque con estados y transiciones resulta

- más *computacional*, y
- es como se las suele estudiar en el campo de la verificación de modelos

Cadena de Markov

Definición

Una cadena de Markov es una tupla $\mathcal{M} = (S, \mathbf{T}, \iota_{init})$ *donde*

- S *es un conjunto de estados numerable y no vacío*
- $\mathbf{T} : S \times S \to [0, 1]$ *es la función de probabilidad de transición y es tal que para todo estado* $s \in S$

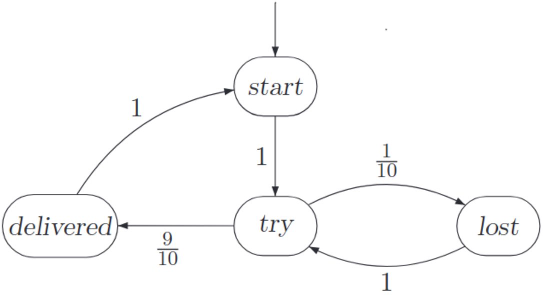
$$\sum_{s' \in S} \mathbf{T}(s, s') = 1$$

- $\iota_{init} : S \to [0, 1]$ *es la distribución inicial y es tal que* $\sum_{s \in S} \iota_{init}(s) = 1$

Algunos comentario sobre cómo estudiaremos a las CMs

- Es usual identificar la función de probabilidad de transición $\mathbf{T} : S \times S \to [0,1]$ con la matriz $(\mathbf{T}(s,t))_{s,t \in S}$. La fila $\mathbf{T}(s,\cdot)$ para el estado s contiene las probabilidades de pasar del estado s a sus sucesores, mientras que la columna $\mathbf{T}(\cdot,s)$ para el estado s especifica la probabilidad de entrar al estado s desde cualquier otro estado.
- Similarmente, se suele ver a la distribución inicial como el vector $(\iota_{init}(s))_{s \in S}$.
- Una cadena de Markov induce un grafo dirigido subyacente, donde los estados actúan como vértices y habrá una arista de s a s' sii $\mathbf{T}(s,s') > 0$, y generalmente son ilustradas de esta manera.
- Hablaremos de caminos en CMs. Un camino es una secuencia infinita de estados $\pi = s_0 s_1 s_2 \cdots \in S^\omega$ tal que $\mathbf{T}(s_i, s_{i+1}) > 0$ para todo $i \geq 0$. Notaremos al conjunto de caminos en una CM $\mathcal{M}$ con $Paths(\mathcal{M})$.

Ejemplos de cadenas de Markov - Protocolo simple de comunicación



$$\mathbf{T} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{10} & \frac{9}{10} \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \qquad \iota_{init} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

¿Cómo hablar de probabilidades en una CM?

- Queremos razonar sobre la probabilidad de ciertos conjuntos de caminos.
- Para ello necesitamos un marco matemático que nos permita asignar probabilidades a conjuntos (eventos) de manera rigurosa.
- Este marco será la teoría de la medida a través de *espacios de probabilidad* y *σ-álgebras*.
- Antes de poder asignarle probabilidades a este, un evento debe poder ser *probabilizable*.
- Un evento será *probabilizable* o, en la teoría de la medida, medible, solo si pertenece a un álgebra o σ-álgebra.

Definición de σ-álgebra

Sea Outc un conjunto no vacío de resultados y $\mathcal{A} \subseteq 2^{\mathrm{Outc}}$ un subconjunto de este cuyos elementos serán los que queremos medir (y a los que llamaremos eventos). El par $(\mathrm{Outc}, \mathcal{A})$ es una σ-**álgebra** si cumple:

1. $\emptyset \in \mathcal{A}$.
2. Si $A \in \mathcal{A}$ entonces $\overline{A} = \mathrm{Outc} \setminus A \in \mathcal{A}$ (cierre por complemento).
3. Si $A_1, A_2, \cdots \in \mathcal{A}$ entonces $\bigcup_{n \geq 1} A_n \in \mathcal{A}$ (cierre por uniones contables).

