

Modelos Estocásticos I

Notas de Curso

Joaquín Ortega Sánchez Víctor Rivero Mercado

Cimat, A.C.

Índice general

1. Introducción a la Teoría de Probabilidad	1
1.1. Introducción	1
1.2. Probabilidad Condicional	3
1.3. Variables Aleatorias	4
1.4. Distribución de una Variable Aleatoria	6
1.5. Funciones de Distribución	6
1.5.1. Variables Discretas	7
1.5.2. Variables Continuas	7
1.6. Valores Esperados y Momentos	8
1.7. Distribuciones Conjuntas e Independencia	10
1.8. Algunas Distribuciones Importantes	13
1.8.1. Distribuciones Discretas	13
1.8.2. Distribuciones Continuas	15
1.9. Probabilidad y Esperanza Condicional	19
1.9.1. El Caso Discreto	19
1.9.2. El Caso Continuo	22
1.9.3. El Caso Mixto	23
1.9.4. Sumas Aleatorias	25
1.10. Funciones Generadoras de Probabilidad	28
1.10.1. Funciones Generadoras de Probabilidad y Sumas de V. A. I.	29
1.11. Funciones Generadoras de Momentos.	31
1.12. Simulación de Variables Aleatorias	33
1.12.1. Método de la Distribución Inversa	33
1.12.2. Método de Rechazo	35
1.12.3. Métodos Particulares	37
1.12.4. Generación de Variables Aleatorias en R	39
1.13. Convergencia de Variables Aleatorias	39
1.13.1. Relación entre los Distintos Tipos de Convergencia	42
2. Cadenas de Markov	47
2.1. Introducción	47
2.2. Definiciones	47
2.2.1. Consecuencias de la Propiedad de Markov	50
2.2.2. Ejemplos	53
2.3. Matrices de Transición	58
2.4. Clasificación de los Estados	61
2.5. Descomposición del Espacio de Estados	65
2.6. Estudio de las Transiciones Iniciales	70

2.7. Paseo al Azar	74
2.8. Procesos de Ramificación	82
2.9. Cadenas de Nacimiento y Muerte.	89
2.10. Simulación de Cadenas de Markov	92
3. Propiedades Asintóticas	95
3.1. Distribuciones Estacionarias	95
3.2. Visitas a un Estado Recurrente	98
3.3. Estados Recurrentes	102
3.4. Existencia y Unicidad de Distribuciones Estacionarias.	104
3.5. Cadenas Reducibles	106
3.6. Convergencia a la Distribución Estacionaria	107
3.7. Invertibilidad	110
3.8. Teorema Ergódico	112
3.9. Ejemplos	113
3.9.1. Cadenas de Nacimiento y Muerte	115
3.10. Inferencia en Cadenas de Markov	122
3.10.1. Comportamiento asintótico	123
3.10.2. Pruebas de independencia	124
3.10.3. Orden de la cadena	125
4. Procesos de Poisson	127
4.1. Distribución Exponencial	127
4.1.1. Falta de Memoria	127
4.1.2. Mínimo de Variables Exponenciales	128
4.2. La Distribución de Poisson	130
4.3. El Proceso de Poisson	131
4.4. Postulados para el Proceso de Poisson	136
4.5. Distribuciones Asociadas a un Proceso de Poisson	138
4.6. Procesos de Poisson Compuestos	139
4.7. Descomposición de un Proceso de Poisson	141
4.8. Superposición de Procesos de Poisson	142
4.9. Procesos No Homogéneos	142
4.9.1. Postulados para un proceso de Poisson no-homogéneo	143
4.9.2. Procesos de Cox	145
4.10. La Distribución Uniforme	146
4.11. Procesos Espaciales de Poisson	150
4.11.1. Procesos no homogéneos en el plano	152

Capítulo 1

Introducción a la Teoría de Probabilidad

1.1. Introducción

El objetivo de la Teoría de Probabilidades es desarrollar modelos para experimentos que están gobernados por el azar y estudiar sus propiedades y aplicaciones. El modelo fundamental para un experimento de este tipo, como el lanzamiento de un dado, es el Espacio de Probabilidad, que describimos a continuación.

En primer lugar tenemos un conjunto Ω , conocido como el *espacio muestral*, que contiene todos los resultados posibles del experimento. Por ejemplo, si el experimento consiste en lanzar un dado, el espacio muestral es $\Omega = \{1, 2, 3, 4, 5, 6\}$. Si seleccionamos un punto al azar en el intervalo $[0, 1]$, el espacio muestral es $\Omega = [0, 1]$. Si consideramos una sucesión infinita de experimentos con dos resultados posibles: 0 ó 1, el espacio muestral es el conjunto de todas las sucesiones de ceros y unos: $\Omega = \{(a_n)_{n \geq 1}, a_n = 0 \text{ ó } 1\}$.

Los elementos del espacio muestral se denotan por ω y se conocen como los sucesos o eventos elementales.

La segunda componente de nuestro modelo es la clase de los eventos o sucesos \mathcal{F} . Esta clase está compuesta por subconjuntos del espacio muestral y debe satisfacer las siguientes propiedades

A1. $\Omega \in \mathcal{F}$.

A2. Si $A \in \mathcal{F}$ entonces $A^c \in \mathcal{F}$.

A3. Si $A_n \in \mathcal{F}$ para $n \geq 1$ entonces $\bigcup_{n \geq 1} A_n \in \mathcal{F}$.

Una colección de subconjuntos de Ω que satisface estas tres condiciones se conoce como una σ -álgebra. Es sencillo demostrar que si A es cualquier conjunto, el conjunto de partes de A , $\mathcal{P}(A)$ es una σ -álgebra.

En el caso de experimentos sencillos, por ejemplo experimentos con un conjunto finito de resultados, normalmente tomamos como σ -álgebra de eventos el conjunto de partes de Ω . En experimentos más complicados, con una cantidad no-numerable de resultados posibles, no siempre es posible tomar esta opción, y es necesario considerar σ -álgebras más pequeñas.

