

Mini-Curso EMALCA 2013
Escuela de Matemáticas de América Latina y del Caribe
Morelia, México, julio de 2013

**INTRODUCCIÓN A LOS MÉTODOS DE MONTE
CARLO**

(versión revisada)

Eliane R. Rodrigues
Instituto de Matemáticas
Universidad Nacional Autónoma de México
E-mail: eliane@math.unam.mx

Apoyo con la solución de problemas y programación:

Juan M. Barrios

Facultad de Ciencias

Universidad Nacional Autónoma de México

E-mail: jbarrios@ciencias.unam.mx

Programas de cómputo pueden ser obtenidos en la página:

<http://sites.google.com/site/jmbarrios/cursos/emalca-2013>

Índice general

1. Introducción a la teoría de probabilidad	3
1.1. Fenómenos aleatorios	3
1.2. Funciones de Probabilidad	5
1.3. Variables aleatorias	7
1.4. Esperanza, varianza y covarianza	12
1.5. Ley de los grandes números	13
1.6. Lista de problemas	14
2. Introducción a las cadenas de Markov	16
2.1. Definiciones básicas	16
2.2. Clasificación de los estados de una cadena de Markov	19
2.3. Resultados básicos	20
2.4. Cadenas reversibles en el tiempo	21
2.5. Lista de problemas	22
3. Métodos de Monte Carlo	24
3.1. Generación de números aleatorios	25
3.2. Métodos de Monte Carlo vía cadenas de Markov	27
3.3. Lista de problemas	28
4. Algoritmos	30
4.1. Algoritmo de Metropolis	30
4.2. Algoritmo de Metropolis-Hastings	31
4.3. Muestreador de Gibbs	32
4.4. Lista de problemas	33

Capítulo 1

Introducción a la teoría de probabilidad

En este capítulo algunos conceptos básicos de la teoría de probabilidad serán presentados. Aunque existen varios resultados conocidos, solamente una pequeña parte será abordada en este capítulo. Los resultados presentados son los que directamente serán utilizados durante la presentación de los métodos de Monte Carlo. La información presentada aquí puede ser encontrada en Feller (1968), Grimmett y Stirzaker (2001) y Ross(2009).

1.1. Fenómenos aleatorios

Un *experimento determinísta* es un experimento que si lo realizamos varias veces bajo las mismas condiciones obtendremos los mismos resultados. Un ejemplo de un experimento determinísta es el experimento que se realiza bajo las hipótesis de la física clásica. Por ejemplo, si un carro inicia en la posición cero, con una velocidad inicial de 0km/h que aumenta de acuerdo a una aceleración de 0.92m/s^2 . ¿Cuanto habrá recorrido este carro después de 30s?

De las ecuaciones de la física clásica tenemos que la posición del carro al tiempo t , indicado por $x(t)$, esta dada por la fórmula

$$x(t) = x(0) + v(0)t + \frac{1}{2}at^2$$

donde $x(0)$ es la posición inicial del carro, $v(0)$ es su velocidad inicial y a es la aceleración. De esta forma, tenemos que $x(30) = 4,12\text{km}$

Bajo las leyes de la física clásica, no importa cuantas veces repetimos este experimento, si usamos la misma información de posición inicial, velocidad inicial y aceleración, siempre obtendremos los mismos resultados.

Un *experimento aleatorio* es un experimento donde podemos decir cuales son los resultados posibles en cada repetición del mismo, pero no podemos decir con seguridad cual de estos resultados obtendremos.

Como ejemplos sencillos de eventos aleatorios tenemos:

1. *Lanzamiento de un dado*. Al lanzar un dado, sabemos que podemos obtener uno de los siguientes resultados: 1, 2, 3, 4, 5, 6, pero a cada lanzamiento no podemos decir con seguridad cual será el valor del lado volteado hacia arriba.
2. *Lanzamiento de tres monedas distinguibles*. Si lanzamos tres monedas simultáneamente e indicamos por a el lado *aguila* y por s el lado *sol*, tenemos que a cada lanzamiento podemos tener

uno de los elementos del siguiente conjunto:

$$\{(a, a, a), (a, a, s), (a, s, a), (s, a, a), (a, s, s), (s, a, s), (s, s, a), (s, s, s)\}.$$

3. *Medición del tiempo de servicio de un aparato electrónico hasta que se descomponga.* El resultado en cada realización de una medición es un valor en el conjunto: $\{t \in \mathbb{R} : t \geq 0 \text{ and } t < \infty\}$

El conjunto de todos los posibles resultados de un experimento aleatorio es llamado *espacio muestral* y será indicado por \mathcal{S} . Cualquier subconjunto de \mathcal{S} es llamado *evento* y será indicado por una letra maiúscula del alfabeto latino.

Para ver que forma toman los eventos en un espacio muestral, considere el experimento de lanzar tres monedas distinguibles. De esta forma,

$$\mathcal{S} = \{(a, a, a), (a, a, s), (a, s, a), (s, a, a), (a, s, s), (s, a, s), (s, s, a), (s, s, s)\},$$

y como ejemplos de eventos tenemos:

1. $A = \{\text{en la primera moneda sale aguila}\} = \{(a, a, a), (a, a, s), (a, s, a), (a, s, s)\}$
2. $B = \{\text{en la segunda moneda sale sol}\} = \{(a, s, a), (a, s, s), (s, s, a), (s, s, s)\}$
3. $C = \{\text{en la tercera moneda sale sol}\} = \{(a, a, s), (a, s, s), (s, a, s), (s, s, s)\}$

Note que las operaciones entre conjuntos, tales como lo son la unión y la intersección, el complemento, entre otras, son válidas cuando aplicadas a eventos. De esta forma, tenemos

1. $A \cup B = \mathcal{S} \setminus \{(s, a, a), (s, a, s)\}$
2. $A \cap C = \{(a, a, s), (a, s, s)\}$

Si dos eventos A y B de un mismo espacio muestral \mathcal{S} son tales que $A \cap B = \emptyset$, entonces decimos que A y B son *mutuamente excluyentes*. Note que el evento A dado arriba y el evento

$$D = \{\text{en la primera y en la tercera moneda salen sol}\} = \{(s, a, s), (s, s, s)\}$$

son mutuamente excluyentes.

La *frecuencia relativa* de un evento es la proporción de veces que el evento ocurre en una serie de realizaciones de un experimento. Formalmente, sea \mathcal{E} un experimento aleatorio que es repetido un número N de veces. Sea A un evento en el espacio muestral de este experimento. Indique por $n_A(N)$ el número de veces en N realizaciones del experimento que el resultado es el evento A . La *frecuencia relativa* del evento A en N realizaciones de un experimento \mathcal{E} , indicada por $f_A(N)$ esta dada por

$$f_A(N) = \frac{n_A(N)}{N} = \frac{\text{número de casos favorables}}{\text{número total de casos}}.$$

Para ilustrar esta definición considere el experimento que registra el color de los carros que pasan en un determinado punto en una calle durante un período fijo de tiempo. Suponga que el número total de carros que pasaron por este punto es 1000, de los cuales 300 eran azules, 400 eran blancos y 300 rojos. Considere el evento

$$A = \{\text{carro azul pasa por el punto}\}.$$

De esta forma, la frecuencia relativa de A es $f_A(1000) = 300/1000 = 0,3$.

Observación. Note que se un experimento aleatorio tienes K posibles resultados, indicados por $r_i, i = 1, 2, \dots, K$, entonces si el experimento es repetido N veces e indicamos por A_i el evento que registra la ocurrencia del resultado r_i , tenemos que

$$\sum_{i=1}^K f_{A_i}(N) = 1.$$

La noción de frecuencia relativa de un evento tiene una relación cercana con el concepto de *probabilidad* del evento. Esto se puede ver en el siguiente concepto informal. Para un experimento \mathcal{E} y un evento E en el espacio muestral de \mathcal{E} se puede definir la probabilidad de E , indicada por $P(E)$, por

$$P(E) = \lim_{N \rightarrow \infty} \frac{n_E(N)}{N}$$

si el límite existe, donde $n_E(N)$ es el número de veces que el evento E ocurre en N realizaciones del experimento.

Observación. Esta definición de probabilidad es consecuencia directa de la conocida *Ley de los Grandes Números* que veremos con mas detalles mas adelante.

1.2. Funciones de Probabilidad

La definición axiomática de *probabilidad de un evento* es: considere que para cada evento A en el espacio muestral \mathcal{S} existe un valor $P(A)$ que satisface las siguientes propiedades,

(a) $0 \leq P(A) \leq 1$,

(b) $P(\mathcal{S}) = 1$,

(c) para A_1, A_2, \dots, A_n , eventos en \mathcal{S} que son mutuamente excluyentes, es decir, $A_i \cap A_j = \emptyset, i \neq j$, vale

$$P(\cup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i),$$

entonces $P(A)$ es llamado *probabilidad de A*.

Sin embargo, existe una definición mas formal que será presentada mas adelante en el contexto de variables aleatorias.

Un ejemplo de una función de probabilidad es el siguiente: sea \mathcal{S} un espacio muestral finito de tamaño M . Suponga, sin pérdida de generalidad, que $\mathcal{S} = \{1, 2, \dots, M\}$, entonces la función definida por $P(\{i\}) = 1/M, i = 1, 2, \dots, M$ es una función de probabilidad y es conocida por función de probabilidad uniforme en \mathcal{S} .

Algunas de las propiedades de una función de probabilidad son:

1. para A en el espacio muestral \mathcal{S} y A^c su complemento, se tiene que $P(A^c) = 1 - P(A)$,
2. para A y B en el espacio muestral \mathcal{S} , si $A \subset B$, entonces $P(A) \leq P(B)$,
3. para cualesquiera A y B en un espacio muestral \mathcal{S} , se tiene que

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Todas estas afirmaciones pueden ser demostradas utilizando propiedades y operación entre conjuntos. Se utiliza el diagrama de Venn para ilustrar las diversas afirmaciones y con esto demostrarlas.

La afirmación (3) dada arriba puede ser generalizada, para más de dos eventos. Esto es realizado de la siguiente forma: sean $A_i, i = 1, 2, \dots, n$, eventos en un espacio de estados \mathcal{S} , entonces vale

$$\begin{aligned} P(\cup_{i=1}^n A_i) &= \sum_{i=1}^n P(A_i) - \sum_{i_1=1}^n \sum_{i_2>i_1}^n P(A_{i_1} \cap A_{i_2}) \\ &+ \sum_{r=3}^{n-1} (-1)^{r+1} \sum_{i_1=1}^n \sum_{i_2>i_1}^n \cdots \sum_{i_r>i_{r-1}}^n P(A_{i_1} \cap \dots \cap A_{i_r}) \\ &+ (-1)^{n+1} P(A_1 \cap \dots \cap A_n), \end{aligned}$$

donde la suma $\sum_{i_1=1}^n \sum_{i_2>i_1}^n \cdots \sum_{i_r>i_{r-1}}^n P(A_{i_1} \cap \dots \cap A_{i_r})$ es sobre todos los subconjuntos de $\{1, 2, \dots, n\}$ de tamaño r .

Una otra función de probabilidad de gran interés es la llamada *probabilidad condicional*. Esta probabilidad consiste en usar alguna información dada para calcular la probabilidad de algún evento en particular. Para ver esto, considere el siguiente ejemplo (Ross, 2009). Considere un edificio donde viven 30 personas. De estas 30 personas, 10 son mujeres y el restante son hombres. De las 10 mujeres, 2 miden mas de 1.70m de altura. De los 20 hombres, 12 miden mas de 1.70m. Pregunta: ¿cuál es la probabilidad de que un entrevistador del INEGI, seleccionando un individuo de este edificio, seleccione un hombre que mide mas de 1.70m? Considere que selección es igualmente probables (uniforme en el espacio de estado).