Y tendremos:

- $\mathrm{Outc} \in \mathcal{A}$ como $\mathrm{Outc} = \overline{\emptyset}$.
- Cierre también por intersecciones contables como $\bigcap_{n \geq 0} A_n = \overline{\cup_{n \geq 0} \overline{A_n}}$

Ejemplos de σ-álgebras

- **Potencia completa:** $\mathcal{A} = 2^{\mathrm{Outc}}$ (todos los subconjuntos son eventos).
- **Trivial:** $\mathcal{A} = \{\emptyset, \mathrm{Outc}\}$ (solo el vacío y el universo como eventos).

Espacio de probabilidad

Un **espacio de probabilidad** es una tripla $(\mathrm{Outc}, \mathcal{A}, \mathsf{Pr})$ donde:

- $(\mathrm{Outc}, \mathcal{A})$ es un σ-álgebra.
- $\mathsf{Pr} : \mathcal{A} \to [0, 1]$ es una medida de probabilidad, es decir, es tal que:
  1. $\mathsf{Pr}(\mathrm{Outc}) = 1$.
  2. Si A_i disjuntos dos a dos, $\mathsf{Pr}(\bigcup_i A_i) = \sum_i \mathsf{Pr}(A_i)$.

Ejemplo: Moneda justa

- $\mathrm{Outc} = \{\mathsf{cara}, \mathsf{cruz}\}$, $\mathcal{A} = 2^{\mathrm{Outc}}$.
- Una medida de probabilidad podría ser:

$$\mathsf{Pr}(\{\mathsf{cara}\}) = \mathsf{Pr}(\{\mathsf{cruz}\}) = \frac{1}{2}, \mathsf{Pr}(\emptyset) = 0, \mathsf{Pr}(\mathrm{Outc}) = 1.$$

σ-álgebra en cadenas de Markov

Para definir una σ-ágebra apropiada para una cadena de Markov, usaremos el hecho de que para cada conjunto Outc y cada subconjunto Π de 2^{Outc} **existe una σ-álgebra maś pequeña que contiene a Π.**

Esto es debido a las observaciones de que:

- el conjunto potencia 2^{Outc} es una σ-álgebra, y
- la intersección de σ-álgebras es una σ-álgebra.

Consecuentemente, la intersección $\mathcal{A}_\Pi = \cap_{\mathcal{A}} \mathcal{A}$ donde $\mathcal{A}$ itera sobre todas las σ-álgebras en Outc que contienen a Π es una σ-álgebra y está contenida en cualquier σ-álgebra $\mathcal{A}$ tal que $\Pi \subseteq \mathcal{A}$. $\mathcal{A}_\Pi$ se llama la σ-**álgebra generada por** Π y Π es la base para $\mathcal{A}_\Pi$.

σ-álgebra en cadenas de Markov

- Para una cadena de Markov $\mathcal{M}$, los eventos serán los caminos infinitos: $\mathrm{Outc}^{\mathcal{M}} = \mathrm{Paths}(\mathcal{M})$.
- La σ-álgebra asociada con $\mathcal{M}$ será la generada por los conjuntos cilindros asociados a los fragmentos de caminos finitos en $\mathcal{M}$.

Definición (Conjunto cilindro)

El conjunto cilindro de $\hat{\pi} = s_0 \ldots s_n \in Paths_{fin}(\mathcal{M})$ se define como

$$\mathrm{Cyl}(\hat{\pi}) = \{\pi \in Paths(\mathcal{M}) \mid \hat{\pi} \textit{ es prefijo de } \pi\}.$$

Los conjuntos cilindros sirven entonces como los eventos base de la σ-álgebra $\mathcal{A}^{\mathcal{M}}$ asociada a $\mathcal{M}$.

Sigma álgebra de una cadena de Markov

Definición

La σ-álgebra $\mathcal{A}^{\mathcal{M}}$ asociada a $\mathcal{M}$ es la σ-álgebra más pequeña que contiene a todos los conjuntos cilindro $\mathrm{Cyl}(\hat{\pi})$ donde $\hat{\pi}$ itera sobre todos los fragmentos de camino finitos en $\mathcal{M}$.

