

Probabilidad y Estadística

Procesos Estocásticos y Cadenas de Markov

Bautista José Peirone

2023

Apunte Unidad VII

Índice

1	Procesos Estocásticos	2
2	Cadenas de Markov	4
2.1	¿Qué son?	4
2.2	Definiciones	4
2.3	Propiedades y Características	5
2.4	Clasificación de Estados	8
2.5	Matrices Canónicas y Fundamental	10
2.6	Distribución Límite, Invariante	13
2.7	Tiempo de Retorno	16
3	Procesos de Bernoulli	17
3.1	¿Qué son?	17
3.2	Cantidad de ensayos exitosos	17
3.3	Distribución de éxitos	18
4	Procesos de Poisson	19
4.1	¿Qué son?	19
4.2	Características de la distribución	19
4.3	Tiempos de arribo	20
5	Bibliografía	22
A	Demostraciones	23

Introducción

Los procesos estocásticos son una herramienta de la teoría de la probabilidad ampliamente utilizada en Ciencias de la Computación, Física, Química, Economía, etc, para modelar procesos aleatorios que dependen del tiempo o espacio. En unidades anteriores se estudiaron las variables aleatorias, útiles para modelar o predecir el resultado de un proceso fijo en el tiempo como por ejemplo la tirada de n monedas, mientras que con los procesos estocásticos podremos modelar la sucesión de tiradas y no solo el resultado final.

1 Procesos Estocásticos

Un proceso estocástico se define formalmente como una sucesión de variables aleatorias X_t (continuas o discretas) parametrizadas por un $t \in T$, donde T es el llamado espacio paramétrico. Las variables X_t del proceso pueden, como no, ser dependientes unas de otras o tener distribuciones distintas. Por razones obvias, es más sencillo el estudio de algunos procesos que cumplen ciertas condiciones, como por ejemplo variables independientes entre sí, con idénticas distribuciones, etc.

T por lo general toma el rol del tiempo o espacio (para nosotros será el tiempo) u otro factor que pueda determinar un cambio en el comportamiento de nuestro proceso. Además, existirá el llamado espacio de estados E definido como el conjunto de todos los posibles resultados del proceso. Cualquier resultado de una variable X_t formará parte de este espacio de estados. Se pueden categorizar procesos según si E es discreto o continuo (dependerá de si las variables X_t son discretas o continuas), o si T es discreto o continuo (dependerá de como se modifique el comportamiento del proceso en cuestión).

A modo de síntesis, dado un experimento ϵ con espacio muestral Ω , un proceso estocástico es:

$$\{X_t : \Omega \rightarrow \mathbb{R}, t \in T\}$$

Otra forma de pensar los procesos estocásticos es tomar en vez de infinitas variables aleatorias una sola variable aleatoria de dos argumentos. En este caso tendríamos $X : \Omega \times T \rightarrow \mathbb{R}$, y entre ambas interpretaciones vale que $X_t(\omega) = X(\omega, t)$.

El espacio de estados antes descripto tiene la siguiente definición: $\{X_t(\omega) | \omega \in \Omega \wedge t \in T\}$

Fijando el suceso tenemos una variable parametrizada por T . Esto es, para un suceso en particular ω_0 , $X_t(\omega_0)$ equivale al estado del proceso en el momento t . Esto es totalmente determinista, ya que el suceso ya está definido, y solo veremos como evoluciona con el paso del tiempo. De forma análoga, para t_0 fijo, $X_{t_0}(\omega)$ es la variable aleatoria que ocurre en el instante t . Esto ya no es determinista, porque depende del suceso aleatorio ω .

Como detalle final, los siguientes ejemplos muestran la amplia variedad que pueden abarcar los procesos estocásticos, y las formas con que se modela cada uno.

1. T discreto y E discreto

Estos problemas suelen modelarse con las llamadas Cadenas de Markov o Procesos de Bernoulli. Un ejemplo de aplicación es contar cuantas caras ocurren en varios lanzamientos de moneda.

2. T discreto y E continuo

Pueden modelarse con sucesiones de variables aleatorias continuas. Ejemplo de aplicación podría ser medir la cantidad de lluvia a lo largo de los días.

3. T continuo y E discreto

Se utilizan los procesos puntuales, como los Procesos de Poisson. Estos procesos dan pie a la llamada Teoría de Colas.

4. T continuo y E continuo

Se utilizan los procesos continuos para problemas como medir la velocidad del viento a cada instante.

2 Cadenas de Markov

2.1 ¿Qué son?

Las cadenas de Markov son un modelo ampliamente utilizado en Machine Learning para simulación de probabilidades (el algoritmo PageRank de Google se basa en estas).

Estas cadenas modelan procesos estocásticos con las siguientes propiedades:

- Espacio paramétrico T infinito y numerable (discreto)
- Espacio de estados E discreto
- Propiedad Markoviana: El próximo estado solo depende del estado actual del sistema y de nada más.

La propiedad Markoviana es la más importante sin duda ya que agrega una condición muy fuerte al problema pero a su vez es muy útil. En un sistema con esta propiedad podríamos decir que hay una falta de memoria, o que el futuro solo esta determinado por el presente y no por el pasado. La forma de expresar formalmente esta propiedad es $\forall n \in T \forall j_1, j_2, \dots, j_n \in E$ vale:

$$P(X_{n+1} = j_{n+1} | X_1 = j_1, X_2 = j_2, \dots, X_n = j_n) = P(X_{n+1} = j_{n+1} | X_n = j_n)$$

Siguiendo con la analogía, X_{n+1} es el futuro, X_n es el presente, y X_1, X_2, \dots, X_{n-1} es el pasado.

NOTA SOBRE LA PROPIEDAD MARKOVIANA: La interpretación propia que le doy a esta propiedad es que solo la información más reciente es útil. Luego, en el momento n , no solo se puede conocer las probabilidades del instante $n+1$ sino también las del momento $n+2$ ya que ahora tenemos información sobre el nuevo presente que es $n+1$, y así se puede "reconstruir" hasta cualquier instante futuro $n+m$.

Al tener espacio de estados y paramétrico discretos podemos pensar $X_n = i$ como estar en el estado i -ésimo en el instante n , y nos será de particular interés en las cadenas de Markov las siguientes probabilidades, tanto que le asignamos la notación

$$P(i, j) := P(X_{n+1} = j | X_n = i)$$

para $i, j \in E$.