La tercera y última componente del modelo es una *probabilidad* P definida sobre la clase de conjuntos \mathcal{F} que toma valores sobre el intervalo $[0, 1]$, y que satisface las siguientes propiedades:

P1. Para cualquier evento $A \in \mathcal{F}$,

$$0 = P(\emptyset) \leq P(A) \leq P(\Omega) = 1.$$

P2. Si A_n , $n \geq 1$ es una colección de conjuntos disjuntos 2 a 2, es decir, $A_i \cap A_j = \emptyset$ si $i \neq j$, entonces

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

La terna (Ω, \mathcal{F}, P) se llama un *espacio de probabilidad*. La función P es una (*medida de*) *probabilidad* y como está definida sobre \mathcal{F} , sólo podemos determinar la probabilidad de los conjuntos que están en esta clase; por eso decimos que estos son los conjuntos *medibles*. Las propiedades de \mathcal{F} garantizan que si hacemos las operaciones usuales (unión, intersección, complementos, diferencias, diferencias simétricas) con conjuntos medibles, obtenemos conjuntos medibles. Por ejemplo, si $A \subset B$ son conjuntos medibles entonces $B \setminus A$ también es medible. Como consecuencia de esto y la aditividad de las medidas de probabilidad tenemos que estas son monótonas: si $A \subset B$ son conjuntos medibles, $P(A) \leq P(B)$, ya que $B = A \cup (B \setminus A)$, los eventos en esta unión son disjuntos y por la aditividad

$$P(B) = P(A) + P(B \setminus A) \geq P(A).$$

Ejemplo 1.1

En el caso del lanzamiento de un dado, el espacio muestral es $\Omega = \{1, 2, 3, 4, 5, 6\}$ y los conjuntos medibles \mathcal{F} son todos los subconjuntos de Ω . Para definir la probabilidad P basta decir que todos los eventos elementales tienen la misma probabilidad (que por lo tanto es $1/6$):

$$P(\{i\}) = 1/6, \quad \text{para } i = 1, \dots, 6.$$

Con esta definición, si A es cualquier subconjunto de Ω entonces

$$P(A) = \frac{\text{Card}(A)}{6} = \frac{\text{Card}(A)}{\text{Card}(\Omega)},$$

donde $\text{card}(A)$ denota el cardinal del conjunto A . ▲

Ejemplo 1.2

Si tenemos un experimento que tiene una cantidad numerable de resultados posibles, $\Omega = \{\omega_1, \omega_2, \dots\}$, podemos tomar \mathcal{F} como la colección de todos los subconjuntos de Ω . Si p_1, p_2, \dots son números no-negativos que satisfacen $\sum_n p_n = 1$, podemos definir la probabilidad P por

$$P(A) = \sum_{\omega_i \in A} p_i, \quad \text{para } A \in \mathcal{F}.$$
▲

Ejemplo 1.3

Si el experimento consiste en escoger al azar un número en el intervalo $[0, 1]$, entonces la probabilidad de escoger un número en el intervalo $[c, d] \subset [0, 1]$ debe ser proporcional a la longitud del intervalo, pero como la probabilidad de que el número escogido caiga en el intervalo $[0, 1]$ es 1, vemos que no sólo es proporcional sino que es igual a la longitud del intervalo:

$$P([c, d]) = d - c, \quad \text{para todo } [c, d] \subset [0, 1]. \quad (1.1)$$

Lamentablemente, no es posible definir una medida de probabilidad sobre todos los subconjuntos de $[0, 1]$ que satisfaga la propiedad (1.1). La demostración de este hecho está fuera de los objetivos de este curso, pero esto implica que hay conjuntos que no son 'medibles', es decir, a los cuales no podemos asignarles una probabilidad.

Por lo tanto, es necesario restringirse a una clase más pequeña \mathcal{F} de subconjuntos de $[0, 1]$, que sea una σ -álgebra, es decir, que satisfaga las condiciones A1, A2 y A3. Una posibilidad es usar la clase de los conjuntos de Borel o σ -álgebra de Borel o borelianos en $[0, 1]$, que es la menor σ -álgebra generada por los subintervalos de $[0, 1]$. Sin embargo, es importante observar que en este caso hay otras σ -álgebras que pueden considerarse. ▲

Dada cualquier colección \mathcal{C} de subconjuntos de Ω , es posible demostrar que existe una σ -álgebra, que denotaremos por $\sigma(\mathcal{C})$, que contiene a \mathcal{C} y que es la menor de todas las σ -álgebras que contienen a \mathcal{C} en el siguiente sentido: Si \mathcal{D} es otra σ -álgebra que contiene a \mathcal{C} , entonces se cumple que $\sigma(\mathcal{C}) \subset \mathcal{D}$. $\sigma(\mathcal{C})$ se conoce como la σ -álgebra generada por \mathcal{C} y es posible demostrar que siempre existe y es única.

En el ejemplo 1.3 mencionamos a la σ -álgebra de Borel \mathcal{B} en $[0, 1]$, que tiene gran importancia en el desarrollo de la teoría de la medida, y que introducimos como la σ -álgebra generada por los subintervalos de $[0, 1]$. De manera equivalente se puede definir como la σ -álgebra generada por la colección de los intervalos abiertos (a, b) , $0 \leq a < b \leq 1$, o los intervalos cerrados $[a, b]$ o los de la forma $(a, b]$, o de la forma $[a, b)$. Es posible demostrar que todas estas definiciones son equivalentes.

También es posible definir la σ -álgebra de Borel como la σ -álgebra generada por los subconjuntos abiertos de $[0, 1]$, y se puede demostrar que esta definición es equivalente a cualquiera de las anteriores. Esta definición tiene la ventaja de que podemos usarla en cualquier espacio que tenga una topología, por ejemplo, en cualquier espacio métrico.

1.2. Probabilidad Condicional

Definición 1.1 La *probabilidad condicional* $P(A|B)$ del evento A dado el evento B se define por

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{si } P(B) > 0, \quad (1.2)$$

y no está definida, o se le asigna un valor arbitrario, cuando $P(B) = 0$.

A partir de esta definición tenemos la relación

$$P(A \cap B) = P(A|B)P(B). \quad (1.3)$$

Supongamos que saber que el evento B ocurrió no cambia la probabilidad de que A ocurra, es decir $P(A|B) = P(A)$. Entonces la relación (1.3) se convierte en

$$P(A \cap B) = P(A)P(B). \quad (1.4)$$

Definición 1.2 Si (1.4) se satisface decimos que los eventos A y B son *independientes*.

Ley de la Probabilidad Total y el Teorema de Bayes

Sea $\{B_i, i \geq 1\}$ una partición de Ω , es decir, una colección de subconjuntos de Ω que satisface

$$B_i \cap B_j = \emptyset \text{ siempre que } i \neq j \quad \text{y} \quad \Omega = \bigcup_1^{\infty} B_i.$$

Entonces, para cualquier evento A , teniendo en cuenta que los subconjuntos $(A \cap B_i), i \geq 1$ son disjuntos dos a dos,

$$P(A) = P(A \cap \Omega) = P(A \cap (\bigcup_1^{\infty} B_i)) = \sum_1^{\infty} P(A \cap B_i)$$

y ahora, usando la relación (1.3) en cada sumando obtenemos

$$P(A) = \sum_1^{\infty} P(A|B_i)P(B_i) \quad (1.5)$$

que se conoce como la ley de la probabilidad total.