Por resultados clásicos de teoría de probabilidad, existe una única medida $\Pr^{\mathcal{M}}$ en $(Paths(\mathcal{M}), \mathcal{A}^{\mathcal{M}})$ tal que:

$$\Pr^{\mathcal{M}}(\mathrm{Cyl}(s_0 \ldots s_n)) = \iota_{\mathrm{init}}(s_0) \cdot \mathbf{T}(s_0 s_1 \ldots s_n).$$

donde

$$\mathbf{T}(s_0 s_1 \ldots s_n) = \Pi_{0 \leq i < n} \mathbf{T}(s_i, s_{i+1})$$

y para fragmentos de longitud 0, $\mathbf{T}(s_0) = 1$

Notación LTL

Antes de seguir, haremos unas aclaraciones sobre la notación. Para describir ciertos conjuntos de caminos en una cadena de Markov usaremos notación LTL. Lo importante a saber de la notación LTL para comprender la presentación será:

- Al evento de alcanzar eventualmente, es decir, en una cantidad finita de pasos, algún estado de un conjunto de estados B lo notaremos $\Diamond B$.
- Al evento de solo pasar por estados en un conjunto C hasta eventualmente llegar a un estado de un conjunto B lo notaremos $C \mathcal{U} B$.
- A ambas notaciones para eventos las podremos anotar con restricciones en esa cantidad de pasos finitos hasta llegar al conjunto de estados deseados. Por ejemplo:
  - $\Diamond^{=n} B$ notará el evento de alcanzar B en exactamente n pasos.
  - $C \mathcal{U}^{\leq n} B$ notará el evento de alcanzar B antes solo habiendo pasado por estados en C en a lo sumo n pasos.

Probabilidades de alcanzabilidad I

Una de las preguntas elementales para el análisis de sistemas es "¿cuál es la probabilidad de eventualmente llegar a cierto conjunto B de estados?". Es decir, nos preguntamos por $\Pr^{\mathcal{M}}(\Diamond B)$, usando notación LTL.

Pensemos en el ejemplo del protocolo simple de comunicación que vimos y pensemos que nuestro objetivo es computar cuál es la probabilidad de llegar al estado *delivered*.

Para este evento, los fragmentos de caminos iniciales $s_0 \dots s_n$ con $s_i \neq \mathit{delivered}$ para $0 \leq i < n$ y $s_n = \mathit{delivered}$ son de interés. Estos fragmentos de caminos tendrán la forma

$$\hat{\pi}_n = \mathit{start}\ \mathit{try}\ (\mathit{lost}\ \mathit{try})^n\ \mathit{delivered}$$

donde n es un número natural arbitrario. La probabilidad de $\hat{\pi}_n$ ocurra será $\left(\frac{1}{10}\right)^n \cdot \frac{9}{10}$. Entonces, $\Pr^{\mathcal{M}}(\Diamond\ \mathit{delivered}) = \sum_{n=0}^{\infty} \left(\frac{1}{10}\right)^n \cdot \frac{9}{10} = \frac{\frac{9}{10}}{1-\frac{1}{10}} = 1$.

Probabilidades de alcanzabilidad II

El ejemplo anterior muestra cómo las probabilidades de llegar a cierto conjunto de estados pueden ser calculadas a través de sumas infinitas. Esto en ejemplos complejos puede ser muy difícil de calcular. Por eso, veremos cómo podemos computar esto de manera más eficiente.

Llamaremos x_s a la variable que denota la probabilidad de llegar a B desde s, para algún s arbitrario. I.e., $x_s = \Pr(s \models \Diamond B)$ en notación LTL.

Si B no es alcanzable desde s en el grafo subyacente, $x_s = 0$, mientras que si es alcanzable $x_s > 0$ y si $s \in B, x_s = 1$.