Construya la siguiente tabla (Tabla 1.1) donde tenemos la especificación de los diversos habitantes del edificio:

	hombre	mujer	total
$> 1,70m$	12	2	14
$\leq 1,70m$	8	8	16
total	20	10	30

Cuadro 1.1: Tabla indicando la división por género y altura de los habitantes del edificio.

Utilizando la interpretación frecuentista, se tiene que de las 30 personas 12 satisfacen la característica de interés, es decir, hombre com mas de 1.70m de altura. De esta forma,

$$P(\text{hombres} > 1,70m) = \frac{12}{30},$$

es la probabilidad buscada.

Considere ahora que un hombre es seleccionado (igualmente probables), ¿cuál es la probabilidad que mida mas de 1.70m de altura? Note que ahora estamos dando la información que un hombre fue seleccionado. De esta forma, estamos reduciendo el espacio de estados para un espacio de estados compuesto solamente de hombre. Así que la probabilidad es ahora 12/20. Este último valor es la *probabilidad condicional* de que un individuo mida mas de 1.70m, dado que es un hombre.

La definición formal de una probabilidad condicional es la siguiente. Sean \mathcal{S} un espacio muestral y $A, B \subset \mathcal{S}$ dos eventos tales que $P(B) > 0$. La *probabilidad condicional* de A dado B , indicada por $P(A|B)$ esta dada por,

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

Observación. La probabilidad condicional satisface todas las propiedades que una función de probabilidad.

Note que de la definición de probabilidad condicional se tiene que

$$P(A \cap B) = P(A | B) P(B).$$

Esta expresión es muy útil para calcular la probabilidad de un evento dado. Para ver esto considere primero la siguiente definición. Decimos que B_1, B_2, \dots, B_n , $B_i \subset \mathcal{S}$, $i = 1, 2, \dots, n$, forman una partición de \mathcal{S} , si:

1. $B_i \cap B_j = \emptyset$, $i \neq j$,
2. $\bigcup_{i=1}^n B_i = \mathcal{S}$,
3. $P(B_i) > 0$.

Ahora podemos dar el siguiente resultado: sea $\{B_1, B_2, \dots, B_n\}$ una partición del espacio muestral \mathcal{S} y sea $B \subset \mathcal{S}$, entonces

$$P(B) = \sum_{i=1}^n P(B \cap B_i) = \sum_{i=1}^n P(B | B_i) P(B_i).$$

La demostración de esta afirmación es inmediata de las propiedades de conjuntos.

De esta fórmula general sale la conocida *fórmula de Bayes* que damos a seguir. Para $\{B_1, B_2, \dots, B_n\}$ una partición del espacio muestral \mathcal{S} y $B \subset \mathcal{S}$ tal que $P(B) > 0$, se tiene que

$$P(B_i | B) = \frac{P(B | B_i) P(B_i)}{\sum_{j=1}^n P(B | B_j) P(B_j)},$$

para toda $i = 1, 2, \dots, n$.

Decimos que dos eventos A y B son llamados *independientes* si $P(A | B) = P(A)$, o equivalentemente, si $P(A \cap B) = P(A) P(B)$. En general, decimos que A_i , $i = 1, 2, \dots, n$ son independientes si para cada subconjunto A_{α_j} , $j = 1, 2, \dots, k$, $k \leq n$, de estos eventos se tiene que

$$P\left(\bigcap_{i=1}^k A_{\alpha_i}\right) = \prod_{i=1}^k P(A_{\alpha_i}).$$

1.3. Variables aleatorias

Una forma de definir una *variable aleatoria* es la siguiente: sea \mathcal{S} un espacio muestral asociado a un experimento aleatorio \mathcal{E} . Una *variable aleatoria* X es una función que asocia a cada evento $e \in \mathcal{S}$ un valor real. La imagen de X es indicada por $S_X = S$ y es llamada *espacio de estados* de X . (La definición formal será dada mas adelante.)

Como un ejemplo, considere el siguiente: una moneda honesta (igual probabilidad de salir águila o sol) es lanzada dos veces. El espacio muestral de este experimento es $\mathcal{S} = \{(a, a), (a, s), (s, a), (s, s)\}$. La función definida por

$$X(e) = \text{cantidad de águilas en } e,$$

es tal que $X : \mathcal{S} \rightarrow \{0, 1, 2\}$ y $X((a, a)) = 2$, $X((a, s)) = X((s, a)) = 1$ y $X((s, s)) = 0$.

Observación. El conjunto S_X puede ser finito o infinito numerable, pero también puede ser identificado con un intervalo real.

Para dar una definición más formal de una variable aleatoria vamos primero definir una σ -álgebra. De esta forma, una colección \mathcal{F} de subconjuntos de un conjunto \mathcal{S} es llamado una σ -álgebra, si las siguientes condiciones son satisfechas:

- (i) $\emptyset \in \mathcal{F}$,
- (ii) si $A_i \in \mathcal{F}$, $i = 1, 2, \dots$, entonces $\bigcup_{i=1}^n A_i \in \mathcal{F}$,
- (iii) si $A \in \mathcal{F}$, entonces $A^c \in \mathcal{F}$.

Ejemplos de una σ -álgebra de un conjunto Ω , son $\mathcal{F} = \{\emptyset, \Omega\}$, $\mathcal{F} = \{\emptyset, \Omega, A, A^c\}$, y $\mathcal{F} = 2^\Omega = \{\text{todos los subconjuntos de } \Omega\}$.

Sea \mathcal{F} la σ -álgebra asociada al espacio muestral \mathcal{S} de un experimento aleatorio \mathcal{E} . Una *variable aleatoria* es una función $X : \mathcal{S} \rightarrow \mathbb{R}$ tal que $\{e \in \mathcal{S} : X(e) \leq x\} \in \mathcal{F}$, para cada $x \in \mathbb{R}$.

A una variable aleatoria podemos asociar una función de probabilidad. Esto se debe al siguiente: sea \mathcal{E} un experimento aleatorio con espacio muestral \mathcal{S} . Sea $X : \mathcal{S} \rightarrow S_X = S$. Tome $B \subset S_X$ y defina $A = \{e \in \mathcal{S} : X(e) \in B\}$. En este caso decimos que A y B son *equivalentes*.

Como un ejemplo, considere el experimento de lanzar una moneda dos veces y sea la variable aleatoria $X(e) = \text{número de águilas en } e$. En este caso tenemos que $S_X = \{0, 1, 2\}$ y $B = \{1\} \in S_X$ y $A = \{(a, s), (s, a)\}$ son equivalentes.

Dado la equivalencia entre los conjuntos en \mathcal{S} y S_X , es posible definir una función de probabilidad en S_X inducida por la variable aleatoria X . De esta forma, sea B un evento en S_X . La *probabilidad de B*, indicada por $P_X(B)$, es definida por

$$P_X(B) = P(\{e \in \mathcal{S} : X(e) \in B\}).$$

Así, dado que para $B = \{1\} \in S_X$ y $A = \{(a, s), (s, a)\}$ son equivalentes, tenemos

$$P_X(B) = P(\{e \in \mathcal{S} : X(e) \in B\}) = P(\{(a, s), (s, a)\}) = P(A) = \frac{1}{2}.$$

De esta forma, cuando consideramos la variable aleatoria X , su imagen S_X y la probabilidad $P(\cdot)$ en \mathcal{S} , estamos induciendo una probabilidad en S_X que es determinada por eventos en \mathcal{S} .

A las variables aleatorias se pueden asociar lo que llamamos *funciones de distribución* que describen el comportamiento de esta variable aleatoria. La función de distribución esta conectada con la probabilidad de ocurrencia de determinados resultados. Formalmente, la *función de distribución* de una variable aleatoria X es una función $F : \mathbb{R} \rightarrow [0, 1]$ dada por

$$F(x) = P(X \leq x).$$

Algunas de las propiedades de una función de distribución son:

1. F es no decreciente, es decir, si $x < y$, entonces $F(x) \leq F(y)$,
2. $\lim_{x \rightarrow \infty} F(x) = 1$,
3. $\lim_{x \rightarrow -\infty} F(x) = 0$,
4. F es continua por la derecha.

Vimos que para el caso donde la variable aleatoria X tiene como imagen un conjunto finito o infinito enumerable, podemos asociar una función de probabilidad. Esta *función de probabilidad* puede ser formalmente definida como sigue. Sea X una variable aleatoria con imagen $S_X = \{x_1, x_2, \dots\}$. Para cada $i = 1, 2, \dots$ sea $p(x_i) = P(X = x_i)$, donde $p(x_i)$ es tal que

$$(a) p(x_i) \geq 0, i = 1, 2, \dots,$$

$$(b) \sum_{i=1}^{\infty} p(x_i) = 1.$$

La función $p : S_X \rightarrow [0, 1]$ es llamada *función de probabilidad de X* .

Como ejemplo de una función de probabilidad para una variable aleatoria X , tenemos lo siguiente: considere el experimento de lanzar una moneda honesta dos veces y defina la variable aleatoria X por

$$X = \text{número de águilas.}$$

De esta forma tenemos que $S_X = \{0, 1, 2\}$ y $p(0) = P(X = 0) = 1/4$, $p(1) = P(X = 1) = 1/2$ y $p(2) = P(X = 2) = 1/4$.

Com base en la definición de función de distribución, tenemos que para una variable aleatoria discreta, su función de distribución es definida por

$$F(x) = \sum_{x_i \leq x} P(X = x_i) = \sum_{x_i \leq x} p(x_i).$$

Observación. Note que esta función es una función escada, donde para $x_1 < x_2 < \dots$, la función tiene valor constante en $[x_i, x_{i+1})$, $i = 1, 2, \dots$, y tiene un salto de tamaño $p(x_{i+1})$ en x_{i+1} .

Considere el caso de la variable aleatoria contando el número de águilas en el lanzamiento de dos monedas honesta. En este caso, tenemos que su función de distribución esta dada por

$$F(x) = \begin{cases} 0, & x < 0 \\ 1/4, & 0 \leq x < 1 \\ 3/4, & 1 \leq x < 2 \\ 1, & x \geq 2. \end{cases}$$

Como hemos visto, la imagen de una variable aleatoria también puede ser un intervalo de \mathbb{R} . En este caso existen funciones que satisfacen algunas propiedades y son tales que podemos escribir la función de distribución correspondiente a esta variable aleatoria en términos de esta función especial. En este caso tenemos una variable aleatoria continua. Formalmente, una variable aleatoria X es llamada *variable aleatoria continua* si existe una función $f(\cdot)$, tal que $f(x) \geq 0$, $x \in \mathbb{R}$, y tal que para $B \subseteq \mathbb{R}$ vale

$$P(X \in B) = \int_B f(x) dx.$$

La función $f(\cdot)$ es llamada *función de densidad de X* .

Como propiedades de esta función tenemos:

1. $P(X \in \mathbb{R}) = \int_{\mathbb{R}} f(x) dx = 1$,
2. para $B = [a, b] \subset \mathbb{R}$, tenemos $P(X \in B) = P(a \leq X \leq b) = \int_a^b f(x) dx$, de donde tenemos que para $a = b$, $P(X = a) = \int_a^a f(x) dx = 0$.
3. $F(x) = P(X \leq x) = \int_{-\infty}^x f(y) dy = P(X < x)$.