Una consecuencia importante del resultado anterior es el Teorema de Bayes.

Teorema 1.1 (Bayes) *Si los eventos $\{B_i, i \geq 1\}$ forman una partición de Ω , para cualquier evento A con $P(A) > 0$ y cualquier índice j ,*

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^{\infty} P(A|B_i)P(B_i)}. \quad (1.6)$$

Demostración. A partir de la definición de probabilidad condicional obtenemos

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

Ahora usando (1.3) en el numerador y (1.5) en el denominador obtenemos el resultado. ■

Ejemplo 1.4

Una prueba médica para detectar cierta enfermedad tiene una efectividad de 98 % (si una persona padece la enfermedad la prueba es positiva con probabilidad 0.98) y da falsos positivos en 10 % de los casos. Por investigaciones epidemiológicas se sabe que un 2 % de la población padece la enfermedad. Si un paciente seleccionado al azar resulta positivo en la prueba, ¿cuál es la probabilidad de que tenga la enfermedad?

Usemos las notaciones E y S para un paciente enfermo o sano, respectivamente, y $+$ para una prueba positiva. E y S son eventos complementarios. Nuestra información inicial es:

$$P(+|E) = 0.98; \quad P(+|S) = 0.1; \quad P(E) = 0.02.$$

Queremos calcular $P(E|+)$ y usando el teorema de Bayes tenemos

$$\begin{aligned} P(E|+) &= \frac{P(+|E)P(E)}{P(+|E)P(E) + P(+|S)P(S)} \\ &= \frac{0.98 \times 0.02}{0.98 \times 0.02 + 0.1 \times 0.98} = 0.16 \end{aligned}$$

▲

Otro resultado importante sobre probabilidades condicionales que resulta útil para diversos cálculos es el siguiente. La demostración queda como ejercicio.

Proposición 1.1 *Sea A_1, \dots, A_n eventos cualesquiera. Entonces*

$$P(A_1 \cap \dots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_2 \cap A_1) \cdots P(A_n|A_1 \cap \dots \cap A_{n-1})$$

1.3. Variables Aleatorias

En este contexto, una variable aleatoria X es una función real definida sobre Ω que satisface ciertas condiciones de medibilidad que describimos a continuación. Está claro que si X toma valores reales, nos va a interesar calcular probabilidades del tipo $P(X \leq a)$, donde $a \in \mathbb{R}$. Por ejemplo, si X representa el

ingreso de una familia, o el número de piezas defectuosas en un lote, o el nivel máximo de un río durante cierto año, las probabilidades anteriores son obviamente de interés.

Ahora bien, para que estas probabilidades existan, es necesario que los conjuntos cuyas probabilidades deseamos calcular sean ‘medibles’, es decir, estén en la σ -álgebra \mathcal{F} . Estos conjuntos son de la forma $\{\omega : X(\omega) \leq a\}$, para $a \in \mathbb{R}$. Por lo tanto, la condición que tenemos que pedirle a X para garantizar que podemos asignar una probabilidad a todos estos conjuntos es la siguiente:

M1. Para todo número real a ,

$$\{\omega : X(\omega) \leq a\} \in \mathcal{F}.$$

Una función que satisface esta propiedad se dice que es *medible*. Por lo tanto, definimos una *variable aleatoria* como una función $X : \Omega \rightarrow \mathbb{R}$ que es medible. Usaremos las letras v.a. para abreviar variable aleatoria.

La medibilidad de una función $X : (\Omega, \mathcal{F}) \rightarrow \mathbb{R}$ depende de la σ -álgebra \mathcal{F} . Una función puede ser medible respecto a una σ -álgebra \mathcal{F}_1 y no respecto a otra \mathcal{F}_2 . Sin embargo, está claro a partir de la definición, que si X es medible respecto a \mathcal{F}_1 y $\mathcal{F}_1 \subset \mathcal{F}_2$, entonces X también es medible respecto a \mathcal{F}_2 . La menor σ -álgebra respecto a la cual una variable aleatoria X es medible se conoce como la σ -álgebra *generada por X* , y se denota por $\sigma(X)$. Esta σ -álgebra no es otra cosa que la intersección de todas las σ -álgebras respecto a las cuales X es medible.

Por ejemplo, si X sólo tiene una cantidad numerable de valores posibles x_1, x_2, \dots , los conjuntos

$$A_i = \{\omega : X(\omega) = x_i\}, \quad i = 1, 2, \dots$$

forman una partición numerable de Ω , es decir,

$$\Omega = \bigcup_{i=1}^{\infty} A_i, \quad \text{y} \quad A_i \cap A_j = \emptyset \quad \text{si } i \neq j.$$

En este caso \mathcal{F} está compuesta por los conjuntos \emptyset, Ω y por todos los conjuntos que sean unión de algunos de los A_i .

Ejemplo 1.5

Veamos un ejemplo sencillo. Consideremos el lanzamiento de un dado con el espacio de probabilidad descrito en el ejemplo 1.1 y consideremos la función $X : \Omega \rightarrow \{0, 1\}$ definida de la siguiente manera:

$$X(\omega) = \begin{cases} 1 & \text{si } \omega \text{ es par,} \\ 0 & \text{si } \omega \text{ es impar.} \end{cases}$$

La σ -álgebra generada por X está formada por los conjuntos \emptyset y Ω y por las preimágenes de los valores de la función, que son, respectivamente, los números pares y los impares en Ω :

$$X^{-1}(0) = \{1, 3, 5\}, \quad X^{-1}(1) = \{2, 4, 6\}.$$

Por lo tanto

$$\sigma(X) = \{\emptyset; \{1, 3, 5\}; \{2, 4, 6\}; \Omega\}.$$

▲

Ejemplo 1.6

Consideremos el mismo espacio Ω del ejemplo anterior y sean

$$\mathcal{A} = \mathcal{P}(\Omega), \quad \text{y} \quad \mathcal{F} = \{\emptyset; \{1, 3, 5\}; \{2, 4, 6\}; \Omega\}$$

dos σ -álgebras de subconjuntos de Ω . Sea $X : \Omega \rightarrow \{0, 1\}$ la función definida por $X = \mathbf{1}_{\{1,2,3\}}$ donde $\mathbf{1}_A$ es la función indicadora del conjunto A ($\mathbf{1}_A(\omega) = 1$ si $\omega \in A$ y $\mathbf{1}_A(\omega) = 0$ si $\omega \notin A$). Entonces X es medible respecto a \mathcal{A} pero no respecto a \mathcal{F} .