Para los estados $s \in S \setminus B$ para los cuales B es alcanzable vale que:

$$x_s = \underbrace{\sum_{t \in S \setminus B} \mathbf{T}(s,t) \cdot x_t}_{\text{doy un paso y veo la probabilidad desde ahí}} + \underbrace{\sum_{u \in B} \mathbf{T}(s,u)}_{\text{llego en un paso}}$$

Probabilidades de alcanzabilidad III

Denotemos ahora con $\tilde{S}$ al conjunto de estados $s \in S \setminus B$ que alcanzan B, i.e., para los que existe un fragmento de camino finito $s_0 s_1 \ldots s_n$ $(n > 0)$ con $s_0 = s$ y $s_n \in B$.
Entonces para el vector $\mathbf{x} = (x_s)_{s \in \tilde{S}}$ tenemos que

$$\mathbf{x} = \mathbf{A}\mathbf{x} + \mathbf{b},$$

donde

- la matriz $\mathbf{A}$ contiene las probabilidades de transición para los estados en $\tilde{S}$, i.e., $\mathbf{A} = (\mathbf{T}(s,t))_{s,t \in \tilde{S}}$, y
- el vector $\mathbf{b} = (b_s)_{s \in \tilde{S}}$ contiene las probabilidades de alcanzar B en un paso, i.e., $b_s = \mathbf{T}(s,B) = \sum_{u \in B} \mathbf{T}(s,u)$

El sistema de ecuaciones puede ser reescrito como el sistema lineal

$$(\mathbf{I} - \mathbf{A}) \cdot \mathbf{x} = \mathbf{b},$$

donde $\mathbf{I}$ es la matriz de identidad $|\tilde{S}| \times |\tilde{S}|$.

Probabilidades de alcanzabilidad IV

Volvamos entonces a nuestro ejemplo de protocolo simple de comunicación y al análisis de las probabilidades de $\Diamond B$ con $B = \{delivered\}$.
$x_s > 0$ para todo estado s, dado que *delivered* es alcanzable desde todos los estados, con lo que $\tilde{S} = \{start, try, lost\}$ y obtenemos las siguientes ecuaciones

$$\begin{array}{lcl} x_{start} & = & x_{try} \\ x_{try} & = & \frac{1}{10} \cdot x_{lost} + \frac{9}{10} \\ x_{lost} & = & x_{try}. \end{array}$$

Estas ecuaciones pueden ser reescritas como

$$\begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -\frac{1}{10} \\ 0 & -1 & 1 \end{pmatrix} \cdot \mathbf{x} = \begin{pmatrix} 0 \\ \frac{9}{10} \\ 0 \end{pmatrix}$$

que resulta en una única solución $x_{start} = x_{try} = x_{lost} = 1$.

Probabilidades de alcanzabilidad V

La técnica descripta se puede entender como un algoritmo de dos partes:

1. realizamos un análisis del grafo para computar el conjunto de estados que pueden alcanzar B (con un backward DFS- o un BFS)
2. generamos la matriz $\mathbf{A}$ y el vector $\mathbf{b}$ y resolvemos el sistema de ecuaciones lineales $(\mathbf{I} - \mathbf{A}) \cdot \mathbf{x} = \mathbf{b}$

Sin embargo, si $\mathbf{I} - \mathbf{A}$ no tiene inversa, $(\mathbf{I} - \mathbf{A}) \cdot \mathbf{x} = \mathbf{b}$ puede tener más de una solución.

Este problema se aborda caracterizando al vector deseado como la menor solución en $[0,1]^{\tilde{S}}$. Esta caracterización posibilita computar el vector de probabilidad con un método de aproximación iterativa.

Vamos a presentar un algoritmo formal pero lo haremos para un tipo de problema más general el de **alcanzabilidad restringida** o de "propiedades until".

Probabilidades de alcanzabilidad VI - until operator

- El evento de alcanzar B a través de un fragmento de camino que termina en un estado $s \in B$, y antes solo visita estados en en conjunto C usando notación LTL, se denota $C \mathcal{U} B$.
- El evento $\Diamond B$ coincide con $S \mathcal{U} B$
- Para $n \geq 0$, el evento $C \mathcal{U}^n B$ tiene el mismo significado que $C \mathcal{U} B$ excepto que se requiere alcanzar B en cuanto mucho n pasos. Formalmente, $C \mathcal{U}^n B$ es la unión de los cilindros básicos producidos por los fragmentos de caminos $s_0 s_1 \ldots s_k$ tales que $k \leq n$ y $s_i \in C$ para todo $0 \leq i < k$ y $s_k \in B$.