2.2 Definiciones

La matriz cuadrada P de $n \times n$ tal que $P_{ij} = P(i, j)$ se llama matriz de probabilidad de transiciones o matriz de transición, donde $n = |E|$. Esta matriz tiene varias propiedades

importantes, destacando dos principales:

$$P(i, j) \geq 0, \sum_{j \in E} P(i, j) = 1, \forall i, j \in E$$

CONVENCIÓN DE NOTACIÓN: Las matrices A_{ij} se notarán como $A(i, j)$. De la misma manera, la componente π_j de un vector se escribirá como $\pi(j)$.

Si una matriz cumple estas propiedades la llamamos matriz estocástica. Aun más, un vector fila que cumple con estas propiedades se llama vector de probabilidades, y no es difícil observar que una matriz estocástica se conforma en cada fila por un vector de probabilidades. Una distribución inicial es un vector de probabilidades que representa las probabilidades de que X_0 esté en cada estado. En otras palabras, si π_0 es una distribución inicial, entonces $P(X_0 = i) = \pi_0(i)$

El diagrama de transiciones de una cadena de Markov es un grafo dirigido ponderado, con un vértice por cada estado $i \in E$, y donde para cualquier par de vértices i, j existe una arista de i a j con peso $P(i, j)$. Si $P(i, j) = 0$ no se suele dibujar la arista.

2.3 Propiedades y Características

Las cadenas de Markov cuentan con varias propiedades para el cálculo de probabilidades de evolución de estados. Esto es, es sencillo calcular la probabilidad de transicionar entre cualquier par de estados, en cualquier número de pasos, incluso incluyendo condiciones intermedias (por ejemplo, la probabilidad de transicionar del estado 1 a 5 en 10 pasos a y pasando a mitad de camino por el estado 2). Las propiedades mas destacables son:

1. $P(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = P(i_{n-1}, i_n) \cdot \dots \cdot P(i_0, i_1) \cdot P(X_0 = i_0) \forall n \in \mathbb{N}, \forall i_0, i_1, \dots, i_n \in E$

Lo que nos dice esto es que si conocemos todas las probabilidades de transición y la distribución inicial, entonces podemos calcular de manera exacta la probabilidad con que se dará cada posible evolución de X .

2. $\forall n, m \in \mathbb{N}_0, \forall i, j \in E, P(X_{m+n} = j | X_m = i) = P^n(i, j)$.

Esto es, la probabilidad de transicionar en n pasos de i a j es igual a la suma de las probabilidades de todos los caminos de longitud n que van de i a j . Esta suma de probabilidades de caminos es lo que $P^n(i, j)$ representa, surge de un resultado común de las matrices de adyacencia de la Teoría de Grafos.

3. $\forall n, m \in \mathbb{N}_0 \forall i, j \in E, P^{n+m}(i, j) = \sum_{k \in E} P^n(i, k) \cdot P^m(k, j)$

Esta propiedad nos dice que la probabilidad de una transición en $n + m$ pasos se puede separar en $|E|$ pares de transiciones, donde en cada par la primera transición es en n

Probabilidad y Estadística

Unidad VII

LCC - FCEIA - UNR

5to cuatrimestre

pasos desde i hasta k , y la segunda va desde k hasta j en m pasos. Esta propiedad lleva el nombre de Chapman-Kolmogorov.

4. Sea π_0 un vector de probabilidades, luego $P(X_n = i) = \pi_n(i)$ donde $\pi_n = \pi_0 \cdot P^n$

A continuación se definen varios conceptos importantes, asociados con probabilidades entre transiciones de estados. Sea $X = \{X_n : n \in \mathbb{N}_0\}$ una cadena de Markov.

Si $j \in E$, llamamos N_j a la variable aleatoria que cuenta cuantas veces se pasa por el estado j en una sucesión infinita de estados de la cadena X . Esto es, para un evento $\omega \in \Omega$, $N_j(\omega) = m \in \mathbb{N}_0$ si y solo si en el suceso ω se pasó exactamente m veces por j .

Sea $n \in \mathbb{N}$, $F_n(i, j)$ es la probabilidad de llegar a j desde i por **primera vez** en exactamente n pasos. Es importante notar que $F_n(i, j)$ difiere de $P^n(i, j)$ en que este último representa la probabilidad de transicionar desde i a j en n pasos sin ninguna restricción sobre la cantidad de veces que podemos pasar j . Luego

$$F_1(i, j) = P(i, j)$$
$$F_n(i, j) = \sum_{b \in E \setminus \{j\}} P(i, b) \cdot F_{n-1}(b, j) \quad \text{si } n \geq 2$$

La idea de esta definición de F_n es que todas representen entre sí probabilidades de eventos disjuntos, para así poder definir

$$F(i, j) = \sum_{n=1}^{\infty} F_n(i, j)$$

$F(i, j)$ representa la probabilidad de transicionar desde i a j en una cantidad finita de pasos, y $1 - F(i, j)$ es entonces la probabilidad de nunca llegar a j partiendo de i . Una propiedad que se tiene es: $F(i, j) = P(i, j) + \sum_{b \in E \setminus \{j\}} P(i, b) \cdot F(b, j)$
La demostración se puede encontrar en [A](#).

Con la definición de $F(i, j)$ podemos ver que $\forall i, j \in E, i \neq j$

$$P(N_j = m | X_0 = i) = \begin{cases} F(i, j)F(j, j)^{m-1}(1 - F(j, j)) & \text{si } m \geq 1 \\ 1 - F(i, j) & \text{si } m = 0 \end{cases}$$
$$P(N_j = m | X_0 = j) = \begin{cases} F(j, j)^{m-1}(1 - F(j, j)) & \text{si } m \geq 1 \\ 0 & \text{si } m = 0 \end{cases}$$

Además, nos interesará conocer en general la probabilidad de transitar un estado una cantidad finita de pasos. Esto lo podemos expresar como

$$P(N_j < \infty | X_0 = j) = \sum_{m=1}^{\infty} P(N_j = m | X_0 = j) = \sum_{m=1}^{\infty} F(j, j)^{m-1}(1 - F(j, j))$$
$$= (1 - F(j, j)) \sum_{m=1}^{\infty} F(j, j)^{m-1}$$

Probabilidad y Estadística

Unidad VII

LCC - FCEIA - UNR

5to cuatrimestre

Luego, si $F(j, j) = 1$, es directo que $P(N_j < \infty | X_0 = j) = 0$, ya que siempre volveremos a j en una cantidad finita de pasos, lo que se traduce a que pasaremos infinitas veces por j considerando una cantidad infinita de transiciones. Por otro lado, si $F(j, j) < 1$, la serie geométrica $\sum_{m=1}^{\infty} F(j, j)^{m-1}$ converge a $\frac{1}{1-F(j, j)}$, luego $P(N_j < \infty | X_0 = j) = (1 - F(j, j)) \sum_{m=1}^{\infty} F(j, j)^{m-1} = 1$. Para sintetizar

$$P(N_j < \infty | X_0 = j) = \begin{cases} 1 & \text{si } F(j, j) < 1 \\ 0 & \text{si } F(j, j) = 1 \end{cases}$$