▲

No es difícil demostrar que si la condición M1 se satisface entonces para cualquier intervalo real I se tiene que

$$\{\omega : X(\omega) \in I\} \in \mathcal{F}, \quad (1.7)$$

y recíprocamente, si (1.7) vale entonces la condición M1 es cierta. Es un poco más difícil demostrar, pero también es cierto, que la condición M1 es equivalente a (1.7) si reemplazamos el intervalo I por un boreliano B . Resumiendo tenemos las siguientes equivalencias:

$$\begin{aligned} \{\omega : X(\omega) \leq a\} \in \mathcal{F}, \text{ para todo } a \in \mathbb{R} &\Leftrightarrow \{\omega : X(\omega) \in I\} \in \mathcal{F}, \text{ para todo intervalo } I \subset \mathbb{R} \\ &\Leftrightarrow \{\omega : X(\omega) \in B\} \in \mathcal{F}, \text{ para todo } B \in \mathcal{B}. \end{aligned}$$

1.4. Distribución de una Variable Aleatoria

Consideremos un espacio de probabilidad (Ω, \mathcal{F}, P) sobre el cual hemos definido una variable aleatoria X con valores reales. Si A es un intervalo de \mathbb{R} y queremos calcular la probabilidad de que la variable X tome valores en A , tenemos que considerar el conjunto $\{\omega : X(\omega) \in A\} = X^{-1}(A)$, que es la pre-imagen de A por la función X . Como la función X es medible, este conjunto está en la colección \mathcal{F} de los conjuntos medibles y en consecuencia podemos calcular su probabilidad. Por lo tanto, si A es un intervalo

$$P(X \in A) = P(\{\omega \in \Omega : X(\omega) \in A\}) = P(X^{-1}(A)). \quad (1.8)$$

Es posible demostrar que esta definición también funciona para conjuntos más complicados. Si llamamos \mathcal{B} a la σ -álgebra generada por los intervalos de \mathbb{R} (como mencionamos anteriormente, \mathcal{B} se conoce como la σ -álgebra de Borel y sus conjuntos son los borelianos de \mathbb{R}) entonces la relación (1.8) vale para todo $A \in \mathcal{B}$.

Esta relación nos permite definir una (medida de) probabilidad P_X sobre $(\mathbb{R}, \mathcal{B})$ inducida por la variable X , de la siguiente manera: Para todo $A \in \mathcal{B}$,

$$P_X(A) = P(X \in A) = P(\{\omega \in \Omega : X(\omega) \in A\}). \quad (1.9)$$

Esta (medida de) probabilidad se conoce como la *distribución* o la *ley* de X y en ocasiones se usa la notación $\mathcal{L}(X)$. Esta probabilidad contiene toda la información probabilística sobre la variable X .

1.5. Funciones de Distribución

Si en la relación (1.9) consideramos subconjuntos A de \mathbb{R} de la forma $(-\infty, x]$ obtenemos la siguiente función $F : \mathbb{R} \rightarrow \mathbb{R}$, que se conoce como la *función de distribución* de X :

$$F(x) = P_X((-\infty, x]) = P(X \leq x).$$

Si tenemos varias variables aleatorias en el mismo contexto, puede resultar útil distinguir sus funciones de distribución usando la notación F_X para la función de distribución de X . Usaremos las letras f.d. para abreviar función de distribución.

Está claro que si conocemos la distribución de una variable aleatoria, entonces podemos determinar la función de distribución correspondiente. El recíproco también es cierto, pero la demostración de este hecho requiere herramientas de teoría de la medida que no están a nuestra disposición en este curso.

Una función de distribución F tiene las siguientes tres propiedades,

FD1. F es continua por la derecha y tiene límites por la izquierda.

FD2. F es creciente (en sentido amplio).

FD3. Si F es una función de distribución,

$$\lim_{x \rightarrow -\infty} F(x) = 0, \quad \lim_{x \rightarrow \infty} F(x) = 1.$$

Estas tres propiedades caracterizan a las funciones de distribución: Si una función F satisface estas tres propiedades entonces existe una variable aleatoria X definida sobre un espacio de probabilidad (Ω, \mathcal{F}, P) tal que F es la función de distribución de X . Más aún, es posible tomar $\Omega = \mathbb{R}$.

1.5.1. Variables Discretas

Una variable aleatoria X es *discreta* si toma valores en un conjunto finito $\{x_1, x_2, \dots, x_n\}$ o numerable $\{x_1, x_2, \dots\}$. En el primer caso existen números positivos p_1, \dots, p_n con $p_1 + \dots + p_n = 1$, tales que

$$P(X = x_i) = p_i. \quad (1.10)$$

para $1 \leq i \leq n$. Llamaremos a esta función $p_X(x_i) = p_i, 1 \leq i \leq n$, la *función de probabilidad* o *densidad* de X . De manera similar, en el segundo caso tenemos números positivos $p_i, i \geq 1$ con $\sum_{i=1}^{\infty} p_i = 1$ que satisfacen (1.10) para $i \geq 1$.

En ambos casos las funciones de distribución son funciones de saltos, es decir, funciones que sólo crecen por saltos y que son constantes entre saltos consecutivos. Los saltos están dados por

$$p(x_i) = F(x_i) - F(x_i^-)$$

es decir, que los saltos ocurren en los valores de la función en los puntos x_i y su altura es igual a la probabilidad de este valor. Además

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

Ejemplo 1.7

Una variable con distribución uniforme en el conjunto $\{1, 2, 3, 4\}$ tiene función de probabilidad

$$p(j) = P(X = j) = \frac{1}{4} \quad \text{para } j = 1, 2, 3, 4.$$

En este caso la función de distribución F es una función escalera

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

▲

1.5.2. Variables Continuas

Una variable X es *continua* si su función de distribución F es continua. Una definición equivalente es que una v.a. X es continua si para cualquier valor x de la variable se tiene que $P(X = x) = 0$. Para la mayoría de estas variables existe una función $f : \mathbb{R} \rightarrow \mathbb{R}$ con $f(x) \geq 0$ para todo x y $\int_{\mathbb{R}} f(x) dx = 1$ que satisface

$$F(x) = \int_{-\infty}^x f(t) dt, \quad (1.11)$$

para todo $x \in \mathbb{R}$. La función f se conoce como la *densidad* de la variable X o de su f.d. F_X .

Hay algunas variables continuas que no tienen esta propiedad, es decir, que su función de distribución no puede obtenerse como la integral de otra función, pero su interés es más bien teórico y no las consideraremos en este curso. Se conocen como variables (o funciones de distribución) continuas singulares. De ahora en adelante todas las f.d. continuas que consideraremos satisfacen (1.11).