Probabilidades de alcanzabilidad VII

Sean $S_{=0}$, $S_{=1}$ y $S_?$ una partición de S tal que

- $B \subseteq S_{=1} \subseteq \{s \in S | \Pr(s \models C \mathcal{U} B) = 1\}$,
- $S \setminus (C \cup B) \subseteq S_{=0} \subseteq \{s \in S | \Pr(s \models C \mathcal{U} B) = 0\}$, y
- $S_? = S \setminus (S_{=1} \cup S_{=0})$.

Y sean

- $\mathbf{A} = (\mathbf{T}(s,t))_{s,t \in S_?}$
- $\mathbf{b} = (b_s)_{s \in S_?}$ donde $b_s = \mathbf{T}(s, S_{=1})$

Ahora presentaremos una caracterización de menor punto fijo para calcular el vector de probabilidad $(\Pr(s \models C \mathcal{U} B))_{s \in S_?}$.
Para calcular el menor punto fijo el conjunto $[0,1]^{S_?}$ viene equipado con un orden parcial $\leq$ dado por $\mathbf{y} \leq \mathbf{y'}$ sii $y_s \leq y'_s$ para todo $s \in S_?$, donde $\mathbf{y} = (y_s)_{s \in S_?}$ e $\mathbf{y'} = (y'_s)_{s \in S_?}$.

Probabilidades de alcanzabilidad VIII - punto fijo

Teorema (Caracterización por menor punto fijo)

El vector $\mathbf{x} = (\Pr(s \models C\mathcal{U}B))_{s\in S_?}$ *el el menor punto fijo del operador* $\Upsilon : [0,1]^{S_?} \to [0,1]^{S_?}$ *que está dado por*

$$\Upsilon(\mathbf{y}) = \mathbf{A}\cdot\mathbf{y} + \mathbf{b}.$$

Además, si $\mathbf{x}^{(0)} = \mathbf{0}$ *es el vector nulo, y* $\mathbf{x}^{(n+1)} = \Upsilon(\mathbf{x}^{(n)})$ *para* $n \geq 0$, *entonces:*

- $\mathbf{x}^{(n)} = (x_s^{(n)})_{s\in S_?}$, *donde* $x_s^{(n)} = \Pr(s \models C\mathcal{U}^{\leq n} S_{=1})$ *para cada estado* $s \in S_?$
- $\mathbf{x}^{(0)} \leq \mathbf{x}^{(1)} \leq \mathbf{x}^{(2)} \leq \cdots \leq \mathbf{x}$, *y*
- $\mathbf{x} = \lim_{n\to\infty} \mathbf{x}^{(n)}$.

La prueba de este teorema la podemos charlar pero está en la página 762

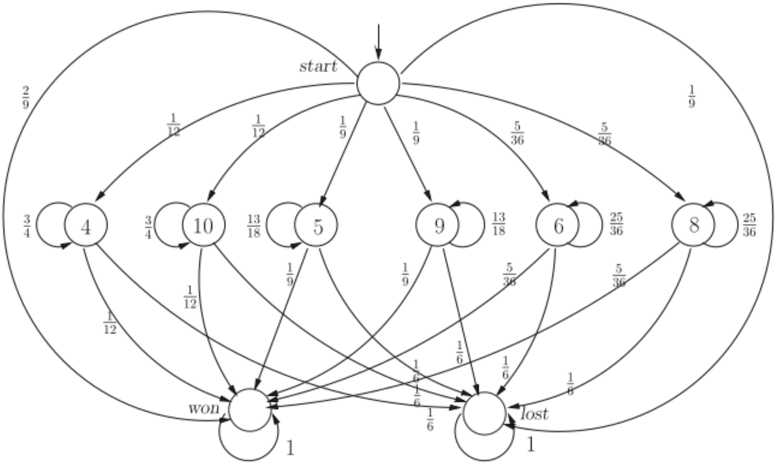
Sobre el algoritmo que se deriva del punto fijo

- Lo obtendremos calculando iterativamente los $\mathbf{x}^{(n+1)}$ y frenando cuando $\max_{s \in S_?} |x_s^{(n+1)} - x_s^{(n)}| < \varepsilon$ para alguna tolerancia ε.
- También lo podemos usar para calcular problemas de alcanzabilidad con pasos limitados.
- Aunque la convergencia de este método está asegurada, generalmente es menos eficiente que otros métodos iterativos para resolver sistemas de ecuaciones como Jacobi o Gauss-Seidel (si han cursado métodos, en otro momento podemos charlar de qué hay que tener en cuenta al usar esos métodos para la convergencia).
- Claramente, la elección de conjuntos $S_{=0}$ y $S_{=1}$ más grandes genera una ejecución más eficiente y los conjuntos cota superior con los que los describimos pueden ser calculados por algoritmos de análisis de grafos que son lineales en el tamaño de la cadena de Markov.