Por otro lado, es de interés calcular la cantidad esperada de visitas a j partiendo desde i . Este número lo denotamos $R(i, j)$, y sabiendo que $\sum_{m=0}^{\infty} m F(j, j)^{m-1} = \frac{1}{(1 - F(j, j))^2}$, si $i \neq j$:

$$\begin{aligned} R(i, j) &= E(N_j | X_0 = i) = \sum_{m=0}^{\infty} m P(N_j = m | X_0 = i) \\ &= (1 - F(j, j)) F(i, j) \sum_{m=0}^{\infty} m F(j, j)^{m-1} = \frac{F(i, j)}{1 - F(j, j)} \\ R(j, j) &= E(N_j | X_0 = j) = \sum_{m=0}^{\infty} m P(N_j = m | X_0 = j) \\ &= (1 - F(j, j)) \sum_{m=0}^{\infty} m F(j, j)^{m-1} = \frac{1}{1 - F(j, j)} \end{aligned}$$

Luego podemos observar propiedades como que $R(i, j) = F(i, j) R(j, j)$ y si $F(j, j) = 1$, entonces $R(j, j) = \infty$

Si ahora se define la variable aleatoria

$$Y_m = \begin{cases} 1 & \text{si } X_m = j \\ 0 & \text{si } X_m \neq j \end{cases}$$

entonces $N_j = \sum_{m=0}^{\infty} Y_m$, y

$$E(Y_m | X_0 = i) = \sum_{k=0}^1 k \cdot P(Y_m = k | X_0 = i) = P(Y_m = 1 | X_0 = i) = P(X_m = j | X_0 = i)$$

por lo que

$$\begin{aligned} R(i, j) &= E(N_j | X_0 = i) \\ &= \sum_{m=0}^{\infty} E(Y_m | X_0 = i) \\ &= \sum_{m=0}^{\infty} P(X_m = j | X_0 = i) = \sum_{m=0}^{\infty} P^m(i, j) \end{aligned}$$

y además resulta que $R = (I - P)^{-1}$ cuando esta última exista. Otra consecuencia inmediata es que si $R(i, j)$ es finito, entonces $\sum_{m=0}^{\infty} P^m(i, j)$ converge, lo que implica $P^m(i, j) \rightarrow 0$ cuando $m \rightarrow \infty$.

2.4 Clasificación de Estados

Estados recurrentes y transitorios

Decimos que un estado j es recurrente si $F(j, j) = 1$. Caso contrario, si $F(j, j) < 1$ se dice que j es un estado transitorio. Los estados recurrentes se dividen a su vez en dos categorías, los positivamente recurrentes y nulo recurrentes. Un estado j es positivamente recurrente si y solo si el tiempo promedio entre dos visitas consecutivas a j es finito, mientras que en los estados nulo recurrentes este tiempo promedio es infinito.

Un estado transitorio tiene la propiedad de que solo se transita un número finito de veces, mientras que cuando un estado recurrente es visitado, entonces se lo volverá a visitar una cantidad infinita de veces. Esto es así porque un estado j es transitorio si $F(j, j) < 1$, luego $1 - F(j, j) > 0$ por lo que la probabilidad de no volver a j es no nula, lo cual significa que a la larga dejaremos de pasar por este.

Deriva de aquí que el número de visitas esperado a j es también finito cuando j es transitorio, o sea que $R(j, j) < \infty$. Además, si $i \neq j$, sabemos que $R(i, j) = F(i, j)R(j, j) \leq R(j, j)$, luego $R(i, j) < \infty$. A su vez, si j es recurrente, $R(j, j) = \infty$.

El siguiente teorema nos da un resultado importante que relaciona la recurrencia y/o transitividad de un estado con las probabilidades de a la larga pasar por ellos.

TEOREMA:

- Si j es transitivo o recurrente nulo, entonces $\forall i \in E, \lim_{n \rightarrow \infty} P^n(i, j) = 0$. De no ser así, a la larga terminaríamos regresando a j en una cantidad finita de pasos.
- Si j es recurrente positivo y aperiódico (véase más adelante su definición) e i está en algún conjunto cerrado con j , entonces vale que $\lim_{n \rightarrow \infty} P^n(i, j) > 0$. Más aún, para cualquier $i \in E$ se tiene $\lim_{n \rightarrow \infty} P^n(i, j) = F(i, j) \cdot \lim_{n \rightarrow \infty} P^n(j, j)$.

Absorción

Un estado j es absorbente si $P(j, j) = 1$. Esto implica que $\forall i \neq j, P(j, i) = 0$. En el grafo, un estado absorbente se caracteriza por tener un lazo en j y ningún otro arco saliente de j .

Accesibilidad

Decimos que j es accesible desde i si $\exists n \in \mathbb{N}_0 : P^n(i, j) > 0$. Esto es, j es accesible desde i

Probabilidad y Estadística

Unidad VII

LCC - FCEIA - UNR

5to cuatrimestre

si hay algún camino dirigido en el grafo de transiciones desde i hasta j de cualquier longitud finita. Lo simbolizamos como $i \rightarrow j$.

Si i es accesible desde j y j es accesible desde i entonces se dice que i y j están comunicados, y se simboliza $i \leftrightarrow j$.

El siguiente teorema relaciona la alcanzabilidad desde un estado recurrente.

TEOREMA: Sean $i, j \in E$ tales que $i \rightarrow j$ y además i es recurrente. Entonces j es recurrente. La intuición de por qué esto es cierto es sencilla. Si j fuese transitivo, entonces es porque eventualmente no se vuelve a j , y por ende no se debe poder volver a i . Pero i es recurrente, luego se tiene un absurdo.

Conjuntos cerrados e irreducibles

Un conjunto $C \subseteq E$ es cerrado si ningún estado en $E \setminus C$ es accesible desde algún estado en j . Puesto algebraicamente:

$$i \in C \wedge j \in E \setminus C \implies \neg(i \rightarrow j)$$

El conjunto E es por ende siempre cerrado.

Un conjunto cerrado C se dice irreducible si $\forall S \subset C$, S no es cerrado. Cuando E sea irreducible, diremos que la cadena es irreducible, y esta propiedad se refleja en el grafo cuando para cada par de estados i, j se tiene $i \rightarrow j$.

