MODELIZACIÓN DE PROBLEMAS ESTADÍSTICOS MEDIANTE GRAFOS

Patricia Caro - Teresa Braicovich - Raquel Cognigni
caropatriciaj@yahoo.com.ar, teresabraicovich@gmail.com, rcognigni@gmail.com
Universidad Nacional del Comahue. Argentina

Tema: Pensamiento probabilístico-estadístico
Modalidad: CB
Nivel educativo: Formación y actualización docente
Palabras clave: Probabilidad de estados – Grafos balanceados- Procesos estocásticos.

Resumen

El concepto de probabilidad puede ser abordado de distintas maneras de acuerdo a las edades evolutivas de los estudiantes. En nuestro caso, trabajaremos con el concepto de probabilidad "a priori" ya que está pensado para estudiantes universitarios. Es interesante descubrir las numerosas aplicaciones de los grafos y cuánto aportan a la comprensión y simplicidad en distintos temas matemáticos. Determinados problemas aleatorios encuentran en los grafos una herramienta sencilla y práctica para llegar a la solución. Los grafos pueden modelizar problemas estocásticos en los que se quiere calcular la probabilidad de que el sistema se encuentre en un determinado estado transcurrido cierto tiempo, ya que el árbol estadístico no permitiría una representación adecuada, pues tendría una cantidad infinita de ramas. Estudiando las propiedades que relacionan las Teorías de Grafos y de Probabilidades, se encontró que en problemas que puedan ser representados por un grafo balanceado y las probabilidades de transición que parten de un mismo estado son iguales, se pueden obtener las probabilidades de cada estado haciendo los cocientes entre el número de arcos que llegan a un determinado vértice y el número total de arcos. Así, el modelar procesos estocásticos infinitos con grafos simplifica enormemente el trabajo algebraico.

INTRODUCCIÓN

El concepto de probabilidad puede ser abordado en los distintos niveles educativos de acuerdo a las edades evolutivas de los estudiantes. En algunas situaciones se plantea desde lo netamente experimental, relacionándose así con el concepto de probabilidad "a posteriori" y en otras se trabaja el concepto de probabilidad "a priori", llegando así a istintos grados de formalización. Las nociones de grafo y de digrafo pueden ser trabajadas como un ente matemático, tienen numerosas aplicaciones y aportan a la comprensión y simplicidad en distintos temas matemáticos. En este trabajo, se muestra que, determinados problemas aleatorios, que pueden ser resueltos por diversos caminos, encuentran en los grafos una herramienta sencilla y práctica para llegar a la solución. La simulación de procesos aleatorios a través de grafos hace accesible a los alumnos problemas cuyo tratamiento formal o teórico es difícil o inadecuado, esto ocurre en procesos que dependen del tiempo y que a veces requieren de un intervalo temporal infinito.

Nuestra propuesta es la modelización mediante grafos de situaciones en las que cada estado o suceso se caracteriza por su distancia al origen. Cada uno de estos estados serán identificados con un vértice del grafo que representará al problema a trabajar y entre dos estados se dibuja una flecha que indica el paso de uno al otro, originándose los arcos del grafo correspondiente. En el próximo apartado se presenta un detalle de la construcción del grafo, de cómo se determinan las probabilidades de transición a cada estado, quedando disponible de una manera clara y accesible toda la información acerca de la situación problemática modelizada. Cabe aclarar que trabajaremos con el concepto de probabilidad "a priori" ya que está pensado para estudiantes universitarios.

DESARROLLO DEL TRABAJO

En primer lugar daremos una serie de definiciones, propiedades y ejemplos que resultarán necesarios para presentar la modelización de procesos estadísticos mediantes grafos.