Un ejemplo más complejo - el juego de Craps

El juego de Craps es un juego de un jugador en el que se tiran 2 dados. La primer tirada, llamada "*come-out roll*", determina si se necesitará otra tirada. Si la primer tirada resulta en un 7 o un 11, el juego termina y el jugador gana, si resulta en un 2, 3 o 12, estos son "*craps*", el juego termina y el jugador pierde, mientras que si sale cualquier otro número el dado se tira otra vez pero el primer resultado es recordado como "el punto". Si la próxima tirada es un 7 o el punto, el juego termina. Si salió 7 el jugador pierde, si salió el punto, gana. En cualquier otro caso, se vuelve a tirar el dado hasta que salga 7 o el punto.
Intentemos armar una cadena de Markov a partir de esta descripción.

Cadena de Markov del juego de Craps



Alcanzabilidad restringida en el juego de Craps

Nos interesará saber la probabilidad de $C\mathcal{U}^{\leq n}B$ donde $B = \{won\}$ y $C = \{start, 4, 5, 6\}$. Con lo que podríamos tener:

$$S_{=0} = \{8, 9, 10, lost\},\ S_{=1} = \{won\},\ \text{y}\ S_{?} = \{start, 4, 5, 6\}$$

Usando el orden de estados $start < 4 < 5 < 6$, la matriz **A** y el vector **b** están dados por

$$\mathbf{A} = \frac{1}{36}\begin{pmatrix} 0 & 3 & 4 & 5 \\ 0 & 27 & 0 & 0 \\ 0 & 0 & 26 & 0 \\ 0 & 0 & 0 & 25 \end{pmatrix} \qquad \mathbf{b} = \frac{1}{36}\begin{pmatrix} 8 \\ 3 \\ 4 \\ 5 \end{pmatrix}.$$

Alcanzabilidad restringida en el juego de Craps - cont.

La caracterización nos propone seguir el siguiente esquema iterativo:

$$\mathbf{x}^{(0)} = \mathbf{0} \text{ y } \mathbf{x}^{(i+1)} = \mathbf{A}\mathbf{x}^{(i)} + \mathbf{b} \text{ para } 0 \leq i < n,$$

donde $\mathbf{x}^i$ guarda para todo estado $s \in S_?$ la probabilidad del evento $C\mathcal{U}^{\leq n}B$. Aplicando este esquema iterativo tenemos que $\mathbf{x}^{(1)} = \mathbf{b}$ y

$$\mathbf{x}^{(2)} = \frac{1}{36}\begin{pmatrix} 0 & 3 & 4 & 5 \\ 0 & 27 & 0 & 0 \\ 0 & 0 & 26 & 0 \\ 0 & 0 & 0 & 25 \end{pmatrix} \cdot \frac{1}{36}\begin{pmatrix} 8 \\ 3 \\ 4 \\ 5 \end{pmatrix} + \frac{1}{36}\begin{pmatrix} 8 \\ 3 \\ 4 \\ 5 \end{pmatrix} = \left(\frac{1}{36}\right)^2 \begin{pmatrix} 338 \\ 189 \\ 248 \\ 305 \end{pmatrix}$$

Luego, por ejemplo $\Pr(\mathit{start} \models C\mathcal{U}^{\leq 2}B) = \frac{338}{36}$. De la misma manera se pueden calcular $\mathbf{x}^{(3)}$, $\mathbf{x}^{(4)}$, etc.

Clasificación de estados

A partir de los cálculos que vimos, existe una manera de clasificar a los estados.