Un ejemplo de variable aleatoria continua es la que tiene como imagen un intervalo $(a, b) \subset \mathbb{R}$, $a < b$, y tal que su función de densidad es

$$f(x) = \frac{1}{b-a}, \quad x \in (a, b).$$

Esta variable aleatoria es conocida como la variable aleatoria con *función de densidad uniforme en (a, b)* y esto se indica por $X \stackrel{\mathcal{D}}{=} U(a, b)$. Note que la función de distribución asociado a esta variable aleatoria es

$$F(x) = \begin{cases} 0, & x \leq a \\ (x-a)/(b-a), & a < x < b \\ 1, & x \geq b. \end{cases}$$

En general, asociamos un nombre a la variable aleatoria de acuerdo con la forma de su función de distribución. Algunas de las más conocidas son:

1. Variables aleatorias discretas

a) **Variable aleatoria uniforme.** Sea X una variable aleatoria discreta con espacio de estados $S_X = \{1, 2, \dots, N\}$. Sea $p(k) = P(X = k)$, $k \in S_X$, su función de probabilidad. Si para toda $k \in S_X$ se tiene que $p(k) = |S_X|^{-1} = 1/N$, donde $|A|$ es la cardinalidad del conjunto A , entonces X es llamada una *variable aleatoria con distribución uniforme en S_X* , indicada por $U(S_X)$.

b) **Variable aleatoria de Bernoulli.** Sea X una variable aleatoria que indica el suceso o el fracaso de un experimento. De esta forma, podemos asociar $X = 0$ si un fracaso ocurre y $X = 1$ si ocurre un suceso. Suponga que el experimento sea un suceso con probabilidad $p \in [0, 1]$. De esta forma, tenemos que la función de probabilidad de X esta dada por

$$p(k) = P(X = k) = \begin{cases} p, & \text{si } k = 1, \\ 1 - p, & \text{si } k = 0, \end{cases}$$

y la variable aleatoria X es llamada una *variable aleatoria con distribución Bernoulli(p)*. Su función de distribución esta dada por

$$F(x) = \begin{cases} 0, & \text{si } x < 0 \\ 1 - p, & \text{si } 0 \leq x < 1 \\ 1, & \text{si } x \geq 1. \end{cases}$$

c) **Variable aleatoria Binomial.** Considere que un experimento es repetido un número n de veces independientemente y que a cada repetición el resultado puede ser un suceso con probabilidad $p \in [0, 1]$ y un fracaso con probabilidad $(1 - p)$. Sea X la variable aleatoria que registra el número de veces que un suceso ocurre en las n repeticiones del experimento. De esta forma, la variable aleatoria X tiene función de probabilidad dada por

$$p(k) = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n,$$

y X es llamada una *variable aleatoria Binomial con parámetros n y p* y esto será indicado por $\text{Binomial}(n, p)$.

Observación. Note que una variable aleatoria Bernoulli(p) es una variable aleatoria Binomial ($1, p$). Adicionalmente, note que para X una variable aleatoria Binomial(n, p),

$$\sum_{k=1}^n P(X = k) = \sum_{k=1}^n \binom{n}{k} p^k (1-p)^{n-k} = [p + (1-p)]^n = 1.$$

- d) **Variable aleatoria geométrica.** Una variable aleatoria X con espacio de estados $S_X = \{1, 2, \dots\}$ tiene *distribución geométrica* con parámetro $p \in (0, 1)$, indicada por Geométrica(p), si su función de probabilidad es

$$P(X = k) = (1-p)^{k-1} p, \quad k = 1, 2, \dots$$

Observación. Una variable aleatoria con distribución Geométrica(p) es una variable aleatoria que registra el número de repeticiones independientes de un experimento con probabilidad $p \in (0, 1)$ de suceso, hasta que se observe el primer suceso. Adicionalmente, note que

$$\sum_{k=1}^{\infty} P(X = k) = \sum_{k=1}^{\infty} (1-p)^{k-1} p = \frac{p}{1-(1-p)} = 1.$$

2. Variables aleatorias continuas

- a) **Variable aleatoria exponencial.** Una variable aleatoria X con espacio de estados $S_X = [0, \infty)$, tiene función de distribución *exponencial con parámetro* $\lambda > 0$, indicado por Exp(λ), si su función de densidad está dada por

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{si } x \geq 0 \\ 0, & \text{si } x < 0. \end{cases}$$

Observación. Note que

$$\int_{-\infty}^{\infty} f(x) dx = \int_0^{\infty} \lambda e^{-\lambda x} dx = 1.$$

- b) **Variable aleatoria Beta.** Una variable aleatoria X con espacio de estados $S_X = (0, 1)$, tiene una distribución *Beta con parámetros* α y β , indicada por Beta(α, β), si su función de densidad es

$$f(x) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}, & \text{si } x \in (0, 1) \\ 0, & \text{en otro caso,} \end{cases}$$

donde $\Gamma(a) = \int_0^{\infty} e^{-x} x^{a-1} dx$ es la función gamma.

- c) **Variable aleatoria normal.** Una variable aleatoria X con espacio de estados \mathbb{R} tiene función de distribución *normal con parámetros* $\mu \in \mathbb{R}$ y $\sigma^2 > 0$, indicada por N(μ, σ^2), si su función de densidad es

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \quad x \in \mathbb{R}.$$

Todos los conceptos vistos hasta ahora con respecto a variables aleatorias pueden ser extrapolados para dimensiones mayores que uno. Con esto tenemos las versiones correspondientes para vectores aleatorios. En el caso de vectores aleatorios tenemos la *función de densidad conjunta* que es la

función de densidad asociada a los vectores aleatorios. De esta forma, para (X_1, X_2, \dots, X_n) un vector aleatorio, la función de *distribución conjunta* asociada a él es $F(x_1, x_2, \dots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n)$. En el caso continuo la *función de densidad conjunta*, indicada por $f(x_1, x_2, \dots, x_n)$, esta dada por

$$f(x_1, x_2, \dots, x_n) = \frac{d^n}{dx_1 dx_2 \dots dx_n} F(x_1, x_2, \dots, x_n).$$

Como consecuencia de la definición de función de densidad conjunta, tenemos el concepto de *función de densidad marginal*. Dada una función de densidad conjunta $f(x_1, x_2, \dots, x_n)$, la función de densidad marginal de la i -ésima coordenada es

$$f_i(x) = \int_{S_{X_1}} \int_{S_{X_2}} \dots \int_{S_{X_n}} f(x_1, x_2, \dots, x_{i-1}, x, x_{i+1}, \dots, x_n) dx_1 dx_2 \dots dx_{i-1} dx_{i+1} \dots dx_n.$$

Observación. Definiciones similares son válidas en el caso que las coordenadas del vector aleatorio sean variables aleatorias discretas. En este caso integrales son remplazadas por sumas y derivadas por diferencias.

Duas variables aleatorias X y Y son llamadas *independientes* si $P(X \leq x, Y \leq y) = P(X \leq x) P(Y \leq y)$, es decir, la *función distribución conjunta* de las dos variables aleatorias es el producto de las respectivas funciones de distribución. La generalización para un número K de variables aleatorias es inmediata.

1.4. Esperanza, varianza y covarianza

Como hemos visto, las variables aleatorias están muy relacionadas con resultados de experimentos aleatorios. Sabemos que en los experimentos aleatorios podemos decir cuales los posibles resultados, pero no podemos decir con seguridad cual será el resultados en una determinada realización del experimento. Al repetirlo varias veces diferentes resultados pueden ocurrir. De esta forma, un otro concepto que nos será de utilidad es el concepto de *esperanza matemática de una variable aleatoria*, también conocida como la *media* de esta variable.

La definición formal de la esperanza de una variable aleatoria es como sigue: sea X una variable aleatoria con espacio de estados S_X , entonces la *esperanza* de X , indicada por $E(X)$ esta dada por:

$$E(X) = \begin{cases} \sum_{x \in S_X} x p(x), & \text{si } X \text{ es discreta,} \\ \int_{S_X} x f(x) dx, & \text{si } X \text{ es continua,} \end{cases}$$

donde $p(\cdot)$ y $f(\cdot)$ son las funciones de probabilidad y densidad, para el caso discreto y continuo, respectivamente.

Como ejemplos de esperanza de variables aleatorias tenemos:

1. Si X es una variable aleatoria Bernoulli(p), entonces por definición $S_X = \{0, 1\}$, $p(0) = 1 - p$ y $p(1) = p$. De esta forma, $E(X) = 0 \times p(0) + 1 \times p(1) = 0 \times (1 - p) + 1 \times p = p$.
2. Si X es una variable aleatoria con distribución uniforme $U(a, b)$, $a < b$, entonces $S_X = (a, b)$, $f(x) = 1/(b - a)$ si $x \in (a, b)$ y $f(x) = 0$, en otro caso. Así tenemos que $E(X) = \int_a^b x f(x) dx = \int_a^b x/(b - a) dx = (a + b)/2$.

La esperanza de una variable aleatoria tiene varias propiedades que son bastantes útiles. Algunas de ellas son:

(i) Para X una variable aleatoria con esperanza finita y $a, b \in \mathbb{R}$, tenemos que es válido

$$E(aX + b) = aE(X) + b.$$

(ii) Para $X_i, i = 1, 2, \dots, n$, variables aleatorias tales que $E(X_i) < \infty$, tenemos que

$$E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i).$$

Como podemos ver la esperanza matemática de una variable aleatoria nos dice en media cual es el valor que esta variable aleatoria puede asumir. Otra cantidad de interés cuando queremos describir el comportamiento de una determinada variable aleatoria es como los resultados varian, en media, al rededor de la media. Una forma de medir esta variabilidad es a través de lo que llamamos *varianza* de la variable aleatoria. Formalmente, tenemos que la *varianza* de una variable aleatoria X , indicada por $\text{Var}(X)$, esta dada por

$$\text{Var}(X) = E([X - E(X)]^2),$$

donde $E(X)$ es la esperanza de X .

De esta forma, para obtener la varianza de una variable aleatoria, basta obtener la media de la variable aleatoria $[X - E(X)]^2$. Para hacer esto note que tenemos la siguiente propiedad de esperanza: para X una variable aleatoria con espacio de estados S_X y $g(\cdot)$ una función real vale

$$E(g(X)) = \begin{cases} \sum_{x \in S_X} g(x) p(x), & \text{si } X \text{ es discreta,} \\ \int_{S_X} g(x) f(x) dx, & \text{si } X \text{ es continua,} \end{cases}$$

donde $p(\cdot)$ y $f(\cdot)$ son las funciones de probabilidad y densidad, para el caso discreto y continuo, respectivamente.

Observación. En el caso de vectores aleatorios, definiciones similares aplican, pero ahora las sumas y las integrales son múltiples.

Una propiedad de la varianza de una variable aleatoria es la siguiente: para X una variable aleatoria con varianza $\text{Var}(X)$ y $a, b \in \mathbb{R}$, tenemos que $\text{Var}(aX + b) = a^2 \text{Var}(X)$. Esta propiedad es muy facil de verificar, basta aplicar la definición directamente.

Una medida para estimar la dependencia entre dos variables aleatorias X y Y es la *covarianza entre X y Y* , indicada por $\text{Cov}(X, Y)$. Esta covarianza esta definida por

$$\text{Cov}(X, Y) = E([X - E(X)][Y - E(Y)]) = E(XY) - E(X)E(Y).$$

Note que por definición, si X y Y son independientes, entonces $\text{Cov}(X, Y) = 0$. Para ver esto, solamente use la definición de esperanza del producto de dos variables, la propiedad de las funciones de distribución y la definición de covarianza.