Probabilidad y diagramas de árbol

En las ramas de los diagramas de árbol se hacen constar las probabilidades de transición entre los nudos. En estos diagramas se cumplen las siguientes propiedades:

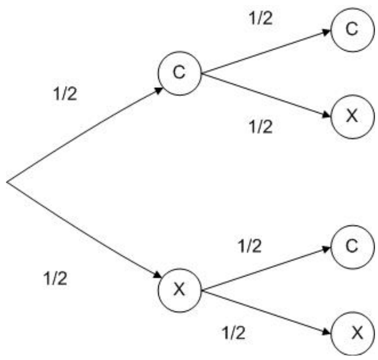
Propiedad 1: La probabilidad de un camino dado es igual al producto de todas las probabilidades a lo largo de dicho camino.

Propiedad 2: La suma de las probabilidades p_i de las ramas que parten de un mismo nudo es igual a 1.

Propiedad 3: La probabilidad de alcanzar un nivel es igual a la suma de las probabilidades de todos los caminos que conducen a ese nivel.

Ejemplo 1: *Calcular la probabilidad de que al lanzar dos monedas se obtenga exactamente una cara.*

El diagrama de árbol correspondiente a este proceso aleatorio es el siguiente:



- Entonces la probabilidad de obtener cara-cruz es $P(CX) = \frac{1}{2}.\frac{1}{2} = \frac{1}{4}$ (Por la propiedad 1).

- La probabilidad de obtener sólo una cara es:

$$P(CX) + P(XC) = \frac{1}{2}.\frac{1}{2} + \frac{1}{2}.\frac{1}{2} = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

(Por la propiedad 3). Además, si nos fijamos en el primer nudo se cumple que $P(C)+P(X)=\frac{1}{2}+\frac{1}{2}=1$, y lo mismo en todos los nudos (Propiedad 2).

Tiempo de espera y grafos

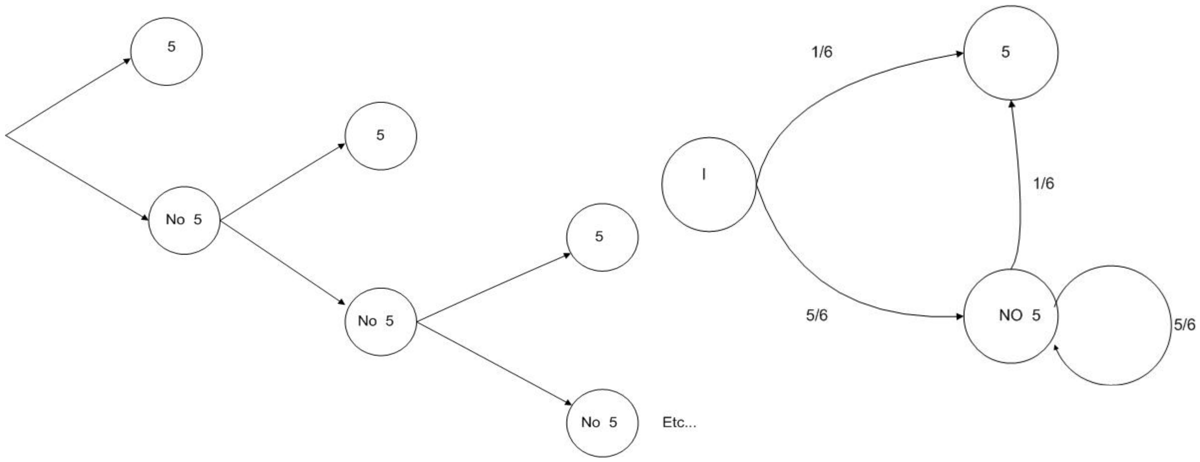
Existen algunos procesos aleatorios simples en los que el árbol deja de ser una forma de representación adecuada ya que, de utilizarse, presentaría un número infinito de ramas. En estos casos, es posible utilizar un grafo para representar el proceso. A continuación definiremos algunos conceptos de grafos.

Definición 1: Un digrafo es una terna $G=(V,U,\Phi)$ que consiste en dos conjuntos no vacíos y disjuntos, *V* y *U*, de elementos llamados *vértices* y *arcos* respectivamente, y de una función Φ, frecuentemente llamada *relación de incidencia*, que asocia a cada arco de *U* un par ordenado de vértices (no necesariamente distintos) de *G*. Si *u* es un arco y *a* y *b* vértices, tales que $\Phi(u)=[a,b]$ se dice que *u* tiene extremo inicial en *a* y extremo final en *b*. Si un arco tiene extremo inicial igual al extremo final se dice que es un *bucle*.