- Un estado s se dice **transitorio** si hay una probabilidad positiva de abandonarlo para siempre. Es decir, si $1 - \Pr(s \models \Diamond s) > 0$.
- Un estado s es **recurrente** si resulta casi seguro que se vuelve a él. Es decir, si $\Pr(s \models \Diamond s) = 1$.

Por otro lado, diremos también que:

- un estado s es **absorbente** si $\mathbf{T}(s, s) = 1$ y $\mathbf{T}(s, s') = 0$ para todo $s' \neq s$.
- un estado s' es **accesible desde** s si $\Pr(s \models \Diamond s') > 0$.
- dos estados s y s' están **comunicados** si s' es accesible desde s y viceversa.

Clasificación de estados - cont.

A su vez, un **estado recurrente** s se dice de **período** δ cuando:

$$\delta = \mathrm{mcd}\{n \geq 1 |\, \Pr(s \models \Diamond^{=n} s) > 0\} \text{ y } \delta \geq 2$$

Si $\delta = 1$ el estado s se dice **aperiódico**.

Pero, ¿cómo calculamos $\Pr(s \models \Diamond^{=n} s)$? O, en general, ¿cómo calculamos alcanzabilidad en exactamente n pasos?

¿Cómo calculamos alcanzabilidad en exactamente n pasos?

La potencia enésima de la matriz $\mathbf{A}$, i.e., la matriz $\mathbf{A}^n$ contiene las probabilidades de llegar a los estados exactamente después de n pasos, pues, la entrada de la matriz $\mathbf{A}^n(s,t)$ es igual a la suma de las probabilidades $\mathbf{T}(s_0 s_1 \ldots s_n)$ de todos los fragmentos de caminos $s_0 s_1 \ldots s_n$ con $s_0 = s$, $s_n = t$ y $s_i \in S_?$ para $0 \leq i \leq n$. Esto es:

$$\mathbf{A}^n(s,t) = \Pr(s \models S_? \mathcal{U}^{=n} t)$$

Ahora bien, si $B = \emptyset$ y $C = S$, entonces $S_{=1} = S_{=0} = \emptyset$, $S_? = S$ y $\mathbf{A} = \mathbf{T}$. La entrada $\mathbf{T}^n(s,t)$ (del la potencia enésima de $\mathbf{T}$) entonces es igual a la probabilidad de estar en el estado t luego de n pasos si empezamos desde el estado s, i.e.,

$$\mathbf{T}^n(s,t) = \Pr(s \models S \mathcal{U}^{=n} t).$$

Pero, ¿cómo calculamos alcanzabilidad en exactamente n pasos?

La probabilidad de que la cadena de Markov $\mathcal{M}$ esté en el estado t después de exactamente n pasos

$$\Theta_n^{\mathcal{M}}(t) = \iota_{init}(s) \cdot \sum_{s \in S} \mathbf{T}^n(s,t)$$

en la literatura se suele llamar *transient state probability* para el estado t y a la función $\Theta_n^{\mathcal{M}}$ se la suele llamar la *transient state distribution*.
Estos nombres podrían resultar confusos con la clasificación de estados transitorios, entonces nosotros los llamaremos probabilidad de alcanzabilidad exacta y distribución de alcanzabilidad exacta.
Al considerar $\Theta_n^{\mathcal{M}}$ como un vector, la ecuación de arriba se puede reescribir como:

$$\Theta_n^{\mathcal{M}} = \iota_{init} \cdot \mathbf{T}^n$$

Ejemplo de cálculo de alcanzabilidad exacta

Volviendo a nuestro ejemplo del protocolo simple de comunicación, ahora queremos saber cuál es la probabilidad de que el mensaje sea enviado en exactamente 4 pasos, es decir, quermeos saber cuánto es $\Pr(start \models \Diamond^{=4} B)$ cuando $B = \{delivered\}$. Tendremos que

$$\Pr(start \models \Diamond^{=4} B) = \Theta_4^{\mathcal{M}}(delivered)$$

$$\Theta_4^{\mathcal{M}} = \iota_{init} \cdot \mathbf{T}^4$$

$$\Theta_4^{\mathcal{M}} = \begin{pmatrix} 1 & 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & \frac{9}{10} & \frac{1}{100} & \frac{9}{100} \\ \frac{9}{100} & \frac{1}{100} & \frac{9}{100} & \frac{81}{100} \\ 0 & \frac{9}{10} & \frac{1}{100} & \frac{9}{100} \\ \frac{9}{10} & \frac{1}{10} & 0 & 0 \end{pmatrix}$$

$$\Theta_4^{\mathcal{M}}(delivered) = \frac{9}{100}$$

Clasificación de estados II – Irreducibilidad y periodicidad global

- Una cadena de Markov es **irreducible** si todos los pares de estados están comunicados entre sí.