1.5. Ley de los grandes números

La *ley de los grandes números* es un resultado que garante que, la llamada *media empírica* de una muestra, se aproxime de la media de la variable aleatoria que geró la muestra, a la medida que el tamaño de la muestra crece. Para x_1, x_2, \dots, x_n una muestra (valores observados) provenientes de una variable aleatoria, la *media empírica* de esta muestra esta dada por

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i$$

y la *varianza empírica* esta dada por

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2.$$

Existen varias versiones de la ley de los grandes números. Presentaremos dos de ellas. Una de ellas es llamada *ley debil de los grandes números* y la otra *ley fuerte de los grandes números*. Los nombres de “debil” y “fuerte” se refieren a la forma de convergencia que se esta tomando en cuenta. Las dos versiones son dadas a seguir:

- (a) **Ley debil de los grandes números.** Sean $X_i, i = 1, 2, \dots$, una sucesión de variables aleatorias independientes e identicamente distribuidas con media $\mu = E(X_i)$, entonces para todo $\epsilon > 0$, vale

$$\lim_{n \rightarrow \infty} P \left(\left| \frac{1}{n} \sum_{i=1}^n X_i - \mu \right| > \epsilon \right) = 0.$$

- (b) **Ley fuerte de los grandes números.** Bajo los mismos supuestos de la ley debil de los grandes números, se tiene que

$$P \left(\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i - \mu \right) = 1.$$

Dado que la idea general de los métodos de Monte Carlo es aproximar integrales por esperanza de variables aleatorias apropiadas, esta ley es de utilidad en la hora de aplicarlos.

1.6. Lista de problemas

1. Demostrar que para cualesquiera A, B y C eventos en un mismo espacio muestral vale:

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C).$$

Sugerencia: usar diagramas de Venn.

2. Suponga que un determinado teste realizado en un laboratorio médico es 95% efectivo en detectar una dada enfermedad en el caso de que esta este presente. Sin embargo, el teste también produce un resultado falso positivo para 1% de las personas saludables que se someten al mismo. Si 0.5% de la población en realidad tiene la enfermedad ¿cuál es la probabilidad que una persona tenga la enfermedad dado que el teste es positivo?

Sugerencia: defina los eventos $A = \{\text{persona que realiza el teste tiene la enfermedad}\}$ y $B = \{\text{teste es positivo}\}$.

3. Sea X una variable aleatoria discreta con espacio de estados $S_X = \{1, 2, 3, 4\}$ y función de probabilidad $p(i) = P(X = i)$, $i \in S_X$, dada por $p(1) = 1/4$, $p(2) = 1/2$, $p(3) = 1/8$ y $p(4) = 1/8$. Hacer la gráfica de la función de probabilidad, obtener la función de distribución y hacer su gráfica.
4. Demostrar la fórmula de Bayes: sean $B_i, i = 1, 2, \dots, n$ una partición del espacio muestral \mathcal{S} y $B \subseteq \mathcal{S}$ un evento tal que $P(B) > 0$, entonces

$$P(B_i | B) = \frac{P(B | B_i) P(B_i)}{\sum_{j=1}^n P(B | B_j) P(B_j)}, \quad i = 1, 2, \dots, n.$$

5. Se sabe que tornillos son producidos en una fábrica de forma que un tornillo defectuoso es fabricado con probabilidad 0.1. La fábrica vende tornillos en cajas con 10 y ofrece la garantía de a lo mas un tornillo defectuoso a cada 10. Si existen dos tornillos o más en la caja, la fábrica se compromete a devolver el dinero pagado por la caja ¿cuál es la proporción de cajas vendidas que deberán tener el dinero de regreso?

Sugerencia: defina la variable aleatoria $X = \text{número de tornillos defectuosos en una caja}$.

6. En una parada de autobús, estos llegan a cada 15min empezando a las 7:00h de la mañana. Si un pasajero llega a la parada en un tiempo que es uniformemente distribuido entre 7:00h y 7:30h, encuentre la probabilidad que espere

- a) menos de 5 minutos,
- b) mas que 10 minutos.

7. El tiempo, en horas, necesario para reparar una máquina es una variable aleatoria con distribución $\text{Exp}(0.5)$.

- a) ¿Cuál es la probabilidad que el reparo dure mas que 2 horas?
- b) ¿Cuál es la probabilidad que dure mas de 10 horas dado que la duración excede 9 horas?

8. Decimos que una variable aleatoria X tiene distribución de Poisson con parámetros $\lambda > 0$ si su espacio de estados es $S_X = \{0, 1, 2, \dots\}$ y su función de probabilidad esta dada por

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad x \in S_X.$$

Obtenga la media y la varianza de X .

Capítulo 2

Introducción a las cadenas de Markov

En este capítulo algunos conceptos básicos sobre cadenas de Markov serán presentados. Nos centraremos en el caso más simple de este tipo de procesos. Presentaremos las definiciones básicas, algunas de las propiedades de cadenas de Markov y algunos ejemplos. Este tipo de proceso es la base para los métodos de Monte Carlo vía cadenas de Markov. La información aquí puede ser encontrada en Ross (2009) y Karlin y Taylor (1975).

2.1. Definiciones básicas

Las *cadenas de Markov* son secuencias de variables aleatorias que satisfacen ciertas propiedades. De entre ellas tenemos que el espacio de estados de las variables, componiendo esta secuencia, es un estado finito o infinito numerable y es el mismo para todas las variables en la secuencia. Este espacio de estados será denotado por S . Adicionalmente, esta secuencia tiene un conjunto de valores que la indexará. Este conjunto de índices será denotado por T y puede ser un subconjunto de los números enteros o un subintervalo de \mathbb{R} . En el caso que T sea un subintervalo de \mathbb{R} se dice que la cadena de Markov es *continua en el tiempo* o simplemente *continua*. En el caso de que T sea un subconjunto de los números enteros, se dice que la cadena de Markov es *discreta en el tiempo* o simplemente *discreta*.

De esta forma, una *cadena de Markov* con espacio de estados común S , y conjunto de índices T es una secuencia de variables aleatorias, indicada por $X = \{X_t : t \in T\}$, que satisface la siguiente propiedad (llamada *propiedad de Markov*): para $i_0, i_1, \dots, i_{n-2}, i, j \in S$ y $t_0 < t_1 < \dots < t_{n-2} < s < t \in T$, vale

$$P(X_t = j \mid X_{t_0} = i_0, X_{t_1} = i_1, \dots, X_{t_{n-2}} = i_{n-2}, X_s = i) = P(X_t = j \mid X_s = i),$$

es decir, dado el valor de la cadena al tiempo presente, el pasado y el futuro son independientes.

Observación. Estaremos considerando solamente cadenas de Markov con espacio de estados finito y T un subconjunto de los enteros positivos.

En el caso considerado aquí, a la probabilidad de Markov será indicada de la siguiente forma: para $i_0, i_1, \dots, i_{n-1}, i, j \in S$ vale

$$P(X_{n+m} = j \mid X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1}, X_n = i) = P(X_{n+m} = j \mid X_n = i),$$

y la probabilidad condicional $P(X_{n+m} = j \mid X_n = i)$ será llamada *probabilidad de transición en m pasos* y la indicaremos por $P_{ij}^{(n, n+m)}$, $i, j \in S$, $n, m \in T$. En el caso que esta probabilidad no dependa del valor de n , pero solamente de la diferencia m , X será llamada *homogénea en el tiempo* y las probabilidades de transición en m pasos serán indicadas por $P_{ij}^{(m)}$, $i, j \in S$, $m \in T$. Cuando $m = 1$,

entonces las probabilidades de transición en m paso serán llamada solamente de *probabilidades de transición*. Note que para $n = 0$, tenemos que $P_{ij}^{(0)} = 1$, si $i = j$ y $P_{ij}^{(0)} = 0$, si $i \neq j$.

Observación. Estaremos solamente considerando cadenas homogéneas en el tiempo.

La matrix $P^{(m)} = \left(P_{ij}^{(m)} \right)_{i,j \in S}$ es llamada *matrix de transición en m pasos de X* . Cuando $m = 1$, tenemos que $P = \left(P_{ij} \right)_{i,j \in S}$ es llamada *matriz de transición de X* .

Como ejemplos de cadenas de Markov tenemos lo siguiente:

1. **Modelo de Wright-Fisher.** El modelo de Wright-Fisher es un modelo estocástico para describir la evolución de una población. La dinámica considerada es como sigue: considere una población de tamaño fijo N . Suponga que en esta población existan dos tipos de individuos, rotulados *Tipo I* y *Tipo II*. Sea $X = \{X_n : n = 0, 1, \dots\}$ la secuencia de variables aleatorias que registra el número de individuos de *Tipo I* en la población a lo largo del tiempo. De esta forma, X_n indica el número de individuos de *Tipo I* en la generación n .

La dinámica de reproducción de la población es como sigue: a cada unidad de tiempo (generación) individuos son independiente y uniformemente seleccionados para reproducir un descendiente. La producción de individuos continua hasta que N descendientes sean producidos. En este momento, la generación de progenitores es sustituida por la generación de los descendientes y el proceso comienza novamente.

Dado la naturaleza del proceso de reproducción es facil ver que X es una cadena de Markov. El espacio de estados es $S = \{0, 1, \dots, N\}$ y las probabilidades de transición son

$$P_{ij} = P(X_{n+1} = j | X_n = i) = \binom{N}{j} \left(\frac{i}{N} \right)^j \left(\frac{N-j}{N} \right)^{N-j}$$

para todas $i, j \in S$.

2. **Sistema auto-organizador.** Considere que tenemos N objetos rotulados $1, 2, \dots, N$, guardados en una lista en un determinado orden. A cada unidad de tiempo, un objeto es seleccionado para ser utilizado y después es regresado a la lista una posición mas cercana del lado izquierdo de la misma. Suponga que la búsqueda del objeto solicitados comienza en el lado izquierdo de la lista y que la selección del objeto es realizada independientemente del pasado. Asuma que el objeto con rótulo i es seleccionado con probabilidad p_i , $p_i > 0$ y $\sum_{i=1}^N p_i = 1$. Sea $X = \{X_n : n = 0, 1, \dots\}$ la secuencia de variables aleatorias donde X_n registra la composición de la lista después que el n -ésimo pedido por un objeto es realizado y éste es regresado a la lista. Note que el espacio de estados de X es el conjunto de todas posibles listas con N objetos, dando un conjunto con $N!$ elementos.

Para $N = 3$, tenemos que $S = \{(1, 2, 3), (1, 3, 2), (2, 1, 3), (2, 3, 1), (3, 1, 2), (3, 2, 1)\}$. Adicionalmente, si al tiempo n tenemos $X_n = (1, 2, 3)$ y si el objeto de rótulo 3 es seleccionado, entonces tenemos $X_{n+1} = (1, 3, 2)$, si en el próximo paso el objeto seleccionado es 1, entonces $X_{n+2} = (1, 3, 2)$, y así por delante.

Note que el estado futuro de la cadena sólo depende del estado presente. Así que X es una cadena de Markov y su matrix de transición esta dada por

$$P(X_{n+1} = (i_1, i_2, i_3) | X_n = (i_2, i_1, i_3)) = p_{i_2},$$

y es cero en otro caso.

El estado de una cadena de Markov al tiempo cero es llamado *estado inicial*. La distribución de una cadena de Markov X al tiempo cero es llamada *distribución inicial* de X y esta dada por

$$\pi_0(x) = P(X_0 = x),$$

y la *distribución al tiempo n* esta dada por

$$\pi_n(x) = P(X_n = x),$$

para toda $x \in S$.