Si F es diferenciable en x entonces la densidad de probabilidad está dada por

$$f(x) = \frac{d}{dx}F(x) = F'(x), \quad -\infty < x < \infty.$$

Ejemplo 1.8

Consideremos una variable X con función de distribución

$$F(x) = \begin{cases} 0 & \text{para } x < 0, \\ x^2 & \text{para } 0 \leq x \leq 1, \\ 1 & \text{para } x \geq 1. \end{cases}$$

Esta función tiene derivada (salvo en el punto 1) que está dada por

$$f(x) = \begin{cases} 2x & \text{en } 0 < x < 1, \\ 0 & \text{en otro caso.} \end{cases}$$

que es la densidad en este caso.

Hay variables aleatorias 'mixtas', cuyas f.d. son la combinación de funciones continuas y saltos. ▲

Ejemplo 1.9

La función de distribución

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

es de este tipo. ▲

Dada cualquier función de distribución F es posible demostrar que existen dos funciones de distribución F_d y F_c , la primera discreta y la segunda continua, y un número $0 \leq \alpha \leq 1$ tales que $F = \alpha F_d + (1 - \alpha)F_c$. Esta descomposición es única.

1.6. Valores Esperados y Momentos

Si X es una v.a. discreta, el *momento de orden n* de X está dado por

$$E[X^n] = \sum_i x_i^n P(X = x_i) \tag{1.12}$$

siempre y cuando la serie en (1.12) converja absolutamente. Si esta serie diverge decimos que el momento no existe.

Si X es una v.a. continua con densidad $f(x)$, el *momento de orden n* de X está dado por

$$E[X^n] = \int_{-\infty}^{\infty} x^n f(x) dx, \quad (1.13)$$

siempre que esta integral converja absolutamente.

El primer momento, que corresponde a $n = 1$, se conoce como la *media*, el *valor esperado* o *esperanza* de X , y lo denotaremos por μ_X . El *momento central* o *centrado* de orden n es el momento de orden n de la variable $X - \mu_X$, siempre que μ_X exista. El primer momento central es cero. El segundo momento central se conoce como la *varianza* de X , denotado por $\text{Var}(X)$. Su raíz cuadrada es la *desviación típica*. Tenemos

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

La *mediana* de una v.a. X es cualquier valor ν con la propiedad de que

$$P(X \geq \nu) \geq \frac{1}{2}, \quad \text{y} \quad P(X \leq \nu) \geq \frac{1}{2}.$$

Si X es una variable aleatoria y g es una función (medible) entonces $g(X)$ también es una variable aleatoria. Si X es discreta y toma valores $x_j, j \geq 1$ entonces el valor esperado de $g(X)$ está dado por

$$E[g(X)] = \sum_{j=1}^{\infty} g(x_j) P(X = x_j) \quad (1.14)$$

siempre que la suma converja absolutamente. Si X es continua y tiene densidad f , el valor esperado de $g(X)$ es

$$E[g(X)] = \int g(x) f_X(x) dx. \quad (1.15)$$

Una fórmula general que abarca ambos casos (y también los casos mixtos) es la siguiente

$$E[g(X)] = \int g(x) dF_X(x), \quad (1.16)$$

donde F_X es la f.d. de la variable X . La integral que aparece en la fórmula (1.16) es una integral de Lebesgue-Stieltjes, cuya definición está más allá del nivel de este curso. Para nuestros efectos, interpretamos esta integral como la suma que aparece en la fórmula (1.14) si la variable es discreta y como la integral de la ecuación (1.15) si es continua.

A continuación presentamos dos desigualdades sencillas pero fundamentales.

Teorema 1.2 (Desigualdad de Markov) *Si X es una variable aleatoria que satisface $X \geq 0$, entonces, para cualquier $a > 0$,*

$$P(X \geq a) \leq \frac{E[X]}{a}.$$

Demostración. Haremos la demostración en el caso continuo. Si X tiene densidad f ,

$$\begin{aligned} E[X] &= \int_0^{\infty} x f(x) dx = \int_0^a x f(x) dx + \int_a^{\infty} x f(x) dx \\ &\geq \int_a^{\infty} x f(x) dx \geq \int_a^{\infty} a f(x) dx \\ &= a \int_a^{\infty} f(x) dx = a P(X \geq a). \end{aligned}$$

■

Corolario 1.1 (Desigualdad de Chebyshev) Si X es una variable aleatoria con media μ y varianza σ^2 , para cualquier valor de $x > 0$,

$$P(|X - \mu| \geq x) \leq \frac{\sigma^2}{x^2}.$$

Demostración. Como $(X - \mu)^2$ es una v.a. no negativa, podemos aplicar la desigualdad de Markov con $a = x^2$ y obtenemos

$$P((X - \mu)^2 \geq x^2) \leq \frac{E[(X - \mu)^2]}{x^2}.$$

Pero como $(X - \mu)^2 \geq x^2$ sí y sólo sí $|X - \mu| > x$, la desigualdad anterior equivale a

$$P(|X - \mu| \geq x) \leq \frac{E[(X - \mu)^2]}{x^2} = \frac{\sigma^2}{x^2}.$$

■

Para concluir esta sección enunciamos dos resultados fundamentales que son válidos para la integral de Lebesgue-Stieltjes. De nuevo, las demostraciones correspondientes están más allá del nivel de este curso.

Teorema 1.3 (Convergencia Monótona) Si X_1, X_2, \dots es una sucesión de variables aleatorias acotadas inferiormente que satisface $X_1 \leq X_2 \leq X_3 \dots$ y $X_n \uparrow X$, entonces

$$E[X_n] \uparrow E[X].$$

Teorema 1.4 (Convergencia Dominada) Sea X_i , $i \geq 1$ una sucesión de variables aleatorias que satisfacen $|X_i| \leq Y$, donde Y es una v.a. con $E[Y] < \infty$. Si

$$\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)$$

para casi todo ω (es decir, para todo $\omega \in \Omega$ fuera de un subconjunto de probabilidad 0), entonces

$$\lim_{n \rightarrow \infty} E[X_n] = E[X].$$

1.7. Distribuciones Conjuntas e Independencia

Si tenemos un par de variables aleatorias (X, Y) definidas sobre un espacio de probabilidad (Ω, \mathcal{F}, P) , su función de distribución conjunta F_{XY} está definida por

$$F_{XY}(x, y) = F(x, y) = P(X \leq x, Y \leq y).$$

Si ambas variables son discretas y toman valores x_i , $i \geq 1$ e y_j , $j \geq 1$ respectivamente, la función de probabilidad conjunta de X e Y es

$$p_{XY}(x_i, y_j) = P(X = x_i, Y = y_j), \quad i \geq 1, j \geq 1.$$

Una función de distribución conjunta tiene densidad (conjunta) si existe una función f_{XY} de dos variables que satisfice

$$F_{XY}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{XY}(s, t) dt ds, \quad \text{para todo } x, y.$$