$$\forall s, t \in S,\ \Pr(s \models \Diamond t) > 0 \text{ y } \Pr(t \models \Diamond s) > 0$$

- La cadena es **aperiódica** si todos sus estados tienen período igual a 1:

$$\forall s \in S,\ \mathrm{mcd}\{n \geq 1 \mid \Pr(s \models \Diamond^{=n} s) > 0\} = 1$$

¿Por qué irreducibilidad y aperiodicidad? - Distribución estacionaria

- **Irreducibilidad** asegura que es posible llegar a cualquier estado desde cualquier otro. Esto garantiza que la cadena no tenga múltiples "bloques aislados" con comportamientos distintos.
- **Aperiodicidad** evita que la cadena oscile de forma rígida entre estados. Esto es crucial para que la distribución converja a una única distribución límite.
- Estas condiciones serán condiciones suficientes para existencia, unicidad y convergencia hacia una **distribución estacionaria** o invariante de probabilidades de transición de una cadena de Markov.

Distribución estacionaria

- Una **distribución estacionaria** de una cadena de Markov con matriz de transición T es un vector de probabilidad π tal que:

$$\pi T = \pi \quad \text{y} \quad \sum_{s \in S} \pi(s) = 1$$

- Si la cadena comienza con distribución π, entonces su distribución se mantiene constante para todos los tiempos.
- Cuando existe, describe el **comportamiento a largo plazo** de la cadena.
- Para obtenerla se puede:
  - Resolver el sistema lineal $\pi T = \pi$ junto con $\sum \pi_i = 1$.
  - Tomar $\pi = \lim_{n \to \infty} \iota_{init} T^n$ si la cadena cumple las condiciones de irreducibilidad y aperiodicidad.

Ejemplo: cálculo de distribución estacionaria

Consideremos la siguiente matriz de transición:

$$T = \begin{pmatrix} 0{,}7 & 0{,}3 \\ 0{,}4 & 0{,}6 \end{pmatrix}$$

Queremos hallar $\pi = (\pi_1, \pi_2)$ tal que:

$$\pi T = \pi \quad \text{y} \quad \pi_1 + \pi_2 = 1$$

Esto da el sistema:

$$\begin{cases} 0{,}7\pi_1 + 0{,}4\pi_2 = \pi_1 \\ 0{,}3\pi_1 + 0{,}6\pi_2 = \pi_2 \\ \pi_1 + \pi_2 = 1 \end{cases} \quad \Rightarrow \pi = \left(\frac{4}{7}, \frac{3}{7}\right)$$

Ejemplo: cálculo de distribución estacionaria

Si con R o algún otro software estadístico calculamos T^n para algún n lo suficientemente grande podemos calcular una aproximación de π. Veamos esto calculando T^{50} en R:

$$T^{50} = \begin{pmatrix} 0{,}5714286 & 0{,}4285714 \\ 0{,}5714286 & 0{,}4285714 \end{pmatrix}$$

Como $\frac{4}{7} \approx 0{,}5714286$ y $\frac{3}{7} \approx 0{,}4285714$, podemos ver que la cadena converge hacia π con cualquier distribución inicial.

¿Qué vimos que podemos hacer con cadenas de Markov?

- ver la probabilidad de llegar a un estado en un conjunto B.
- ver la probabilidad de llegar a un estado en un conjunto B solo pasando antes por estados en un conjunto C.
- ver la probabilidad de llegar a un estado en un conjunto B en n o menos pasos.
- ver la probabilidad de llegar a un estado en un conjunto B en exactamente n pasos.
- ver el comportamiento a largo plazo de una cadena.