Para S el espacio de estado de una cadena de Markov X , al conjunto de números indicados por $\{\pi(x), x \in S\}$, tal que

$$\pi(x) = \begin{cases} \sum_{y \in S} \pi(y) P_{yx} \\ \sum_{x \in S} \pi(x) = 1, \end{cases} \quad (2.1)$$

le llamamos *distribución estacionaria* de X . El sistema de ecuaciones (2.1) es llamado *ecuaciones de equilibrio total*.

Como un ejemplo de la distribución estacionaria de una cadena de Markov, considere $X = \{X_n : n = 0, 1, \dots\}$, con espacio de estados $S = \{0, 1, 2\}$ y con matriz de transición

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/3 & 1/3 & 1/3 \\ 1/4 & 1/2 & 1/4 \end{pmatrix}. \quad (2.2)$$

Para encontrar la distribución estacionaria de X basta encontrar la solución del siguiente sistema de ecuaciones

$$(\pi(0), \pi(1), \pi(2)) = (\pi(0), \pi(1), \pi(2)) \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/3 & 1/3 & 1/3 \\ 1/4 & 1/2 & 1/4 \end{pmatrix}$$

sujeito a $\pi(0) + \pi(1) + \pi(2) = 1$, es decir, encontrar la solución de

$$\begin{cases} x + y + z & = 1 \\ \frac{1}{2}x + \frac{1}{3}y + \frac{1}{2}z & = y \\ \frac{1}{3}y + \frac{1}{4}z & = z, \end{cases}$$

que es $\pi(0) = 8/21$, $\pi(1) = 3/7$ y $\pi(2) = 4/21$.

Para una cadena de Markov X con espacio de estados S y matriz de transición en n pasos $P^{(n)} = (P_{ij}^{(n)})_{i,j \in S}$, tome $i, j \in S$, si el límite $\lim_{n \rightarrow \infty} P_{ij}^{(n)}$ existe, a la función límite la llamaremos *distribución límite* de X .

Este concepto de distribución límite así como condiciones para su existencia, serán de importancia al utilizarmos los métodos de Monte Carlo vía cadenas de Markov. Usando la ley de los grandes número conjuntamente con este concepto, nos dice el porqué tiene sentido usar métodos de Monte Carlo vía cadenas de Markov para estimar valores de integrales y series.

2.2. Clasificación de los estados de una cadena de Markov

Algunas de las condiciones para la existencia de una distribución límite están relacionadas con el comportamiento de los estados de una cadena de Markov. De esta forma, necesitamos una forma de clasificar los estados de acuerdo con sus propiedades. Algunos de los conceptos que se necesitarán esta dados a seguir.

Decimos que un estado $j \in S$ es *accesible* a partir de un estado $i \in S$, si existe $n \geq 0$ tal que $P_{ij}^{(n)} > 0$. Si i es accesible a partir de j y vice-versa, decimos que los estados i y j se *comunican*.

Por ejemplo, en el caso de la cadena de Markov con matriz de transición (2.2) tenemos que los estados 0 y 1, se comunican dado que $P_{01} = 1/2 > 0$ y $P_{10} = 1/3 > 0$. Por lo tanto, existen $n_1 = 1$ y $n_2 = 1$ tales que $P_{01}^{(n_1)} > 0$ y $P_{10}^{(n_2)} > 0$.

El *período* de un estado i de una cadena de Markov, indicado por $d(i)$, es definido por

$$d(i) = m.c.d\{n \geq 1 : P_{ii}^{(n)} > 0\}$$

donde *m.c.d.* es el máximo común divisor. Cuando $d(i) = 1$, decimos que i es *aperiódico*.

En el caso de la cadena de Markov con matriz de transición (2.2) vemos que el estado 2 es aperiódico dado que $p_{22} = 1/4 > 0$

La condición para que la distribución límite de una cadena de Markov exista es que la cadena sea, lo que se conoce por, una *cadena ergódica*. Para definir lo que es esta cadena, necesitamos primero de una otra definición, la definición de estado *recurrente positivo*. Una forma de definir un estado recurrente positivo es la siguiente: sea X una cadena de Markov con espacio de estados S . Un estado $i \in S$ es llamado *recurrente positivo* si al empezar en el estado i la esperanza del tiempo de retorno a i es finito.

Formalmente, tenemos que un estado i es llamado *recurrente* si

$$P(X_n = i \text{ para alguna } n \geq 1 | X_0 = i) = 1,$$

si esta probabilidad es estrictamente menor que uno, entonces el estado es llamado *transitorio*

Note que la por la propiedad de Markov (falta de memoria), el concepto de recurrencia de un estado, indica que al salir del estado i , con i recurrente, la cadena regresará a este estado un número infinito de veces con probabilidad uno. En el caso de la transitoriedad, para i transitorio, existe una probabilidad positiva que la cadena jamás regrese a i dado que la cadena partió d i . De los estados recurrentes existen los que son llamado *recurrentes nulos* los que son llamados *recurrentes positivos*. La clasificación entre estos dos tipos de recurrencia depende de lo que se conoce como el *tiempo medio de retorno* al estado.

El *tiempo medio de retorno* de un estado i esta dado por

$$\mu_i = E(T_{ii}) = \begin{cases} \sum_{n=1}^{\infty} n f_{ii}(n), & \text{si } i \text{ es recurrente} \\ \infty, & \text{si } i \text{ es transitorio,} \end{cases}$$

donde $T_{ii} = \min\{n \geq 1 : X_n = i | X_0 = i\}$ es el *tiempo de primer regreso* a i y $f_{ii}(n) = P(X_n = i, X_k \neq i, k = 1, 2, \dots, n-1 | X_0 = i)$ es la distribución de este tiempo de regreso.

Un estado recurrente i es llamado *recurrente positivo* si $\mu_{ii} < \infty$. En el caso que $\mu_{ii} = \infty$, el estado i es llamado *recurrente nulo*.

Un estado es llamado *ergódico* si es recurrente positivo y aperiódico. Una cadena de Markov es llamada *ergódica* si todos sus estados son recurrentes positivos y aperiódicos, y todos sus estados se comunican entre si.

2.3. Resultados básicos

Para poder verificar si una cadena de Markov es ergódica usando las definiciones de los tipos de estados puede ser una tarea un poco compleja. Sin embargo, existen algunos resultados que hacen con que esta verificación sea bastante sencilla, principalmente en el caso de cadenas de Markov con espacio de estados finitos, que es el tipo de espacio de estados que estaremos considerando en estas notas.

En esta sección estaremos presentando tanto algunos resultados como propiedades básicas de las cadenas de Markov que serán de utilidad para comprobar el uso de los métodos de Monte Carlo vía cadenas de Markov en la aproximación de integrales y series.

El primero resultado es llamado *ecuaciones de Chapman-Kolmogorov*. Este conjunto de ecuaciones auxilian en el cálculo de las probabilidades de transición en n pasos y con esto poder verificar el período de un estado, la comunicación entre ellos y la recurrencia positiva.

Teorema. (Ecuaciones de Chapman-Kolmogorov) Sea X una cadena de Markov con espacio de estados S y matriz de transición en m pasos $P^{(m)} = \left(P_{ij}^{(m)} \right)_{i,j \in S}$, entonces para toda $i, j \in S$, $m \geq 0$ y $0 \leq r < m$, se tiene que,

$$P_{ij}^{(m)} = \sum_{k \in S} P_{ik}^{(r)} P_{kj}^{(m-r)}.$$

Demostración. Por definición se tiene que

$$P_{ij}^{(m)} = P(X_m = j | X_0 = i).$$

Usando propiedades de probabilidades conjunta y marginal, tenemos,

$$P_{ij}^{(m)} = P(X_m = j | X_0 = i) = \sum_{k \in S} P(X_m = j, X_r = k | X_0 = i),$$

y usando la definición de probabilidad condicional, podemos escribir

$$\sum_{k \in S} P(X_m = j, X_r = k | X_0 = i) = \sum_{k \in S} P(X_m = j | X_r = k, X_0 = i) P(X_r = k | X_0 = i).$$

Por la propiedad de Markov, se tiene que

$$P(X_m = j | X_r = k, X_0 = i) = P(X_m = j | X_r = k),$$

y el resultados sigue.

De las ecuaciones de Chapman-Kolmogorov, tenemos que para S finito, $P^{(m)} = P^m$, donde P^m es la matriz de transición P a su m -ésima potencia, es decir, podemos obtener la matriz de transición en m pasos, multiplicando la matriz de transición de un paso por ella misma, m veces.

Proposición. Sea S el espacio de estados de una cadena de Markov X . Entonces:

- (i) si $i \in S$ es recurrente positivo e i and $j \in S$ se comunican, entonces j es recurrente positivo,
- (ii) si $i \in S$ tiene período d e i and $j \in S$ se comunican, entonces j también tiene periodo d .

Demostración. Ver Ross (2009) y Karlin y Taylor (1975).

Teorema. Sea X una cadena de Markov con espacio de estados S y matriz de transición en m pasos $P^{(m)} = \left(P_{ij}^{(m)} \right)_{i,j \in S}$. Si X es una cadena ergódica, entonces el límite $\lim_{m \rightarrow \infty} P_{ij}^{(m)}$ existe, es único, estrictamente positivo y no depende del estado inicial i . Adicionalmente,

$$\lim_{m \rightarrow \infty} P_{ij}^{(m)} = \pi(j),$$

donde $\{\pi(i) : i \in S\}$ es la distribución estacionaria de X .

Demostración. Ver Karlin y Taylor (1975) y Ross(2009).

Este resultado garantiza que para una cadena ergódica, la distribución límite existe, es única y coincide con la distribución estacionaria de la cadena. Sin embargo, la existencia de la distribución estacionaria $\pi(\cdot)$ de una cadena, en general, no implica en la existencia de la distribución límite para esta cadena. Un caso donde esto pasa es el siguiente: sea X una cadena con espacio de estados $S = \{1, 2\}$ y tal que su matriz de transición esta dada por

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

En este caso la distribución estacionaria $\{\pi(1), \pi(2)\}$ esta dada por $\pi(1) = \pi(2) = 1/2$, dado que esta es la solución del sistema de ecuaciones

$$\begin{cases} \pi(j) = \sum_{i \in S} \pi(i) P_{ij} \\ \sum_{i \in S} \pi(i) = 1. \end{cases}$$

Así que, por definición, la distribución estacionaria de esta cadena existe. Sin embargo, la distribución límite no. Para ver esto note que, para $n = 1, 2, 3, \dots$ vale,

$$P^{(2n)} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{y que} \quad P^{(2n+1)} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Por lo tanto, dependiendo si n es par o ímpar, tenemos un valor diferente para el límite $\lim_{n \rightarrow \infty} P_{ij}^{(n)}$. De esta forma, el límite no existe y por lo tanto no existe la distribución límite.

Observación. Algunas de las condiciones que deben ser verificadas para poder clasificar una cadena como ergódica o no, pueden no ser tan directas de verificar, como por ejemplo, la recurrencia positiva. Sin embargo, en el caso de espacio de estados finito esto se simplifica. Basta verificar si todos los estados de la cadena se comunican y si ellos son aperiódicos. La propiedad de recurrencia positiva es inmediata, dado que para una cadena con espacio de estados finitos, por lo menos unos de los estados tiene que ser recurrente, y este estado es recurrente positivo (ver Ross, 2009). Dado que la recurrencia es una propiedad de clase, todos los estados son recurrentes positivos (por la comunicación entre todos los estados). La aperiodicidad de todos los estados también es consecuencia inmediata de la comunicación entre ellos y del hecho que esta también es una propiedad de clase.