Una propiedad importante que surge es: Si $C \subseteq E$ cerrado e irreducible y finito, entonces $\forall i \in C$, i es positivamente recurrente. Surge de aquí que para una cadena con cantidad finita de estados, no hay estados recurrentes nulos y hay algún estado positivamente recurrente (o de forma equivalente, no todos los estados pueden ser transitivos).

Los siguientes dos resultados relacionan los conjuntos irreducibles con los estados recurrentes.

LEMA: Sea $j \in E$ un estado recurrente, luego existe $C \subseteq E$ irreducible tal que $j \in C$.

TEOREMA: Sea $C = \{j \in E : j \text{ es recurrente}\}$. C se puede particionar de manera única en conjuntos irreducibles. Además, si C_k es una componente de esta partición, entonces C_k es una cadena de Markov irreducible. A cada una de estas C_j las llamaremos clases de recurrencia ya que forman una clase de equivalencia bajo la relación de alcanzabilidad.

TEOREMA: Si se tiene una cadena de Markov irreducible, entonces sus estados son todos recurrentes o todos transitivos. En cada caso decimos que la cadena es recurrente o transitiva respectivamente.

Habiendo dado estas definiciones, podemos relacionar los estados recurrentes según si pertenecen

a un mismo conjunto irreducible

TEOREMA:

- Sean i, j estados recurrentes que pertenecen al mismo conjunto irreducible, entonces $F(i, j) = 1$.
- Sean i, j estados recurrentes que pertenecen a distintos conjuntos irreducibles, entonces $F(i, j) = 0$
- Sea i un estado transitivo y $C \subset E$ un conjunto irreducible. Entonces $F(i, j) = F(i, k) \forall j, k \in C$

COROLARIO: Tenemos que si existe una única clase recurrente C e i es un estado transitivo del cual solo un numero finito de estados transitivos pueden ser alcanzados, luego $F(i, j) = 1 \forall j \in C$.

Período

Sea j un estado recurrente, se define $\delta = \text{mcd} \{n \geq 1 : P^n(j, j) > 0\}$. Si $\delta \geq 2$ entonces decimos que j es un estado periódico con período δ . Por otro lado, cuando $\delta = 1$ decimos que j es un estado aperiódico.

Si se tiene una cadena irreducible y un estado de período δ , entonces todos los estados tienen período δ y se dice que la cadena es periódica de período δ . Si un estado es aperiódico, todos los estados lo serán y se dice que la cadena es aperiódica.

Ergodicidad

Sea X una cadena con cantidad finita de estados. Se dice que X es ergódica si y solo si X es irreducible, recurrente y aperiódica. Notemos que pedir que es recurrente aquí no es necesario, ya que esto se puede deducir de la irreducibilidad y cantidad finita de sus estados.

2.5 Matrices Canónicas y Fundamental

Si i es recurrente y j transitivo, entonces es un hecho que $P(i, j) = 0$, luego si tomamos un orden de los estados $i_1, \dots, i_n, j_1, \dots, j_m$ con i_k recurrente y j_k transitivo, donde n es la cantidad de estados recurrentes y m la de transitivos, si reescribimos P en este orden, se tendrá un bloque de ceros en la región superior derecha. Esto es:

$$P = \left(\begin{array}{c|c} \text{Transiciones entre} & \\ \text{recurrentes} & 0 \\ \hline \text{Transiciones de} & \text{Transiciones} \\ \text{transitivos a recurrentes} & \text{entre transitivos} \end{array} \right)$$

Podemos tomar esta idea de reorganizar la matriz de transiciones aun más al extremo. Una observación muy importante para esto es ver que, en cierto sentido, las clases de recurrencia se comportan como estados absorbentes. No se puede salir de los estados absorbentes al igual que de las clases de recurrencia, y entonces no podemos ir de una clase de recurrencia a otra. Además, para pasar de un estado transitivo a una clase de recurrencia debemos transicionar a cualquier estado de dicha clase. Esta relación entre clases de recurrencia y estados absorbentes se puede observar en los siguientes grafos.

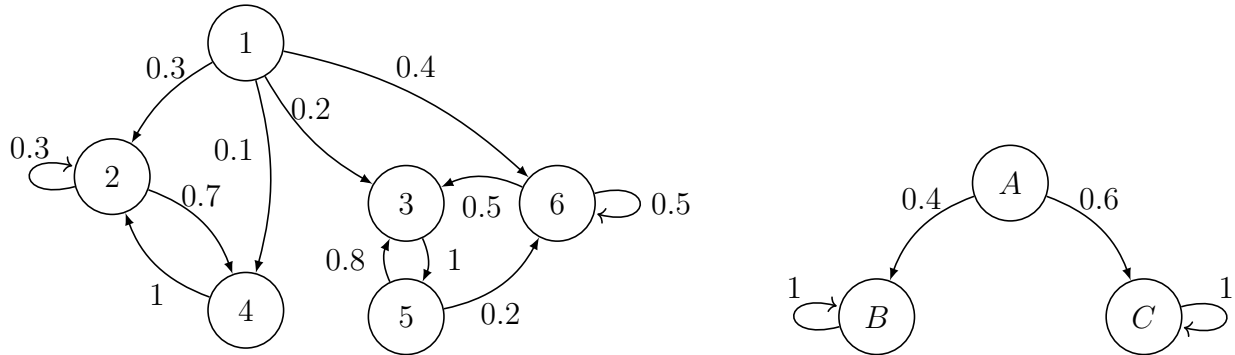


Figura 1: Equivalencia de clases de recurrencia con estados absorbentes

En los grafos de la figura 1 vemos una correspondencia entre los estados 1 y A, la clase de recurrencia $\{2, 4\}$ con B, y $\{3, 5, 6\}$ con el estado C.

A raíz de esto, podemos hacer más sencilla la matriz de transición, considerando clases de conjuntos recurrentes en vez de estados recurrentes. Las clases de recurrencia nos permitirán simplificar cierta información, como media de visitas entre estados de la misma clase, probabilidad de transicionar en cantidad finita de pasos, etc. Para lograrlo, sea D el conjunto de los estados transitivos de una cadena, $m = |D|$ y C_1, \dots, C_n los conjuntos irreducibles de estados recurrentes, donde n es la cantidad de clases de recurrencia.

Se define

$$P' = \begin{pmatrix} I_{n \times n} & \mathbf{0}_{n \times m} \\ B & Q \end{pmatrix}$$

con dimensiones $(n + m) \times (n + m)$ como la matriz canónica de transición o forma canónica de P . La matriz Q es una matriz de $m \times m$, y es la submatriz de P que consta de las transiciones entre estados transitivos solamente.