La función $F_X(x) = \lim_{y \rightarrow \infty} F(x, y)$ es una f.d. que se conoce como la *función de distribución marginal* de X . Si las variables aleatorias son ambas discretas, las funciones de probabilidad marginales están dadas por

$$p_X(x_i) = \sum_{j=1}^{\infty} p_{XY}(x_i, y_j) \quad y \quad p_Y(y_j) = \sum_{i=1}^{\infty} p_{XY}(x_i, y_j).$$

Si la f.d. F tiene una densidad conjunta f , las densidades marginales de X e Y están dadas, respectivamente, por

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy \quad y \quad f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx.$$

Si X e Y tienen distribución conjunta entonces

$$E[X + Y] = E[X] + E[Y]$$

siempre y cuando todos estos momentos existan.

Independencia

Si para todos los valores de x e y se tiene que $F(x, y) = F_X(x) \times F_Y(y)$ decimos que las variables X e Y son *independientes*. Si las variables son discretas y tienen función de probabilidad conjunta p_{XY} , son independientes si y sólo si

$$p_{XY}(x, y) = p_X(x)p_Y(y). \quad (1.17)$$

De manera similar, si las variables son continuas y tienen densidad conjunta $f_{XY}(x, y)$, son independientes si y sólo si

$$f_{XY}(x, y) = f_X(x)f_Y(y). \quad (1.18)$$

Para una colección X_1, \dots, X_n de variables aleatorias la distribución conjunta se define como

$$F(x_1, \dots, x_n) = F_{X_1 \dots X_n}(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n).$$

Si

$$F_{X_1 \dots X_n}(x_1, \dots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n)$$

para todos los valores posibles de x_1, \dots, x_n decimos que las variables X_1, \dots, X_n son independientes.

Una función de distribución conjunta $F(x_1, \dots, x_n)$ tiene densidad de probabilidad $f(t_1, \dots, t_n)$ si

$$F(x_1, \dots, x_n) = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f(t_1, \dots, t_n) dt_1 \cdots dt_n,$$

para todos los valores de x_1, \dots, x_n .

Para variables X_1, \dots, X_n y funciones arbitrarias h_1, \dots, h_m de n variables,

$$E\left[\sum_{j=1}^m h_j(X_1, \dots, X_n)\right] = \sum_{j=1}^m E[h_j(X_1, \dots, X_n)],$$

siempre que todos estos momentos existan.

Proposición 1.2 *Si X, Y son v.a.i. con primer momento finito, entonces el producto XY también tiene primer momento finito y*

$$E(XY) = E(X)E(Y)$$

Este resultado se extiende a cualquier colección finita de variables independientes.

Demostración. Vamos a hacer la demostración en el caso discreto. Sean $\{x_i, i \geq 1\}$ y $\{y_j, j \geq 1\}$ los conjuntos de valores de X y Y , respectivamente, y sea $p_{XY}(x, y)$ la función de probabilidad conjunta con marginales $p_X(x)$ y $p_Y(y)$. por independencia tenemos que $p_{XY}(x, y) = p_X(x)p_Y(y)$.

El valor esperado del producto XY es

$$\begin{aligned} E(XY) &= \sum_{x,y} xyp_{XY}(x, y) \\ &= \sum_{x,y} xyp_X(x)p_Y(y) \end{aligned}$$

Podemos separar la suma sobre x de la suma sobre y y obtenemos

$$= \sum_x xp_X(x) \sum_y yp_Y(y) = E(X)E(Y)$$

■

Si X e Y son variables con distribución conjunta, medias μ_X , μ_Y , y varianzas finitas σ_X^2 , σ_Y^2 , la covarianza de X e Y , que escribimos σ_{XY} o $\text{Cov}(X, Y)$, está definida por

$$\sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)] = E[XY] - \mu_X\mu_Y,$$

y decimos que X e Y no están correlacionadas si su covarianza es cero, es decir, $\sigma_{XY} = 0$.

Las variables independientes con varianza finita no están correlacionadas, pero el recíproco no es cierto. Hay variables que no están correlacionadas pero no son independientes.

Ejemplo 1.10

Si X es una variable aleatoria con distribución normal típica $\mathcal{N}(0, 1)$ y $Y = X^2$ entonces $\text{Cov}(X, Y) = E(XY) = E(X^3)$ y es fácil demostrar que el tercer momento de una normal típica vale 0. Más aún, todos los momentos impares de una variable normal típica son nulos.

Un segundo ejemplo de variables que no están correlacionadas pero no son independientes es el siguiente: Sea U una variable con distribución uniforme en $[0, 2\pi]$ y definamos $X = \text{sen } U, Y = \text{cos } U$. Entonces $E(X) = E(Y) = 0$ y

$$\text{Cov}(X, Y) = E(XY) = \int_0^{2\pi} \text{sen}(u) \text{cos}(u) du = 0.$$

Dividiendo la covarianza σ_{XY} por las desviaciones típicas σ_X y σ_Y obtenemos el coeficiente de correlación $\rho_{X,Y}$:

$$\rho_{X,Y} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}$$

que satisface $-1 \leq \rho \leq 1$.

Corolario 1.2 Si X e Y son independientes y tienen varianzas respectivas σ_X^2 y σ_Y^2 entonces la varianza de la suma $Z = X + Y$ es la suma de las varianzas:

$$\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2.$$

Esta propiedad se extiende al caso de n variables independientes.