Ejemplo 2: *Lanzar un dado cúbico hasta obtener un 5. La pregunta es: ¿Cuántas tiradas se necesitan para que vuelva a salir otro 5?*

Partiendo del estado inicial I, si sale un 5 la experiencia finaliza y este es un estado final, con una probabilidad de transición de 1/6. Si sale un resultado distinto de 5, la probabilidad de transición es igual a 5/6 y hay que continuar lanzando, con lo cual puede salir 5 o no. Así el proceso se repite infinitamente.

A continuación se muestran las dos representaciones del experimento, el vértice identificado con la letra *I* representa el estado inicial:



Si bien este es un ejemplo elemental y puede ser estudiado a partir de un enfoque frecuencial, la posibilidad de que el proceso sea infinito hace que por lo general no sea considerado en la enseñanza, pero trabajando con la modelización que nos permite el grafo, sí es posible tratar el tema en el aula. El grafo permite efectuar una síntesis de conceptos que pueden ser captados visualmente y se facilita la compresión con respecto a una posible descripción del proceso.

Definición 2: Un *proceso estocástico* es una familia indexada de variables aleatorias {X(t)}, las que representan una sucesión de estados en el tiempo *t*. El índice *t* puede tomar valores continuos o discretos, según si los cambios de estado se producen en todo momento o cada cierto intervalo de tiempo. Para determinar parámetros estadísticos es necesario trabajar con una variable discreta, de modo que el sistema tenga una cantidad finita o infinita numerable de estados. En este caso, el proceso se llama discreto en el tiempo.

Definición 3: El conjunto *S* de valores de las variables X(t) se llama *conjunto o espacio de estados*. Es un conjunto de estados E_1, E_2…, E_k exhaustivos y mutuamente excluyentes de un experimento aleatorio en cualquier tiempo.

Definición 4: Un proceso estocástico se dice que evoluciona *sin memoria* cuando, en cada etapa, el cambio de estado sólo depende del estado en el que se encuentra actualmente.

Definición 5: Un proceso estocástico se llama *homogéneo en el tiempo* si las probabilidades de transición entre los estados no cambian con el tiempo. Se dice que estás probabilidades son estacionarias (no cambian) a través del tiempo.

Los procesos que cumplen con las definiciones 2, 3, 4 y 5 reciben el nombre de cadenas de Markov, discretas, homogéneas en el tiempo y con un número finito o infinito numerable de estados.

Ejemplo 3: En un país lejano sólo existen dos posibilidades en el clima, seco y mojado. Un estudiante de metereología sabe que la probabilidad de que el clima sea seco el 1° de enero del año en curso es a y la probabilidad de que en dos días consecutivos el clima sea el mismo, tiene un valor p, $0 < p < 1$.

Escribamos los elementos que identifican en este problema una cadena de Markov.

Solución:

❖ Solo hay dos posibles estados, E1 es el estado Seco y E2 el estado Mojado. Estos estados son exhaustivos y mutuamente excluyentes en cualquier tiempo. Inicialmente en el tiempo t_0, el sistema puede estar en cualquiera de estos estados.

❖ Sea $a_j^0 (j = 0,1,........,k)$ la probabilidad absoluta de que el sistema se encuentre en el estado E_j en t_0. Definamos p_{ij} como la probabilidad de transición de un paso de ir al estado i en t_{n-1}, al estado j en t_n, es decir, la probabilidad de que en el siguiente periodo (paso) se encuentre en E_j, dado que en el periodo (paso) inmediatamente anterior estuvo en E_j. En nuestro caso $a_1^0 = a$ $\quad a_2^0 = 1 - a$

La matriz P de transición de un paso será:

$$\begin{array}{cc} & \begin{array}{cc}\text{Seco} & \text{Mojado}\end{array} \\ \begin{array}{c}\text{Seco} \\ \text{Mojado}\end{array} & \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix} \end{array}$$

Se observa en P que las probabilidades de clima seco un día, dado que el anterior fue seco; y de mojado un día, dado que el día anterior fue mojado son iguales a p. En cualquier otro caso tenemos una probabilidad igual a (1 - p). Las probabilidades a_1^0 y a_2^0, junto con P, determinan en este ejemplo una cadena de Markov.