B es una matriz de $m \times n$ definida como

$$B(i, C_j) = \sum_{k \in C_j} P(i, k)$$

con $i \in D$ y C_j corresponde a una clase de recurrencia, y es la probabilidad de transicionar a la clase C_j desde i en un paso.

Probabilidad y Estadística

Unidad VII

LCC - FCEIA - UNR

5to cuatrimestre

Más aun, a todas las matrices $(n+m) \times (n+m)$ con entradas solo para clases de recurrencia y estados transitivos diremos que están en su forma canónica. Se estudiará como obtener las formas canónicas de R y F , lo que nos permitirá la reconstrucción de ambas. Para esto, hablaremos de una matriz muy importante.

DEFINICIÓN: La matriz S de $m \times m$ definida como $S = (I - Q)^{-1}$ se llama matriz fundamental.

S tiene la propiedad de que indica la media de visitas entre dos estados transitivos, osea que ¡¡forma parte de la matriz R !! En concreto, se tiene que la forma canónica de R es:

$$R' = \begin{pmatrix} \infty \cdot I_{n \times n} & \mathbf{0}_{n \times m} \\ Y & S \end{pmatrix}$$

¿Quién debe ser aquí Y ? Sabemos que debe representar la media de visitas de un estado transitivo a un clase de recurrencia. Luego, si i es transitivo y C_j es una clase de recurrencia, tendremos que la media es 0 si la clase C_j no es alcanzable desde i , y sino ∞ . Con R' podemos reconstruir fácilmente R . ¿Cómo? La media de visitas entre transitivos sigue siendo la misma, S . La media de visitas de estados recurrentes a transitivos será 0, mientras que de recurrentes a recurrentes será ∞ si y solo si pertenecen ambos a la misma clase, 0 de lo contrario. Por ultimo, similar a Y tendremos que de un estado transitivo i a uno recurrente j la media es ∞ si y solo si $i \rightarrow j$, cuando no valga esto la media será 0. Para sintetizar, el procedimiento para hallar R dado R' es ampliar las matrices del lado izquierdo para considerar estados recurrentes y no clases de recurrencia. Para lograr esto, si C_k es una clase de recurrencia y $j \in C_k$, entonces $R(i, j) = R'(i, C_k)$.

La forma de F es parecida a la de R , y la forma de obtenerla también lo es. Su forma canónica tiene la siguiente forma:

$$F' = \begin{pmatrix} I_{n \times n} & \mathbf{0}_{n \times m} \\ G & Z \end{pmatrix}$$

donde G de $m \times n$ es $G = SB$ y cuenta con la propiedad de que $G(i, j) = F(i, k)$, $\forall k \in C_j$. Falta encontrar Z , para esto debemos ver la probabilidad de transicionar de un estado transitivo a otro en un número finito de pasos. Para esto haremos uso de R , y sean i, j transitivos. Por ser transitivo j , vale $R(j, j) < \infty$, luego aprovecharemos la propiedad de que $R(i, j) = F(i, j)R(j, j) \implies F(i, j) = \frac{R(i, j)}{R(j, j)}$. De esta forma hemos hallado la forma canónica de F , y para obtenerla basta con repetir el procedimiento hecho para pasar de R' a R , expandiendo los bloques de F' .

Por último, sea T_A la variable aleatoria que mide el tiempo de absorción. Esta variable cuenta la cantidad de transiciones hasta caer en una clase de recurrencia, y puede resultar interesante conocer la esperanza de dicha variable, ya que mide la cantidad que se pasa en

los estados transitivos. Tiene sentido entonces pensar que la media de tiempo de absorción partiendo del estado i sea la media de visitas a todos los demás estados transitivos. Recordando que $S(i, j)$ marca para dos estados transitivos i, j la media de visitas a j desde i , el siguiente teorema formaliza este resultado:

TEOREMA:

Sean i, j estados transitivos, S la matriz fundamental, y D el conjunto de estados transitivos, entonces $\forall i \in D$

$$E(T_A | X_0 = i) = \sum_{j \in D} S(i, j)$$

2.6 Distribución Límite, Invariante

Una característica importante de algunas distribuciones es que no son modificadas tras una transición. Estas distribuciones pueden resultar útiles cuando se quiere mantener el estado de algunos sistemas, por ejemplo proporciones o cantidades. Para esto, buscamos π tal que $\pi = \pi P$. Notemos también que cuando una distribución cumpla esto, valdrá inductivamente que $\pi = \pi P^n$ para $n \in \mathbb{N}_0$. Esta distribución es nada más ni nada menos que un autovector a izquierda de la matriz P asociado al autovalor 1. Además, como queremos una distribución, i.e. un vector de probabilidades, queremos que la suma de sus componentes valga 1.

DEFINICIÓN: A partir de esto, se dice que π es una distribución invariante si y solo si verifican

$$\begin{cases} \pi(j) &= \sum_{i \in E} \pi(i)P(i, j), \forall j \in E \\ \sum_{j \in E} \pi(j) &= 1 \end{cases}$$

La primera condición impone la condición de "invarianza", mientras que la segunda asegura que π es un vector de probabilidades.

DEFINICIÓN: Cuando $\pi(j) = \lim_{n \rightarrow \infty} P^n(i, j) > 0$ exista para cada $i, j \in E$ y sea independiente de i , entonces diremos que π es una distribución límite. π resulta ser una distribución invariante de P y único cuando existe.

TEOREMA:

Si la cadena es irreducible y aperiódica, entonces sus estados son positivamente recurrentes si y solo si existe distribución invariante¹.

¹Esto muestra una relación entre la recurrencia positiva de los estados y el 1 como autovalor de P

Probabilidad y Estadística

Unidad VII

LCC - FCEIA - UNR

5to cuatrimestre

COROLARIO: Sigue del teorema anterior que cuando la cantidad de estados es finita, como sus estados son positivamente recurrentes, entonces existe dicha distribución π y además vale que esta es la distribución límite de la cadena.

COROLARIO: Una cadena de Markov ergódica tiene distribución límite.

Hablemos un poco más de la distribución límite. ¿Qué propiedades tiene esta?

Cuando la distribución límite π existe, la matriz de transición en n pasos tiende a una matriz donde cada fila de esta corresponde con π . Es fácil ver entonces que si π_0 es un vector de probabilidades, $\pi_0 P^n$ tiende a π , y también de aquí podemos observar se deriva el hecho de que π es una distribución invariante (se puede encontrar demostrado en [A](#)). Ahora, cuando π no exista, ¿qué sucede con $\lim_{n \rightarrow \infty} P^n$? Para esto pensemos en qué sucede cuando algunas de las hipótesis no se cumplen:

Supongamos una cadena periódica de período δ , entonces P^n no convergerá debido a la oscilación de las probabilidades de los estados.