2.4. Cadenas reversibles en el tiempo

En la sección 2.2 algunas propiedades de las cadenas de Markov fueron presentadas. También fue presentado el concepto de cadena ergódica que garantiza la convergencia de la distribución de m pasos de una cadena de Markov para su distribución estacionaria. Sin embargo, aunque se tenga el sistema de ecuaciones que deben ser resueltas para obtener la expresión de esta distribución límite (cuando existe), en algunos casos particulares esta distribución puede ser obtenida de una forma mas sencilla. Uno de estos casos es el de las cadenas reversibles en el tiempo.

Antes de definir una *cadena reversible en el tiempo*, necesitamos del concepto de *cadena reversa en el tiempo*. Para definir una cadena reversa en el tiempo, considere una cadena de Markov $X = \{X_n : \dots, -2, -1, 0, 1, 2, \dots\}$, con espacio de estados S , matriz de transición $P = (P_{ij})_{i,j \in S}$, y distribución estacionaria $\{\pi(i) : i \in S\}$. Suponga que X ya se encuentre en el *estado estacionario*, es decir, $P(X_n = j) = \pi(j)$, $j \in S$. Defina la cadena $X^* = \{X_n^* : \dots, -2, -1, 0, 1, 2, \dots\}$, tal que si $X_n^* = X_n$,

entonces $X_{n+1}^* = X_{n-1}$. Esta cadena X^* es una cadena de Markov (para verificar esta afirmación basta usar la definición de X^* y la propiedad de Markov de X) con el mismo espacio de estados de X y cuya matriz de transición $P^* = (P_{ij}^*)_{i,j \in S}$ dada por

$$P_{ij}^* = \frac{\pi(j) P_{ji}}{\pi(i)}, \quad i, j \in S.$$

Adicionalmente, X^* también tiene distribución estacionaria $\{\pi(i) : i \in S\}$ (para verificar esta afirmación basta usar las respectivas definiciones y la propiedad de Markov). La cadena X^* es llamada *cadena de Markov reversa en el tiempo* con respecto a P .

Si $P_{ij}^* = P_{ij}$, entonces la cadena X es llamada *reversible en el tiempo*. Note que la condición de reversibilidad puede ser traducida a

$$\pi(i) P_{ij} = \pi(j) P_{ji},$$

para toda $i, j \in S$. Este sistema de ecuaciones es llamado *ecuaciones detalladas de equilibrio*.

Un resultado de bastante utilidad que esta relacionado con estas ecuaciones detalladas de equilibrio es el siguiente:

Proposición. La función de distribución $\{\pi(i) : i \in S\}$ es la distribución estacionaria de una cadena de Markov X reversible en el tiempo, con matriz de transición $P = (P_{ji})_{i,j \in S}$ si, y sólo si, las ecuaciones detalladas de equilibrio son satisfechas.

Demostración. Basta utilizar las definiciones de cadena reversible en el tiempo y distribución estacionaria.

Por lo tanto, para saber si una distribución $\{\pi(i) : i \in S\}$ es la distribución estacionaria de una cadena de Markov X reversible en el tiempo con matriz de transición $P = (P_{ji})_{i,j \in S}$, basta verificar si las ecuaciones detalladas de equilibrio son satisfechas.

Como un ejemplo de una cadena de Markov reversible tenemos la cadena que registra la composición de la lista de objetos en el sistema auto-organizador. En aquél caso la distribución estacionaria esta dada por

$$\pi((i_1, i_2, i_3)) = \frac{1}{c} p_{i_1}^3 p_{i_2}^2 p_{i_3}$$

donde $c = \sum_{\alpha \in S} \pi(\alpha)$ es la constante normalizadora.

Además de los conceptos, definiciones y propiedades presentadas aquí, existe mucho más información al respecto de las cadenas de Markov. Aquí solamente consideramos el caso donde el espacio de estados es finito. Cuando tratamos de espacio de estados infinitos numerables, verificar algunas de las propiedades de las cadenas de Markov, en general, puede no ser tan sencillo.

2.5. Lista de problemas

1. **Modelo de Moran.** Considere una población de tamaño fijo N que consta de dos tipos de individuos A_1 y A_2 que evoluciona segundo la siguiente regla: a cada unidad de tiempo, independientemente de los eventos pasados, un individuo es uniformemente e independientemente seleccionado para reproducir y un individuo es uniformemente e independientemente seleccionado para morir. El individuo que es producido reemplaza el individuo que muere. Suponga que con una probabilidad γ_1 un individuo de tipo A_1 produce un mutante de tipo A_2 y que con probabilidad γ_2 un individuo de tipo A_2 produce un mutante de tipo A_1 . Defina la variable aleatoria X_n como la variable que registra el número de individuos de tipo A_1 al tiempo n . Indique por $X = \{X_n : n = 0, 1, \dots\}$ la cadena de Markov formada por estas variables aleatorias. Obter el espacio de estados de X y sus probabilidades de transición.

2. Considere una cadena de Markov X con espacio de estados $S = \{1, 2, 3\}$ y con matriz de transición dada por

$$P = \begin{pmatrix} 0,9 & 0,075 & 0,025 \\ 0,15 & 0,8 & 0,05 \\ 0,25 & 0,25 & 0,5 \end{pmatrix}.$$

- a) Obtenga la distribución estacionaria $\{\pi(i) : i \in S\}$ de esta cadena.
 b) Es la cadena X ergódica? Si no es ergódica, justificar su respuesta. Si es ergódica, justificar su respuesta y obtener la distribución límite usando la definición de distribución límite.
3. Demuestre que la función de distribución $\{\pi(i) : i \in S\}$ es la distribución estacionaria de una cadena de Markov X , reversible en el tiempo con matriz de transición $P = (P_{ji})_{i,j \in S}$ si, y sólo si, las ecuaciones detalladas de equilibrio son satisfechas.
4. Determine cuales estados se comunican entre si y cual es el período de cada uno de ellos para la cadena de Markov con matriz de transición

$$P = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 \end{pmatrix}.$$

5. Considere una cadena de Markov que clasifica el cambio en la humedad del aire. El valor cero es asignado si está seco y valor uno es asignado se esta húmedo. Sea X_n la variable aleatoria que indica si en el n -ésimo día tenemos un día húmedo o un día seco. Obtenga el espacio de estados de $X = \{X_n : n = 0, 1, \dots\}$. Suponga que la matriz de transición de X es

$$P = \begin{pmatrix} 3/4 & 1/4 \\ 1/3 & 2/3 \end{pmatrix}.$$

Obtenga la distribución límite y la distribución estacionaria de X . ¿ Es la cadena X ergódica? Justifique su respuesta.

6. Sea X una cadena de Markov con espacio de estados $S = \{0, 1, 2\}$ y vector de probabilidad inicial dado por $\pi_0 = (\pi_0(0), \pi_0(1), \pi_0(2)) = (1/4, 1/2, 1/4)$ y matriz de transición

$$P = \begin{pmatrix} 1/4 & 3/4 & 0 \\ 1/3 & 1/3 & 1/3 \\ 0 & 1/4 & 2/4 \end{pmatrix}.$$

Calcule:

- a) $P(X_0 = 0, X_1 = 1, X_2 = 1)$.
 b) $P_{01}^{(2)}$.

Capítulo 3

Métodos de Monte Carlo

Algunos de los resultados presentados aquí pueden ser encontrados en Robert y Casella (1999) entre otros.

Los *métodos de Monte Carlo* son utilizados para aproximar integrales del tipo

$$\int_S g(x) f(x) dx, \quad S \subseteq \mathbb{R}$$

o series del tipo

$$\sum_{x \in S} g(x) p(x), \quad S \subseteq \{\dots, -2, -1, 0, 1, 2, \dots\}$$

donde $g(\cdot)$ es una función con dominio S y con imagen un subconjunto de los números reales, $f : S \rightarrow \mathbb{R}$, $p : S \rightarrow \mathbb{R}$, con $f(x) > 0$, $g(x) > 0$, $x \in S$, tal que la integral (serie) es finita. (El caso multidimensional es similar.)

Si $f(\cdot)$ y $g(\cdot)$ son tales que

$$\int_S f(x) dx < \infty \quad \text{y} \quad \sum_S p(x) < \infty$$

entonces podemos normalizarlas por constantes c_f y c_p , respectivamente, de forma que

$$\int_S \frac{f(x)}{c_f} dx = 1 \quad \text{y} \quad \sum_S \frac{p(x)}{c_p} = 1,$$

entonces el siguiente es válido.

Para $f : S \rightarrow \mathbb{R}$ tal que $\int_S f(x) dx = 1$, existe una variable aleatoria X con espacio de estados S con f es su función de densidad y tal que para toda $g : S \rightarrow \mathbb{R}$, $g(X)$ es un variable aleatoria con $\int_S g(x) f(x) dx < \infty$ (Billingsley, 1995). De esta forma podemos escribir

$$E[g(X)] = \int_S g(x) f(x) dx.$$

Similar argumentación es válida en el caso de la serie.

Por lo tanto, tanto en el caso de la integral como en el caso de la serie, por la *ley de los grandes números* tenemos que para M suficientemente grande

$$E[g(X)] \approx \frac{1}{M} \sum_{i=1}^M g(x_i)$$

donde $x_i, i = 1, 2, \dots, M$ es una muestra de valores generados a partir de $f(\cdot)$.

Observación. De ahora en adelante vamos a estar describiendo el procedimiento para el caso de integrales, pero el mismo es válido para el caso de series.

En general, lo que se quiere es calcular una integral de la forma

$$\int_S g(x) dx.$$

Para realizar este cálculo también usaremos la relación que existe entre esperanza de una variable aleatoria e integrales. De esta forma, considere X una variable aleatoria con función de densidad $f : S \rightarrow \mathbb{R}$ y sea $g : S \rightarrow \mathbb{R}$ tal que $Z = g(X)/f(X)$ es una variable aleatoria, es decir, $f(x) > 0$ and $\{w \in \mathcal{S} : g(X(w))/f(X(w)) \leq x\} \in \mathcal{F}$ (la σ -álgebra asociada al espacio muestral de X). En este caso, podemos escribir

$$\int_S g(x) dx = \int_S \frac{g(x)}{f(x)} f(x) dx = \mathbb{E} \left[\frac{g(X)}{f(X)} \right].$$

Así que por la ley de los grandes números tenemos que para M suficientemente grande

$$\mathbb{E} \left[\frac{g(X)}{f(X)} \right] \approx \frac{1}{M} \sum_{i=1}^M \frac{g(x_i)}{f(x_i)}$$

donde $x_i, i = 1, 2, \dots, M$ es una muestra de valores generados a partir de $f(\cdot)$.

Por lo tanto, vemos que todo se reduce en obtener una muestra de valores simulados a partir de una función de distribución adecuada y usar la ley de los grandes números para aproximar la integral (serie) deseada. De esta forma, necesitamos un algoritmo para generar estos valores. En general, los paquetes de cómputo tienen rutinas que permiten simular valores de un gran número de funciones de distribución. Una de estas distribuciones es la uniforme en el intervalo $[0, 1]$. Con esta función de distribución es posible simular valores de otras funciones de distribución más generales. Algunos de los métodos utilizados son dados en la siguiente sección.

3.1. Generación de números aleatorios

En esta sección vamos a proporcionar algunos algoritmos básicos para simular valores a partir de una función de distribución en particular. Los algoritmos son generales y algunos de ellos pueden ser utilizados para mostrar valores de cualquier función de distribución para la cual se puede obtener la función inversa, tanto la inversa usual, como la llamada *inversa generalizada*.