A continuación llamaremos S al *conjunto de los estados* de un grafo y sea $p_{ij}(n)$ la *probabilidad de transición* desde el estado i al estado j en n pasos. El estado i se comunica con el j si hay al menos un camino de i a j, es decir, si $p_{ij}(n) > 0$. Un estado i de S se llama *absorbente* si $p_{ii} = 1$. El conjunto de los estados absorbentes de una cadena se llama *borde* de S y se representa con B. Puede ocurrir en algunos casos que B sea el conjunto vacío. Los estados que no pertenecen a B se llaman *interiores* y se representan con T.

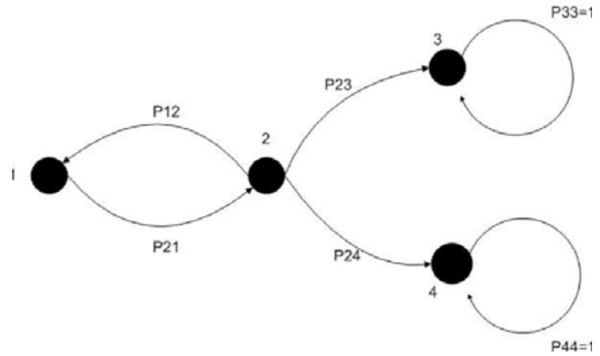
Las probabilidades de transición cumplen las siguientes propiedades:

Propiedad 4: $0 \le p_{ij} \le 1$

Propiedad 5: $\sum p_{ij} = 1$, esto indica que la suma de las probabilidades de todas las aristas que parten de un estado es igual a 1.

Un proceso estocástico está completamente determinado si se conoce además de las probabilidades de transición p_{ij}, el estado inicial I en que se encuentra el sistema y las probabilidad absoluta de que el sistema se encuentre en el estado E_j en t_0.

Ejemplo 4:La siguiente figura muestra el grafo de un proceso aleatorio.



Estados finales como 3 y 4 se llaman absorbentes. En general, un estado *i* es absorbente cuando $p_{ii} = 1$. El borde es el conjunto $B = \{3,4\}$.

Un proceso estocástico se puede representar mediante recorridos en un grafo.

Reglas de los caminos

Propiedad 6: La probabilidad de un camino dado es igual al producto de todas las probabilidades a lo largo de dicho camino.

Propiedad 7: La probabilidad p_i de alcanzar un subconjunto *T* del borde *B* a partir de *i* es igual a la suma de las probabilidades de todos los caminos que conducen desde i hasta *T*.

Propiedad 8: La duración media $\overline{X_i}$ esperada de un recorrido aleatorio desde un estado *i* hasta *B* es la media ponderada de las longitudes de todos los caminos de *i* a *B*. La longitud X_k de cada camino está ponderada por su probabilidad q_k, es decir:

$$\overline{X_i} = \sum_k X_k . p_k$$

Estas reglas son aplicables a aquellos grafos que son diagramas de árbol.

Reglas del valor medio

Propiedad 9: La probabilidad de un estado interior es igual a la media ponderada de las probabilidades de sus estados vecinos. Es decir $p_i = \sum_{k=1}^{n} p_{ik} . p_k$

Propiedad 10: $p_i = 1$ para todos los estados de *T*, $p_i = 0$ para todos los estados de *B* que no están en *T*.

Propiedad 11: El valor de espera de un estado interior es igual a la media ponderada de las probabilidades de sus estados vecinos. Es decir $m_i = 1 + \sum_{k=1}^{n} p_{ik} . m_k$

Propiedad 12: $m_i = 0$ para todos los estados del borde *B*.