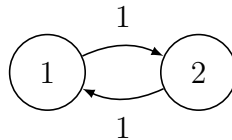


Figura 2: Cadena de Markov de período 2

La cadena 2 tiene período 2, y su matriz de transición es $P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Luego, se puede ver que para $n \geq 0$, $P^{2n} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ y $P^{2n+1} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, por lo tanto $\lim_{n \rightarrow \infty} P^n$ diverge y no hay comportamiento límite.

Por otro lado, si la cadena no es irreducible, entonces habrá un claro problema, y es que el como las probabilidades se distribuyen a largo plazo dependerá de como se han distribuido inicialmente. Esto es, la distribución límite pasará a depender de π_0 . En estos casos $\lim_{n \rightarrow \infty} P^n$ no tendrá cada fila igual, ya que justamente la probabilidad de estar en cualquier estado ahora **sí** depende del estado inicial.

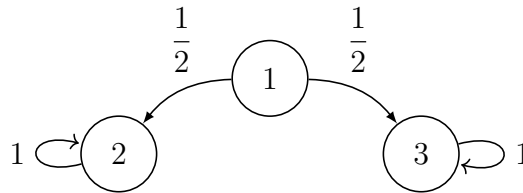


Figura 3: Cadena de Markov irreducible

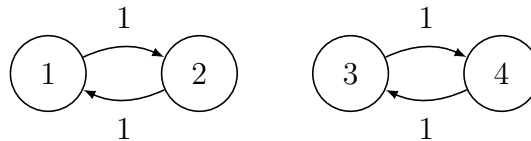
En este caso, la matriz de transición es $P = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, con la peculiaridad de que $P = P^n \forall n > 0$, luego $\lim_{n \rightarrow \infty} P^n = P$. Notemos entonces que si bien la matriz converge, tendremos que para cada π_1, π_2, π_3 definidos como

$$\pi_1 = (1, 0, 0), \pi_2 = (0, 1, 0), \pi_3 = (0, 0, 1)$$

y tenemos que $\pi_1 \cdot \lim_{n \rightarrow \infty} P^n = (0, 0.5, 0.5)$, $\pi_2 \cdot \lim_{n \rightarrow \infty} P^n = (0, 1, 0)$, y $\pi_3 \cdot \lim_{n \rightarrow \infty} P^n = (0, 0, 1)$. Osea, el comportamiento límite ha cambiado según como comenzaron inicialmente distribuidas las probabilidades. Esto no sucede cuando la cadena es irreducible, ya que todos los estados están comunicados.

Además, es importante resaltar la diferencia entre distribución límite y distribución invariante. La distribución límite es única y es además una distribución invariante. Cuando la cadena es irreducible, aperiódica y finita, ambas coinciden (ya que hay solo una distribución invariante). Si la cadena no es aperiódica, no existe distribución límite, y si la cadena no es irreducible entonces podemos encontrar infinitas distribuciones invariantes. La distribución límite es útil para analizar las probabilidades a largo plazo y como tenderán a estabilizarse.

Por ejemplo, si se tiene la cadena de Markov siguiente:



En esta cadena se tiene que no hay distribución límite, y sin embargo hay infinitas distribuciones invariantes, ya que cualquier vector de la forma:

$$\pi = \left(\frac{1}{2}\alpha, \frac{1}{2}\alpha, \frac{1}{2}\beta, \frac{1}{2}\beta \right), \alpha, \beta \geq 0, \alpha + \beta = 1$$

lo es.

Por último, si j es un estado positivamente recurrente y aperiódico en una clase de recurrencia C_j entonces podemos asegurar la existencia de $\lim_{n \rightarrow \infty} P^n(i, j)$ para todo $i \in C_j$

2.7 Tiempo de Retorno

Sea X una Cadena de Markov con espacio de estados E . Tomemos $j \in E$. ¿Cuánto podemos esperar en promedio para retornar a j ? En términos del espacio paramétrico esto es, ¿cuántos pasos discretos deben suceder de media hasta volver a j ?

Esta pregunta se puede reformular planteando la variable aleatoria $T_j : \min \{n \geq 1 : X_n = j\}$, y nos interesa conocer $E(T_j | X_0 = j)$. El siguiente teorema relaciona la distribución límite de una cadena ergódica con el tiempo esperado de retorno.

TEOREMA:

Si se tiene una cadena de Markov ergódica X , con espacio de estados E , entonces:

$$E(T_j | X_0 = j) = \frac{1}{\pi(j)}$$

donde π corresponde a la distribución límite de la cadena.

Lo que este teorema nos dice es bastante intuitivo, a medida que aumenta la probabilidad estacionaria de estar en j , disminuye el tiempo esperado para pasar por dicho estado y viceversa.

3 Procesos de Bernoulli

3.1 ¿Qué son?

Un proceso de Bernoulli es un caso particular de un proceso estocástico. Este tipo de procesos se usa para modelar sucesiones infinitas de ensayos, con las siguientes características:

- Las variables aleatorias del proceso representan ensayos de éxito (E) o fracaso (F).
- Los ensayos son independientes entre sí (las variables son entonces independientes entre sí).
- Los ensayos se realizan bajo condiciones idénticas (las variables están entonces idénticamente distribuidas).

En otras palabras, todas las variables del proceso siguen la distribución de Bernoulli con el mismo parámetro p . Se tiene luego que el espacio paramétrico es \mathbb{N} que representará el número del intento y un espacio de estados discreto, éxito o fracaso.

De aquí en adelante, $p = P(E)$ será la probabilidad del éxito y $q = 1 - p = P(F)$ la del fracaso, y podremos entonces afirmar que $X_n \sim \text{Be}(p) \forall n \in \mathbb{N}$.

El espacio muestral luego consiste de todas las sucesiones posibles de éxitos o fracasos, y la X_i dependerá entonces de la i -ésima componente del suceso. Puesto algebraicamente:

$$\Omega = \{(\omega_1, \omega_2, \dots) : \omega_i \in \{E, F\}, i \in \mathbb{N}\}$$

$$X_n(\omega) = \begin{cases} 1 & \text{si } \omega_n = E \\ 0 & \text{si } \omega_n = F \end{cases}$$

$$P(X_n = 1) = p, P(X_n = 0) = q$$

3.2 Cantidad de ensayos exitosos

Definamos la variable aleatoria $N_n = X_1 + X_2 + \dots + X_n$. Esta variable representa el número de éxitos en los primeros n ensayos. Como es una suma de variables con idéntica distribución de Bernoulli, tendremos $N_n \sim \text{Bi}(n, p)$. Con esto, si definimos $N_0 = 0$, entonces $\{N_n : n \in \mathbb{N}_0\}$ es un proceso estocástico que modela la cantidad de ensayos exitosos, y su espacio de estados es \mathbb{N}_0 .