Demostración.

$$\begin{aligned}\text{Var}(X + Y) &= \text{E} \left[((X + Y) - \text{E}(X + Y))^2 \right] = \text{E} \left[(X - \text{E}(X) + Y - \text{E}(Y))^2 \right] \\ &= \text{E} \left[(X - \text{E}(X))^2 \right] + \text{E} \left[(Y - \text{E}(Y))^2 \right] + 2\text{E} \left[(X - \text{E}(X))(Y - \text{E}(Y)) \right] \\ &= \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y).\end{aligned}$$

Este resultado es general, pues no hemos usado hasta ahora la independencia de las variables. Si las variables no están correlacionadas, y en particular si son independientes, la covarianza vale 0 y se tiene el resultado del teorema. ■

Sumas y Convolutiones

Si X e Y son variables aleatorias independientes con f.d. F_X y F_Y , respectivamente, entonces la f.d. de la suma $Z = X + Y$ es la convolución de F_X y F_Y :

$$F_Z(z) = \int_{-\infty}^{\infty} F_X(z-t) dF_Y(t) = \int_{-\infty}^{\infty} F_Y(z-t) dF_X(t).$$

Si X e Y toman valores en los enteros no-negativos con funciones de probabilidad respectivas p_X y p_Y entonces

$$p_Z(n) = P(Z = n) = \sum_{i=0}^n P(X = i)P(Y = n - i) = \sum_{i=0}^n p_X(i)p_Y(n - i) = \sum_{i=0}^n p_X(n - i)p_Y(i).$$

Si consideramos la situación en la cual X tienen Y densidades f_X y f_Y , respectivamente, la densidad f_Z de la suma es la convolución de las densidades f_X y f_Y :

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z-t)f_Y(t) dt = \int_{-\infty}^{\infty} f_Y(z-t)f_X(t) dt.$$

1.8. Algunas Distribuciones Importantes

1.8.1. Distribuciones Discretas

Distribución de Bernoulli

Una variable aleatoria de Bernoulli toma valores 1 y 0 con probabilidades respectivas p y $q = 1 - p$, donde $0 < p < 1$. Si el resultado del experimento es 1 decimos que ha ocurrido un *éxito* y p es entonces la probabilidad de éxito. La media y varianza son, respectivamente,

$$\text{E}[X] = p, \quad \text{Var}[X] = p(1 - p).$$

Si A es un evento y $\mathbf{1}_A$ es la función indicadora de A , entonces $\mathbf{1}_A$ es una variable de Bernoulli con parámetro $p = \text{E}[\mathbf{1}_A] = P(A)$.

Distribución Binomial

Consideremos una colección X_1, X_2, \dots, X_n de variables independientes de Bernoulli con probabilidad de éxito p . Sea S el total de éxitos en los n experimentos de Bernoulli, es decir, $S = \sum_1^n X_i$. La distribución de S es binomial con parámetros n y p , es decir, la función de probabilidad es

$$p_S(k) = P(S = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad \text{para } k = 0, 1, \dots, n. \quad (1.19)$$

Usaremos la notación $X \sim \text{Bin}(n, p)$ para indicar que la variable X tiene distribución binomial de parámetros n y p .

Podemos determinar el valor esperado usando la definición

$$E[S] = E\left[\sum_1^n X_i\right] = \sum_1^n E[X_i] = np,$$

y usando independencia podemos calcular la varianza,

$$\text{Var}[S] = \text{Var}\left[\sum_1^n X_i\right] = \sum_1^n \text{Var}[X_i] = np(1-p).$$

Distribución Geométrica

En el mismo esquema anterior, sea Z el número de ensayos antes del primer éxito, es decir, $Z = k$ si y sólo si los primeros $k-1$ ensayos resultan en fracaso, cada uno con probabilidad $(1-p)$ y el k -ésimo es un éxito. En este caso Z tiene distribución geométrica con función de probabilidad

$$p_Z(k) = p(1-p)^{k-1}, \quad \text{para } k = 1, 2, \dots \quad (1.20)$$

y los primeros dos momentos son

$$E[Z] = \frac{1}{p}, \quad \text{Var}[Z] = \frac{1-p}{p^2}.$$

En ciertas ocasiones se define la distribución geométrica asociada al número de fracasos hasta el primer éxito, es decir, $Z' = Z - 1$. Con esta definición la función de probabilidad es

$$p_{Z'}(k) = p(1-p)^k, \quad \text{para } k = 1, 2, \dots \quad (1.21)$$

En este caso $E[Z'] = E[Z] - 1 = (1-p)/p$ y $\text{Var}[Z'] = \text{Var}[Z] = (1-p)/p^2$.

Distribución Binomial Negativa

Sea ahora W_k el número de ensayos antes del k -ésimo éxito, para un entero fijo $k \geq 1$. La distribución de esta variable se conoce como la binomial negativa de parámetros k y p . Para que W_k tome el valor r es necesario que haya exactamente $k-1$ éxitos en los primeros $r-1$ ensayos y luego un éxito en el ensayo r . La probabilidad de la primera condición es binomial mientras que la probabilidad de la segunda es p , lo cual nos da la función de probabilidad de W_k :

$$P(W_k = r) = \binom{r-1}{k-1} p^k (1-p)^{r-k}, \quad r = k, k+1, k+2, \dots$$

Otra manera de escribir W_k es como la suma de k variables independientes con distribución geométrica (1.20): $W_k = Z_1 + \dots + Z_k$. Usando esta relación es inmediato que

$$E[W_k] = \frac{k}{p}, \quad \text{Var}[W_k] = \frac{k(1-p)}{p^2}.$$

Distribución de Poisson

La distribución de Poisson de parámetro $\lambda > 0$ tiene función de probabilidad

$$p(k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad \text{para } k = 0, 1, \dots \quad (1.22)$$

Usaremos la notación $X \sim \text{Pois}(\lambda)$ para esta distribución. El desarrollo en serie de potencias de la función exponencial es

$$e^\lambda = 1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \dots$$

y vemos que $\sum_0^\infty p(k) = 1$. Usando de nuevo este desarrollo podemos calcular el valor esperado para una variable X con esta distribución

$$E[X] = \sum_{k=0}^{\infty} k p(k) = \sum_{k=1}^{\infty} k \frac{\lambda^k e^{-\lambda}}{k!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda.$$

La misma idea nos permite calcular

$$E[X(X-1)] = \sum_{k=0}^{\infty} k(k-1)p(k) = \sum_{k=2}^{\infty} k(k-1) \frac{\lambda^k e^{-\lambda}}{k!} = \lambda^2 e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} = \lambda^2.$$

A partir de esta relación obtenemos $E[X(X-1)] = E[X^2] - E[X] = \lambda^2$, de donde $E[X^2] = \lambda^2 + \lambda$ y $\text{Var}[X] = E[X^2] - (E[X])^2 = \lambda$.

Entre otras razones, la distribución de Poisson es importante porque aparece como límite de la distribución binomial cuando $n \rightarrow \infty$ y $p \rightarrow 0$ de modo que $np \rightarrow \lambda > 0$. Este resultado se conoce como la ley de 'eventos raros'.

Distribución Multinomial

Las variables X_1, \dots, X_k , con valores en el conjunto $\{0, 1, \dots, n\}$, tienen una distribución multinomial si su función de probabilidad conjunta es

$$P(X_1 = r_1, \dots, X_k = r_k) = \begin{cases} \frac{n!}{r_1! \dots r_k!} p_1^{r_1} \dots p_k^{r_k}, & \text{si } r_1 + \dots + r_k = n, \\ 0 & \text{si no.} \end{cases}$$

donde $p_i > 0$ para $i = 1, \dots, k$ y $p_1 + \dots + p_k = 1$.