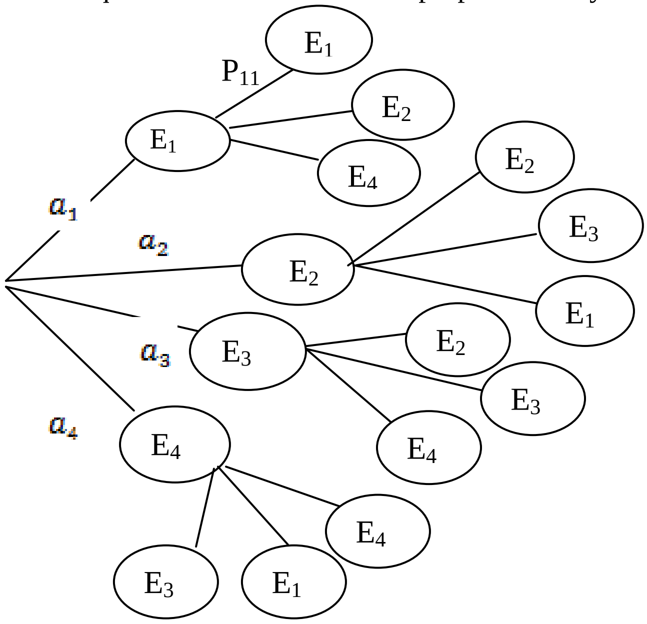
Cadenas de Markov sin estados absorbentes con probabilidades de transición iguales para aquellos arcos que parten de un mismo estado

Definición 6: Un dígrafo $G(V,U)$ es *balanceado* si $gr^+(v) = gr^-(v)$ para todo vértice del conjunto *V*. En caso que el dígrafo sea balanceado y se verifique que $gr^+(v) = gr^-(v) = k$ para todo vértice *v* de *G* se dice que el dígrafo es *k-regular*.

A continuación se mostrarán distintas situaciones que conducen a la resolución a través de grafos balanceados, sean o no *k*-regular.

Ejemplo 5: *calcular la probabilidad de cada estado, partiendo de cualquier estado inicial*

Solución: método tradicional utilizando diagrama de árbol y resolviendo el sistema de ecuaciones que se forma utilizando las propiedades 2 y 9.

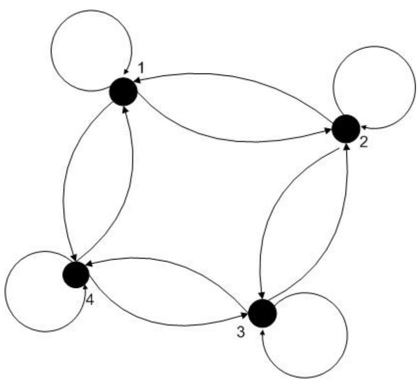


$$a_1 + a_2 + a_3 + a_4 = 1$$
$$a_1 = \frac{1}{3}a_1 + \frac{1}{3}a_2 + \frac{1}{3}a_4$$
$$a_2 = \frac{1}{3}a_2 + \frac{1}{3}a_3 + \frac{1}{3}a_1$$
$$a_3 = \frac{1}{3}a_2 + \frac{1}{3}a_3 + \frac{1}{3}a_4$$
$$a_4 = \frac{1}{3}a_4 + \frac{1}{3}a_1 + \frac{1}{3}a_3$$

Resolviendo el sistema se obtiene:

$$a_1 = a_2 = a_3 = a_4 = \frac{1}{4}$$

Recurriendo a la resolución utilizando grafos nos queda determinado un dígrafo k-regular y balanceado