Notemos que por $N_{n+1} = N_n + X_{n+1}$, y más en general $N_{n+m} = N_n + X_{n+1} + \dots + X_{n+m}$, para todo n, m positivos.

Por otro lado $N_{m+n} - N_m$ representa el número de éxitos entre los n -ésimo y $(m+n)$ -ésimo ensayos. Osea, $N_{m+n} - N_m \sim N_n$ para $n \in \mathbb{N}_0$. Esto nos dice que no importa en qué punto comenzamos a contar, sino la cantidad de ensayos observados. Junto con esto, como sabemos que los ensayos son independientes, tendremos que la cantidad de éxitos hasta el momento m no influye en la cantidad que habrá entre el momento m y $(m+n)$. Osea

$$P(N_{m+n} - N_m = k | N_0 = k_0, N_1 = k_1, \dots, N_m = k_m) = P(N_{m+n} - N_m = k)$$

para cualquiera sean k_0, k_1, \dots, k_m . Por ser la probabilidad independiente del momento, decimos que el proceso es a incrementos estacionarios.

Otra cuestión interesante relacionada a estas variables es que cumplen con la propiedad markoviana. Sean $0 < n_1 < \dots < n_m$, recién se mencionó que $N_{n_1}, N_{n_2} - N_{n_1}, \dots, N_{n_m} - N_{n_{m-1}}$ son independientes, luego la cantidad de éxitos hasta el momento n_m solo puede depender de la cantidad hasta n_{m-1} y no de momentos anteriores. Esto nos dice que $P(N_{m+1} = k_{m+1} | N_1 = k_1, \dots, N_m = k_m) = P(N_{m+1} = k_{m+1} | N_m = k_m)$, lo cual es en efecto la propiedad markoviana.

3.3 Distribución de éxitos

Hemos visto que los éxitos entre observaciones se comportan de forma independiente a los momentos de observación pero ¿qué se puede decir sobre estos valores?. Es claro que si tras n lanzamientos de una moneda obtuve i caras, el total tras la próxima tirada solo podrá ser en total i o $i+1$ tiradas, no más ni tampoco menos, ya que no se puede retroceder la cuenta ni obtener más de una cara en un lanzamiento. Esto es análogo con los ensayos de Bernoulli:

$$P(N_{n+1} = j | N_n = i) = P(N_n + X_{n+1} = j | N_n = i) = P(X_{n+1} = j - i)$$

Ahora, X_{n+1} solo puede ser 0 o 1 con q y p probabilidades respectivamente, luego

$$P(X_{n+1} = j - i) = \begin{cases} p & \text{si } j - i = 1 \\ q & \text{si } j - i = 0 \\ 0 & \text{en otro caso} \end{cases}$$

Más aún, $P(N_{n+m} = j | N_n = i) = P(N_n + \hat{X} = j | N_n = i) = P(\hat{X} = j - i)$ donde $\hat{X} = X_{n+1} + \dots + X_{n+m} \sim \text{Bi}(m, p)$, luego $P(N_{n+m} = j | N_n = i) = P(\hat{X} = j - i) = \binom{m}{j-i} p^{j-i} q^{m-j+i}$

Finalmente, podemos estimar la cantidad esperada de éxitos en una cantidad determinada de ensayos. De hecho, podemos respaldar esta estimación haciendo uso de información sobre ensayos anteriores. Esto es, por la linealidad de la esperanza podemos observar que:

$$E(N_{n+m} | N_n = k_n) = E(N_n + \hat{X} | N_n = k_n) = E(N_n | N_n = k_n) + E(\hat{X} | N_n = k_n) = k_n + mp$$

ya que $E(\hat{X}) = mp$ por ser binomialmente distribuida.

4 Procesos de Poisson

4.1 ¿Qué son?

Un proceso de Poisson es una clase particular de procesos estocásticos tales como vimos que eran los procesos de Bernoulli y las cadenas de Markov. Este tipo de procesos son particularmente útiles para modelar secuencias de arribos, y por eso constituye una parte muy importante de la Teoría de Colas (**Queueing Theory**).

Por esto, las variables aleatorias son de la forma $N_t(\omega)$: numero de arribos en el intervalo $[0, t)$ en el suceso $\omega \in \Omega$. Debe ocurrir entonces que el recorrido de estas variables sea finito y discreto, pero la parametrización es continua. Luego, $N = \{N_t : t \in \mathbb{R}^+\}$ es un proceso de Poisson si el espacio de estados es \mathbb{N}_0 y además se cumplen las siguientes condiciones:

- $N_0(\omega) = 0 \forall \omega \in \Omega$. Esto es, al momento 0 no han ocurrido arribos.
- La función $t \rightarrow N_t(\omega)$ es seccionalmente constante con saltos de magnitud 1. Esto quiere decir que, para un suceso ω fijo, en un instante dado podrá haber como mucho un arribo, y cada salto se corresponde con uno de estos.
- El número de arribos en el intervalo $[s, s+t)$ corresponde con $N_{s+t} - N_s$. Más aun, este número debe ser finito.

Todas estas hipótesis nos permiten asegurar que la cantidad de arribos no dependen del momento en que se comienza a observar sino de la longitud de la observación (identico a lo que sucede con los procesos de Bernoulli), y además nos permiten demostrar que todas las N_t seguirán una distribución de Poisson con la misma tasa.

4.2 Características de la distribución

De las hipótesis anteriores derivan varios lemas importantes. Todos tienen en común el mismo λ :

1. $P(N_t = 0) = e^{-\lambda t} \forall t \geq 0$
2. $\lim_{t \rightarrow 0} \frac{1}{t} P(N_t \geq 2) = 0$
3. $\lim_{t \rightarrow 0} \frac{1}{t} P(N_t = 1) = \lambda$
4. $P(N_t = k) = \frac{e^{-\lambda t} (\lambda t)^k}{k!} \forall k \in \mathbb{N}_0, t \geq 0$
5. $E(N_t) = \lambda t$ y $V(N_t) = \lambda t$

Osea, logramos ver que $N_t \sim \text{Poss}(\lambda t)$ y por ende el parámetro λ lo interpretamos como la tasa de arribos por unidad de tiempo.

Vemos que aquí también se cumple una especie de propiedad markoviana, ya que, si $u \leq s$, entonces $P(N_{s+t} - N_s | N_u = k_u) = P(N_{s+t} - N_s | N_s = k_s)$. Esto es, solo nos importa en ultima instancia la cantidad de arribos que hubo hasta el instante s , y podemos descartar la información de todos los momentos anteriores. (Notemos que la propiedad markoviana dice esto, pero para una parametrización discreta y finita).