1. **Uniforme(0,1) - generador recursivo múltiple.** Este generador es un generados pseudo-aleatorio que es utilizado como base para la obtención de los valores uniformemente distribuidos en $(0,1)$. Tome $a_i, i = 1, 2, \dots, K$ valores en \mathbb{Z}_m el conjunto de los enteros módulo m , es decir, $\mathbb{Z}_m = \{0, 1, 2, \dots, m\}$. Suponga que ya tengamos $x_{n-1}, x_{n-2}, \dots, x_{n-K}$, entonces el valor x_n es obtenido de la siguiente forma

$$x_n = (a_1 x_{n-1} + \dots + a_K x_{n-K}) \bmod(m).$$

Los valores uniformes generados son $u_n = x_n/m$. Este generador tiene un ciclo de tamaño $p = m^K - 1$. Por lo tanto, tenemos que tomar m de forma que la p sea lo más grande posible. Adicionalmente, se toma m un número primo muy grande.

Muchos programas de cómputo traen un generador de números aleatorios con distribución uniforme. De esta forma, basta usar uno de ellos y, para una función de densidad o de probabilidad más general, usar uno de los siguientes algoritmos.

2. **Variable aleatoria discreta X con función de probabilidad p_k , $k \in S_X$.** Sea $S_X = \{a_0, a_1, \dots\}$, donde los elementos son ordenados de forma que $a_i < a_{i+1}$, $i = 0, 1, \dots$, y $P(X = a_k) = p_k > 0$, $k = 0, 1, \dots$ con $\sum_{k=0}^{\infty} p_k = 1$. Queremos generar valores a partir de la función de probabilidad $\bar{p} = \{p_k : k = 0, 1, \dots\}$. Sea $F(\cdot)$ la función de distribución asociada a \bar{p} , entonces por definición tenemos que

$$F(k) = P(X \leq a_k) = \sum_{l=0}^k p_l.$$

Para generar una muestra con función de distribución \bar{p} proceda como sigue:

- a) genere un valor u con distribución $U(0, 1)$,
- b) atribuya valor a_k a X si

$$F(k-1) < u \leq F(k),$$

- c) repetir (a) y (b) hasta que se complete el tamaño de muestra deseado.

Este procedimiento genera una muestra de valores para cualquier función de probabilidad. Las distribuciones discretas más conocidas se encuentran programadas en varios de los paquetes de cómputo.

3. **Función de distribución con inversa usual.** Sea $F(\cdot)$ la función de distribución de una variable aleatoria X tal que su inversa usual $F^{-1}(\cdot)$ existe. Para generar valores de una muestra a partir de esta función de distribución, basta utilizar el siguiente algoritmo:

- a) genere un valor u con distribución $U(0, 1)$,
- b) atribuya valor $F^{-1}(u)$ a X ,
- c) repetir (a) y (b) hasta que se complete la muestra del tamaño deseado.

Usando la definición y propiedades de función de distribución, de distribución $U(0, 1)$, y de función inversa, se puede demostrar que los valores de X generados de esta forma tiene la función de distribución $F(\cdot)$, es decir, $P(X \leq x) = F(x)$.

Para muchas de las funciones de distribución con una inversa usual, el código utilizado para gerar valores a partir de ellas estan programados en varios paquetes de cómputo.

4. **Función de distribución sin inversa usual.** Sea X una variable aleatoria con función de distribución $F(\cdot)$ que es una función sin la inversa usual. En este caso usaremos la *inversa generalizada* de $F(\cdot)$, que es definida de la siguiente forma: para $u \in [0, 1]$, la *inversa generalizada* de $F(\cdot)$ esta dada por

$$F^{-1}(u) = \inf\{x \in \mathbb{R} : F(x) \geq u\}.$$

Note que, cuando X es tal que su función de distribución es una función continua, la inversa generalizada coincide con a inversa usual. Así que para generar valores de X a partir de $F(\cdot)$, basta seguir el algoritmo dado para el caso da inversa usual, sólo que ahora usando a inversa generalizada.

5. **Aceptación/Rechazo.** Este es uno de los métodos que pueden ser utilizados cuando no se puede generar una muestra con una distribución deseada utilizando los métodos (1), (2), (3) y (4) presentados arriba.

Para utilizar este algoritmo y generar una muestra de valores de una variable aleatoria X con función de distribución $F(\cdot)$ y función de densidad (probabilidad) $f(\cdot)$, proceda de la siguiente forma: tome $g(\cdot)$ una función de densidad con función de distribución $G(\cdot)$ tal que para x en el soporte de $f(\cdot)$, $f(x) \leq M g(x)$, para alguna constante M , y tal que generar una muestra de valores con distribución $G(\cdot)$ es relativamente fácil. Entonces,

- a) genere independientemente un valor y de la variable aleatoria Y con función de densidad $g(\cdot)$ y un valor u de la variable aleatoria U con distribución $U(0, 1)$,

- b) si

$$u \leq \frac{f(y)}{M g(y)},$$

entonces defina $X = y$, caso contrario regrese al paso (a).

Además existen otras formas de poder generar valores a partir de una función de distribución de interés, pero aquí sólo vamos a considerar los tipos dados en (1), (2), (3), (4) y (5) dados arriba.

Por lo tanto, cuando la función $g(\cdot)$ es tal que su dominio y su forma, hace con que sea fácil encontrar una función de distribución tal que podemos generar valores de ella de forma sencilla, aproximar la integral de $g(\cdot)$ es algo sencillo. El problema aparece cuando la forma de $g(\cdot)$ hace con que la función de densidad de la cual es necesario generar valores sea una función complicada o entonces el dominio de $g(\cdot)$ es muy grande o entonces $g(\cdot)$ es multivariada. En este caso, una posible solución es utilizar los *métodos de Monte Carlo vía cadenas de Markov*.

3.2. Métodos de Monte Carlo vía cadenas de Markov

Los *métodos de Monte Carlo vía cadenas de Markov* son utilizados cuando queremos obtener una muestra de valores generados a partir de una función de densidad (probabilidad) que tiene la siguiente forma

$$\pi(x) = \frac{h(x)}{c}$$

teniendo como dominio un conjunto S , donde $h(\cdot)$ es una función con forma conocida y c es la constante normalizadora, es decir, c es tal que

$$\int_S \pi(x) dx = 1.$$

Algunos de los casos donde los métodos de Monte Carlo vía cadenas de Markov son de utilidad son cuando el espacio de estados S es muy grande haciendo con que la constante c no sea fácil de calcularse, o cuando $h(\cdot)$ es multidimensional y no existe un algoritmo implementado en algún paquete que se pueda usar para simular valores, o cuando la forma de $h(\cdot)$ es conocida pero no puede ser identificada con alguna función de densidad (probabilidad) conocida que se pueda utilizar como forma de obtener valores muestreados que simulan los que son producidos por la función $h(\cdot)$.

Cuando usamos los métodos de Monte Carlo vía cadenas de Markov el objetivo es construir una cadena (proceso) de Markov ergódico con espacio de estados S y distribución límite (estacionaria)

$\{\pi(x) : x \in S\}$. Por lo tanto, lo que tiene que ser construido es una cadena de Markov $X = \{X_n : n = 0, 1, 2, \dots\}$ con espacio de estados el dominio de $h(\cdot)$, es decir, el conjunto S , tal que X sea ergódica y tal que para $P_{xy}^{(n)}$, $x, y \in S$ las probabilidades de transición en n pasos de X , tengamos

$$\lim_{n \rightarrow \infty} P_{xy}^{(n)} = \lim_{n \rightarrow \infty} P(X_n = y | X_0 = x) = \pi(y),$$

para toda $x, y \in S$. De esta forma, tenemos que existe una n^* tal que

$$P(X_n = x) \approx \pi(x), \quad x \in S,$$

para toda $n > n^*$.

Con esto surge un nuevo problema que es estimar el valor de n^* adecuado. Existen varios métodos para poder realizar esta tarea. Muchos de los programas de cómputo que son utilizados para estimar parámetros de modelos utilizando la estadística Bayesiana, poseen varias rutinas para testar la convergencia a la distribución límite de una cadena de Markov. En esta clase de programas esta el “R”. En el paquete CODA están programados los testes más utilizados.

En el próximo capítulo algunos de los algoritmos de Monte Carlo vía cadenas de Markov serán presentados.

3.3. Lista de problemas

1. Decimos que una variable aleatoria X tiene distribución geométrica con parámetro $p \in (0, 1)$, denotado por $X \stackrel{\mathcal{D}}{=} \text{Geométrica}(p)$, si su función de probabilidad esta dada por

$$P(X = k) = p_k = \begin{cases} (1 - p)^{k-1} p, & k = 1, 2, \dots \\ 0, & \text{en otro caso.} \end{cases}$$

Un algoritmo utilizado para generar variables aleatorias con distribución Geométrica(p) es: genere valores u_1, u_2, \dots a partir de una $U(0, 1)$ hasta que el valor generado es menor o igual a p . Demuestre que la variable aleatoria X que cuenta el número de valores generados hasta que se obtenga un valor menor o igual a p tiene distribución Geométrica(p).

2. Una variable aleatoria X tiene distribución exponencial con parámetro $\lambda > 0$, indicado por $X \stackrel{\mathcal{D}}{=} \text{Exp}(\lambda)$, si su función de densidad esta dada por

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & \text{en otro caso.} \end{cases}$$

Usando el método de la función inversa, demuestre que la variable aleatoria X definida por

$$X = -\frac{1}{\lambda} \log(U),$$

donde $U \stackrel{\mathcal{D}}{=} U(0, 1)$, tiene distribución $\text{Exp}(\lambda)$.

3. Decimos que una variable aleatoria X tiene distribución gama con parámetros $\alpha > 0$ y $\beta > 0$, indicado por $X \stackrel{\mathcal{D}}{=} \text{Gama}(\alpha, \beta)$, si su función de densidad esta dada por

$$f(x) = \begin{cases} (\beta^\alpha / \Gamma(\alpha)) x^{\alpha-1} e^{-\beta x}, & x > 0 \\ 0, & \text{en otro caso,} \end{cases}$$

donde

$$\Gamma(\alpha) = \begin{cases} (\alpha - 1)!, & \text{si } \alpha \text{ es un valor entero,} \\ \int_0^{\infty} x^{\alpha-1} e^{-x} dx, & \text{en otro caso,} \end{cases}$$

es la función gama. Sea $Y \stackrel{\mathcal{D}}{=} \text{Gama}(\alpha, 1)$. Demuestre, usando la definición de función de distribución, que la variable aleatoria definida por $X = Y/\beta$ tiene distribución $\text{Gama}(\alpha, \beta)$.

Capítulo 4

Algoritmos

En este capítulo algunos de los algoritmos que pueden ser utilizados para generar valores a partir de una función distribución a través de una cadena de Markov, serán presentados. Existen varios otros, y algunos de ellos son variantes de los que serán presentados aquí (ver por ejemplo Robert y Casella, 1999). Los *métodos de Monte Carlo vía cadenas de Markov* empezaron a ser más conocidos después de la publicación del artículo de Hastings (1970). En aquél trabajo, una generalización del algoritmo propuesto por Metropolis y Ulam (1949) y Metropolis et al. (1953) fue presentado. El algoritmo original de Metropolis y co-autores, fue utilizado para estudiar problemas en el área de física. Este algoritmo así como el conocido algoritmo de Metropolis-Hastings están dados en la siguientes secciones.

Aunque estamos considerando solamente cadenas de Markov con espacio de estados finitos (o distribuciones con dominio un conjunto finito), en problemas reales el espacio de estados puede ser más general. En este caso, definiciones más generales para ergodicidad deben ser consideradas.

4.1. Algoritmo de Metropolis

Como hemos comentado, el objetivo es construir una cadena de Markov ergódica cuya distribución límite es la distribución de donde queremos muestrear valores. La forma utilizada por el *algoritmo de Metropolis* es la siguiente: sea S el espacio de estados de la función $h(\cdot)$ (definida en el capítulo 3). Este conjunto S será el espacio de estados de la cadena a ser construida. Sea $\{\pi(x) : x \in S\}$ la distribución de la cual queremos simular valores.