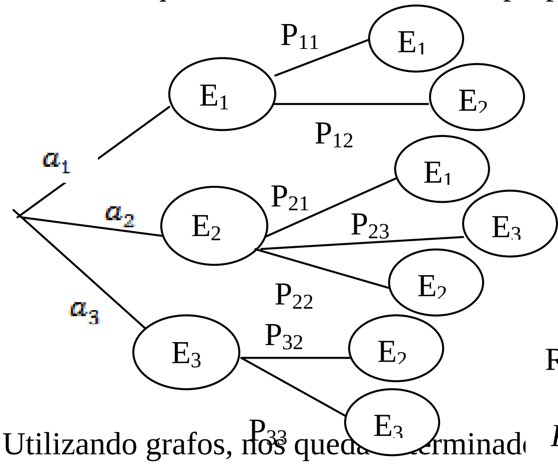


La probabilidad de llegar a cada estado partiendo de cualquier estado se obtiene dividiendo la cantidad de arcos que llegan a cada vértice por la cantidad total de arcos que tiene el dígrafo. Por ejemplo, en el vértice 1, llegan 3 arcos sobre los 12 totales que posee el dígrafo da una probabilidad de 1/4 .Como este dígrafo, además de ser balanceado es *k*-regular, todos los vértices tienen la misma probabilidad.

Ejemplo 6:

Solución: método tradicional utilizando diagrama de árbol y resolviendo el sistema de ecuaciones que se forma utilizando las propiedades 2 y 9

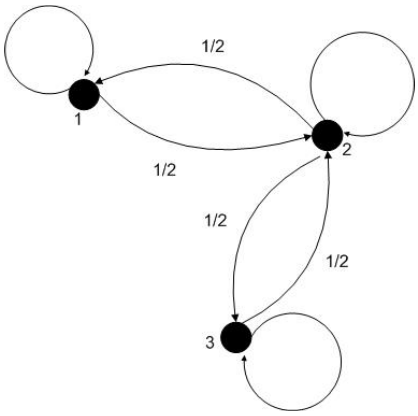
$$a_1 + a_2 + a_3 = 1$$
$$a_1 = \frac{1}{2}a_1 + \frac{1}{3}a_2$$
$$a_2 = \frac{1}{2}a_1 + \frac{1}{3}a_2 + \frac{1}{2}a_3$$
$$a_3 = \frac{1}{2}a_2 + \frac{1}{2}a_3$$



Resolviendo el sistema se obtiene:

Utilizando grafos, nos queda determinad

$$P_1 = \frac{2}{7} \quad , \quad P_2 = \frac{3}{7} \quad , \quad P_3 = \frac{3}{7}$$



Como en el ejemplo 5, este mismo resultado se obtiene si se cuenta la cantidad de arcos que llegan a cada vértice y se divide por la cantidad total de arcos que tiene el dígrafo. Verificando para el vértice 1 se ve que llegan 2 arcos y al dividir por los 7 que posee el dígrafo da una probabilidad de 2/7.

CONCLUSIONES

- Se pueden modelar los procesos estocásticos infinitos con grafos, de manera sencilla.
- Cuando los grafos son balanceados (sean *k*-regular o no) y las probabilidades de transición que parten de un mismo estado son iguales, se pueden obtener las probabilidades de cada estado (vértices) haciendo el cociente entre el número de arcos que llegan a un determinado vértice sobre el número de arcos total del grafo. De esta manera, se simplifica enormemente el trabajo algebraico.
- En función de las conclusiones anteriores podemos decir que sería posible presentar el tema de esta manera en la formación de docentes, por supuesto, haciendo hincapié en la importancia de trabajar con modelización.

Referencias bibliográficas

Braicovich, T.(2009). *Introducción a la Teoría de Grafos.* Buenos Aires: Docuprint S.A.

Contreras, M. (1998). Lenguaje simbólico y pruebas en la enseñanza de las matemáticas: un enfoque sociocognitivo. En G. Mugny y J. Pérez (Eds.), *Psicología social del desarrollo cognitivo*, Capítulo 2, pp. 265-288. Barcelona: Anthropos.

Contreras M. Probabilidad geométrica, grafos y procesos aleatorios.. http://www.mauriciocontreras.es/estadística4.pdf. Consultado el 02/03/2012

Lavalle. A. Rubio N (2003). El ábaco probabilístico en la Enseñanza. XXXI Coloquio Argentino de Estadística.