También similar a los procesos de Bernoulli tenemos que si $0 < t_1 < t_2 < \dots < t_n$, entonces $N_{t_1}, N_{t_2} - N_{t_1}, \dots, N_{t_n} - N_{t_{n-1}}$ son todas variables aleatorias independientes.

4.3 Tiempos de arribo

Consideremos un proceso de Poisson de número de arribos y sean las variables aleatorias T_k : momento del k -ésimo arribo. Entonces $\{T_k : k \in \mathbb{N}\}$ es un proceso estocástico con espacio de parametros discreto (\mathbb{N}) y espacio de estados continuo (\mathbb{R}^+) que representa el proceso de tiempos de arribos. Claramente el proceso de tiempos de arribos está relacionado al proceso de número de arribos, ¿pero de qué forma?

- Si hasta el momento s hubo menos de k arribos entonces el momento del k -ésimo arribo es posterior al instante s , y viceversa. Osea, $N_s < k \iff T_k > s$. De esto surge que

$$P(T_k > s) = P(N_s < k) = \sum_{j=0}^{k-1} \frac{e^{-\lambda s} (\lambda s)^j}{j!} \quad (1)$$

- Si entre el momento T_k y $T_k + s$ no ocurren arribos, entonces el tiempo entre el k -ésimo arribo y el $(k+1)$ -ésimo es mayor a s y viceversa. Esto es $T_{k+1} - T_k > s \iff N_{T_k+s} - N_{T_k} = 0$. Luego

$$\begin{aligned} P(T_{k+1} - T_k > s) &= P(N_{T_k+s} - N_{T_k} = 0) = e^{-\lambda s} \\ &\Downarrow \\ P(T_{k+1} - T_k < s | T_1 < s_1, \dots, T_k < s_k) &= 1 - e^{-\lambda s} \end{aligned} \quad (2)$$

La ecuación 2 nos dice que el tiempo entre dos arribos sucesivos es independiente de los tiempos de arribos previos y del número de arribo esperado, que son todas independientes entre sí y también que:

$$T_1, T_2 - T_1, \dots, T_k - T_{k-1} \sim \exp(\lambda)$$

Esto es un resultado muy importante, nos dice que cuando $\{T_k : k \in \mathbb{N}\}$ sea un proceso de tiempos de arribo en un proceso $N = \{N_t : t \geq 0\}$ tales que se verifica que $T_1, T_2 -$

Probabilidad y Estadística

Unidad VII

LCC - FCEIA - UNR

5to cuatrimestre

$T_1, \dots, T_k - T_{k-1}, \dots$ son variables aleatorias independientes, y distribuidas exponencialmente con parámetro λ , entonces N es un proceso de Poisson de parámetro λ .

Por último, $T_k = T_1 + (T_2 - T_1) + \dots + (T_k - T_{k-1})$, y como todos los sumandos siguen una distribución exponencial, entonces $T_k \sim \text{Erlang}(k, \lambda)$ ².

Como conclusión, las características más importantes de un proceso de Poisson con parámetro λ son:

- La variable $N_{s+t} - N_s$ cuenta el número de arribos en el intervalo $[s, s+t)$, no depende de s y sigue una distribución $\text{Pois}(\lambda t)$.
- La variable T_k corresponde con el instante del k -ésimo arribo y $T_k \sim \text{Erlang}(k, \lambda)$.
- La variable $T_{k+1} - T_k$ mide el tiempo entre el $(k+1)$ -ésimo arribo y el k -ésimo, y sigue una distribución $\exp(\lambda)$. Además, las variables $T_1, T_2 - T_1, \dots, T_k - T_{k-1}, \dots$ son independientes entre si.

²Una pista de esto es la ecuación 1, la cual nos muestra que la función de probabilidad de $P(T_k < s)$ coincide con la de la distribución Erlang

5 Bibliografía

- Slides de la Cátedra - Año 2023
- Cadenas de Markov (en tiempo discreto, con estados discretos) - Universidad de La Rioja, Logroño
- Probabilidad - Seymour Lipschutz, Editorial Schaum
- Video de YouTube: [Queueing Theory and Poisson process](#)

A Demostraciones

Demostración de $F(i, j) = P(i, j) + \sum_{b \in E \setminus \{j\}} P(i, b)F(b, j)$:

$$\begin{aligned}
 F(i, j) &= \sum_{n=1}^{\infty} F_n(i, j) \\
 &= F_1(i, j) + \sum_{n=2}^{\infty} F_n(i, j) \\
 &= P(i, j) + \sum_{n=2}^{\infty} \sum_{b \in E \setminus \{j\}} P(i, b) \cdot F_{n-1}(b, j) \\
 &= P(i, j) + \sum_{b \in E \setminus \{j\}} \sum_{n=2}^{\infty} P(i, b) \cdot F_{n-1}(b, j) \\
 &= P(i, j) + \sum_{b \in E \setminus \{j\}} P(i, b) \sum_{n=1}^{\infty} F_n(b, j) \\
 &= P(i, j) + \sum_{b \in E \setminus \{j\}} P(i, b) F(b, j)
 \end{aligned}$$

□

Demostración de propiedades sobre la distribución límite.

Supongamos que existe tal distribución límite π . Sabemos que es única, y además por definición vale que $\lim_{n \rightarrow \infty} P^n(i, j) = \pi(j) \forall i, j \in E$. Enumeraremos las componentes con números para mayor claridad. Tendremos luego

$$\lim_{n \rightarrow \infty} P^n = \begin{pmatrix} \pi(1) & \pi(2) & \pi(3) & \cdots \\ \pi(1) & \pi(2) & \pi(3) & \cdots \\ \pi(1) & \pi(2) & \pi(3) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

ya que sabemos que los valores no dependen del estado inicial, luego no dependen de la fila. Ahora, para demostrar que si π_0 es un vector de probabilidades, entonces $\pi = \pi_0 \lim_{n \rightarrow \infty} P^n$, recordemos que un vector de probabilidades nos dice que $\sum_{i=1}^n \pi_0(i) = 1$, luego

$$(\pi_0 \lim_{n \rightarrow \infty} P^n)(j) = \sum_{i=1}^n \pi_0(i) (\lim_{n \rightarrow \infty} P^n)(i, j) = \sum_{i=1}^n \pi_0(i) \pi(j) = \pi(j) \sum_{i=1}^n \pi_0(i) = \pi(j)$$

y así vemos que ambos vectores son iguales componente a componente.

□