Para empezar, debemos conseguir una matriz estocástica (transición) $Q = (Q_{xy})_{x,y \in S}$, con $Q_{xy} > 0$, $x, y \in S$, y tal que dado $x \in S$ es relativamente fácil generar valores $y \in S$ a partir de Q_x . Adicionalmente, Q debe ser tal que $Q_{xy} = Q_{yx}$, es decir, debemos tomar matrices simétricas (esta última condición será eliminada en el caso del algoritmo de Metropolis-Hastings).

La descripción del algoritmo es como sigue:

- (a) Tome $X_0 = x^{(0)}$ obtenido del espacio de estado de acuerdo a alguna distribución,
- (b) Para $n = 1, 2, \dots$, suponga que el estado de la cadena es $X_{n-1} = x^{(n-1)}$,
 - (i) gere $y \in S$ a partir de $Q_{x^{(n-1)}}$.
 - (ii) calcule

$$\alpha(x^{(n-1)}, y) = \min \left\{ 1, \frac{\pi(y)}{\pi(x^{(n-1)})} \right\}$$

(iii) haga

$$X_n = x^{(n)} = \begin{cases} y, & \text{con probabilidad } \alpha(x^{(n-1)}, y) \\ x^{(n-1)}, & \text{en otro caso.} \end{cases}$$

Observación. Note que con la construcción descrita arriba, la constante normalizadora c no aparece en la formulación. De esta forma, para utilizar este algoritmo basta tener la forma de $h(\cdot)$ y la matriz Q .

Proposición ().* La cadena de Markov $X = \{X_n : n = 0, 1, 2, \dots\}$ construída a través del algoritmo de Metropolis es una cadena ergódica reversible, con espacio de estados S , matriz de transición $P = (P_{xy})_{x,y \in S}$ dada por:

$$P_{xy} = \begin{cases} Q_{xy} \alpha(x, y), & y \neq x \\ Q_{xx} + \sum_{y \neq x} Q_{xy} [1 - \alpha(x, y)], & x = y, \end{cases}$$

y distribución estacionaria $\{\pi(x) : x \in S\}$.

Demostración. Dejado como uno de los problemas a serem resueltos (ver Problema 1 de la lista de problemas al final de este capítulo).

Observación. Note que si el $y \in S$ que es generado a partir de Q_x es tal que $\pi(y)/\pi(x) \geq 1$, entonces el algoritmo acepta el nuevo valor y la cadena cambia para este nuevo valor y . En el caso que $\pi(y)/\pi(x) < 1$, entonces con probabilidad $\pi(y)/\pi(x)$, el algoritmo acepta el cambio, es decir, generamos una variable aleatoria $U \stackrel{D}{=} U(0, 1)$ y se $U = u < \pi(y)/\pi(x)$, entonces la cadena cambia al nuevo valor y , en otro caso se queda donde esta, es decir, con el valor x .

4.2. Algoritmo de Metropolis-Hastings

Esta generalización del algoritmo de Metropolis fue propuesta por Hastings (1970) y elimina la condición de que la matriz $Q = (Q_{xy})_{x,y \in S}$ sea simétrica. De esta forma, basta que sea una matriz de transición con valores positivos y que sea facil generar valores a partir de ella.

Considere la misma notación usada en la descripción del algoritmo de Metropolis. De esta forma, S , que es el dominio de la función $h(\cdot)$, es el espacio de estados de la cadena a ser construida. También indicaremos por $\{\pi(x) : x \in S\}$ la distribución de la cual queremos simular valores a partir de ella. Sea $Q = (Q_{xy})_{x,y \in S}$, una matriz estocástica (transición) con $Q_{xy} > 0$, $x, y \in S$, y tal que dado $x \in S$, es relativamente facil gerar valores $y \in S$ a partir de Q_x .

Construya una cadena de Markov $X = \{X_n : n = 0, 1, \dots\}$ de forma similar a la que fue construida usando el algoritmo de Metropolis, pero ahora la probabilidad de aceptación $\alpha(x, y)$ esta dada de la siguiente forma: sea $X_{n-1} = x^{(n-1)}$ si el valor propuesto generado es $y \in S$, entonces haga $X_n = y$ con probabilidad

$$\alpha(x^{(n-1)}, y) = \min \left\{ 1, \frac{\pi(y) Q_{y x^{(n-1)}}}{\pi(x^{(n-1)}) Q_{x^{(n-1)} y}} \right\},$$

caso contrario haga $X_n = x^{(n-1)}$.

Proposición. La cadena de Markov $X = \{X_n : n = 0, 1, \dots\}$ generada por la construcción del algoritmo Metropolis-Hastings es una cadena ergódica, reversible en el tiempo, con espacio de estados S y matriz de transición $P = (P_{xy})_{x,y \in S}$, dada de forma similar a la de la cadena producida por el algoritmo de Metropolis pero ahora con la nueva $\alpha(x, y)$. Adicionalmente, la distribución estacionaria de X es $\{\pi(x) : x \in S\}$.

Observación. Note que el algoritmo permite que $X_{n+1} = X_n$, por lo tanto la cadena es aperiódica. La irreducibilidad es satisfecha al tenermos $P_{xy} > 0$ para todas $x, y \in S$ (dado que $Q_{xy} > 0$ para todas $x, y \in S$). La recurrencia positiva viene de que el espacio de estados es finito y de la propiedad de irreducibilidad.

4.3. Muestreador de Gibbs

El *muestreador de Gibbs* es utilizado para obtener una muestra de valores a partir de distribuciones multivariadas que satisfacen ciertas condiciones. Una de estas condiciones es que sea facil simular valores a partir de las llamadas *distribuciones marginales condicionales completas* (que serán definidas mas adelante).

Antes de iniciar la descripción del algoritmo de Gibbs, necesitamos definir algo de notación. Sea $d > 1$ un número entero y sea el vector aleatorio (X_1, X_2, \dots, X_d) con distribución $\pi((x_1, x_2, \dots, x_d))$. Para $\bar{x} = (x_1, x_2, \dots, x_d)$ indique por $\bar{x}_{(-i)}$ el vector \bar{x} sin la i -ésima coordenada, es decir, $\bar{x}_{(-i)} = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_d)$. Las *distribuciones marginales condicionales completas de $\pi(\cdot)$* son las distribuciones $\pi_i(\cdot | \bar{x}_{(-i)})$, $i = 1, 2, \dots, d$. Suponga que es relativamente facil generar valores a partir de $\pi_i(\cdot | \bar{x}_{(-i)})$ para cada i .

El *algoritmo de Gibbs* puede ser descrito de la siguiente forma:

(a) tome $\bar{X}_0 = (x_1^{(0)}, x_2^{(0)}, \dots, x_d^{(0)}) = \bar{x}^{(0)}$ de acuerdo a alguna distribución inicial,

(b) para $n = 1, 2, \dots$, suponga que $X_{n-1} = (x_1^{(n-1)}, x_2^{(n-1)}, \dots, x_d^{(n-1)}) = \bar{x}^{(n-1)}$,

(i) genere $x_1^{(n)}$ utilizando $\pi_1(\cdot | x_2^{(n-1)}, x_3^{(n-1)}, \dots, x_d^{(n-1)})$

■ para $i = 2, 3, \dots, d-1$, genere $x_i^{(n)}$ utilizando

$$\pi_i(\cdot | x_1^{(n)}, \dots, x_{i-1}^{(n)}, x_{i+1}^{(n-1)}, \dots, x_d^{(n-1)})$$

(ii) genere $x_d^{(n)}$ utilizando $\pi_d(\cdot | x_1^{(n)}, x_2^{(n)}, \dots, x_{d-1}^{(n)})$

(c) haga $\bar{X}_n = (x_1^{(n)}, x_2^{(n)}, \dots, x_d^{(n)}) = \bar{x}^{(n)}$.

La cadena $\bar{X} = \{\bar{X}_n = (X_1^{(n)}, X_2^{(n)}, \dots, X_d^{(n)}) : n = 0, 1, \dots\}$ generada de esta forma es una cadena de Markov, así como las sucesiones formadas por sus coordenadas. Adicionalmente, en el caso que X sea ergódica, la distribución límite de esta cadena de Markov X es $\pi(\bar{x}) = \pi((x_1, x_2, \dots, x_d))$.

Observación. Note que X tiene como probabilidades de transición el siguiente:

$$P_{\bar{x}^{(n-1)}\bar{x}^{(n)}} = P(X_n = \bar{x}^{(n)} | X_{n-1} = \bar{x}^{(n-1)}) = \pi_1(x_1^{(n)} | x_2^{(n-1)}, \dots, x_d^{(n-1)}) \left[\prod_{i=2}^{d-1} \pi_i(x_i^{(n)} | x_1^{(n)}, \dots, x_{i-1}^{(n)}, x_{i+1}^{(n-1)}, \dots, x_d^{(n-1)}) \right] \pi_d(x_d^{(n)} | x_1^{(n)}, x_2^{(n)}, \dots, x_{d-1}^{(n)})$$

Una de la condiciones para que se pueda utilizar el muestreador de Gibbs es que sea relativamente facil generar valores de las probabilidades condicionales marginales completas. Sin embargo, muchas veces esto no ocurre cuando uno esta trabajando con problemas reales. Una solución para este problema es considerar un paso del algoritmo de Metropolis-Hastings dentro del muestreador

de Gibbs. Este paso del algoritmo de Metropolis-Hastings sería utilizado en aquellos casos donde muestrear directamente de la probabilidad condicional marginal completa no sea tan simple.

Las condiciones de ergodicidad son de especial interés cuando utilizando el algoritmo de Gibbs, donde aún cuando el espacio de estado es finito, es posible que el algoritmo no pueda producir en un paso de la simulación una configuración que sea compatible con el problema en estudio. Por lo tanto, es muy importante verificar que la condición de ergodicidad sea satisfecha cuando usamos el algoritmo de Gibbs.

4.4. Lista de problemas

1. Demuestre la Proposición (*) dada en la descripción del algoritmo de Metropolis.
2. Obtenga la forma de la matriz de transición de la cadena X formada por el algoritmo de Metropolis-Hastings. Demuestre que X es reversible en el tiempo y que $\{\pi(x) : x \in S\}$ es su distribución estacionaria.

Bibliografía

1. P. Billingsley (1995), *Probability and measure*, John Wiley and Sons, Estados Unidos.
2. W. Feller (1968), *An introduction to probability theory and its applications*, Vol. 1, 3rd Edition, John Wiley and Sons, Estados Unidos.
3. G. Grimmett y D. Stirzaker (2001), *Probability and random processes*, Third Edition, Oxford University Press, Reino Unido.
4. W. K. Hastings (1970), Monte Carlo sampling methods using Markov chains and their applications, *Biometrika* **57**, 97-109.
5. S. Karlin y H. E. Taylor (1981), *A first course in stochastic processes*, Second Edition, Academic Press, Estados Unidos.
6. N. Metropolis y S. Ulam (1949), The Monte Carlo method, *Journal of the American Statistical Association* **44**, 335-341.
7. N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller y E. Teller (1953), Equations of state calculations by fast computing machine, *J. Chem. Phys.* **21**, 1087-1091.
8. C. P. Robert y G. Casella (1999), *Monte Carlo statistical methods*. Springer Text in Statistics. Springer. Estados Unidos.
9. S. Ross (2009), *A first course in probability*, 8th Edition, Pearson, Estados Unidos.