Para esta distribución $E[X_i] = np_i$, $\text{Var}[X_i] = np_i(1 - p_i)$ y $\text{Cov}(X_i X_j) = -np_i p_j$.

1.8.2. Distribuciones Continuas

Distribución Normal o Gaussiana

Una variable X tiene distribución normal de parámetros μ y σ^2 si su densidad es

$$\varphi(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}, \quad -\infty < x < \infty.$$

Usaremos la notación $X \sim \mathcal{N}(\mu, \sigma^2)$. Los parámetros μ y σ^2 representan el valor esperado y la varianza de la variable X . La densidad $\varphi(x; \mu, \sigma^2)$ es simétrica respecto a μ .

El caso $\mu = 0, \sigma^2 = 1$ se conoce como la densidad normal típica o estándar. Si $X \sim \mathcal{N}(\mu, \sigma^2)$ entonces $Z = (X - \mu)/\sigma$ tiene una distribución normal típica. De esta manera los cálculos de probabilidad siempre pueden reducirse al caso estándar. La densidad típica es

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad -\infty < x < \infty,$$

y la función de distribución correspondiente es

$$\Phi(x) = \int_{-\infty}^x \varphi(t) dt, \quad -\infty < x < \infty.$$

Distribución Lognormal

Si $\log V$ tiene una distribución normal, decimos que V tiene distribución lognormal. Recíprocamente, si $X \sim \mathcal{N}(\mu, \sigma^2)$ entonces $V = e^X$ es una variable con distribución lognormal. Haciendo un cambio de variable para la densidad obtenemos

$$f_V(v) = \frac{1}{\sqrt{2\pi\sigma v}} \exp\left\{-\frac{1}{2}\left(\frac{\log v - \mu}{\sigma}\right)^2\right\}, \quad v \geq 0.$$

La media y varianza son, respectivamente,

$$\begin{aligned} E[V] &= \exp\left\{\mu + \frac{1}{2}\sigma^2\right\} \\ \text{Var}[V] &= \exp\left\{2\left(\mu + \frac{1}{2}\sigma^2\right)\right\}(e^{\sigma^2} - 1) \end{aligned}$$

Distribución Exponencial

Una variable T no-negativa tiene una distribución exponencial con parámetro $\lambda > 0$ si su densidad es

$$f_T(t) = \begin{cases} \lambda e^{-\lambda t} & \text{para } t \geq 0, \\ 0 & \text{para } t < 0. \end{cases}$$

La función de distribución correspondiente es

$$F_T(t) = \begin{cases} 1 - e^{-\lambda t} & \text{para } t \geq 0, \\ 0 & \text{para } t < 0. \end{cases}$$

Usaremos la notación $X \sim \mathcal{Exp}(\lambda)$. La media y varianza están dadas por

$$E[T] = \frac{1}{\lambda}, \quad \text{Var}[T] = \frac{1}{\lambda^2}.$$

Una de las propiedades fundamentales de distribución exponencial es la falta de memoria, que explicamos a continuación. Supongamos que T es el tiempo de vida de cierto componente y, dado que el componente ha durado hasta el instante t queremos obtener la distribución condicional del tiempo de vida remanente $T - t$. Equivalentemente, para $x > 0$ queremos determinar la distribución condicional $P(T > t + x | T > t)$. Aplicando la definición de probabilidad condicional obtenemos

$$\begin{aligned} P(T > t + x | T > t) &= \frac{P(T > t + x, T > t)}{P(T > t)} \\ &= \frac{P(T > t + x)}{P(T > t)} \\ &= \frac{e^{-\lambda(t+x)}}{e^{-\lambda t}} = e^{-\lambda x}. \end{aligned}$$

Por lo tanto,

$$P(T > t + x | T > t) = e^{-\lambda x} = P(T > x)$$

y un componente que ha durado hasta el instante t tiene una vida remanente que es estadísticamente igual a la de un componente nuevo.

La *función de riesgo* ('hazard') o *de falla* $r(s)$ de una variable no-negativa S con densidad continua $g(s)$ y f.d. $G(s) < 1$ se define como

$$r(s) = \frac{g(s)}{1 - G(s)}, \quad \text{para } s > 0.$$

Calculemos ahora

$$\begin{aligned} P(s < S \leq s + \Delta s | S > s) &= \frac{P(s < S \leq s + \Delta s)}{P(S > s)} \\ &= \frac{g(s)\Delta s}{1 - G(s)} + o(\Delta s) \\ &= r(s)\Delta s + o(\Delta s). \end{aligned}$$

Por lo tanto un componente que ha durado hasta el tiempo s fallará en el intervalo $(s, s + \Delta s]$ con probabilidad condicional $r(s)\Delta s + o(\Delta s)$, lo que motiva el nombre de función de riesgo o fallas.

Podemos invertir la relación de la definición integrando

$$-r(s) = \frac{-g(s)}{1 - G(s)} = \frac{d[1 - G(s)]/ds}{1 - G(s)} = \frac{d(\log(1 - G(s)))}{ds}$$

para obtener

$$-\int_0^t r(s) ds = \log[1 - G(t)],$$

o

$$G(t) = 1 - \exp \left\{ -\int_0^t r(s) ds \right\}, \quad t \geq 0,$$

que nos da la f.d. explícitamente en términos de la tasa de fallas.

La distribución exponencial puede ser caracterizada como la única distribución continua que tiene tasa de fallas constante $r(t) = \lambda$. La tasa de fallas no cambia en el tiempo, otra consecuencia de la propiedad de falta de memoria.

Distribución Uniforme

Una variable aleatoria U tiene distribución uniforme en el intervalo $[a, b]$, con $a < b$, si tiene la densidad de probabilidad

$$f_U(u) = \begin{cases} \frac{1}{b-a} & \text{para } a \leq u \leq b, \\ 0 & \text{en otro caso.} \end{cases}$$

La función de distribución es

$$F_U(x) = \begin{cases} 0 & \text{para } x \leq a, \\ \frac{x-a}{b-a} & \text{para } a < x \leq b, \\ 1 & \text{para } x > b. \end{cases}$$

Usaremos la notación $U \sim \mathcal{U}[a, b]$. La media y varianza son

$$E[U] = \frac{1}{2}(a + b) \quad \text{y} \quad \text{Var}[U] = \frac{(b - a)^2}{12}.$$

Distribución Gamma

La distribución Gamma con parámetros $\alpha > 0$ y $\lambda > 0$ tiene densidad de probabilidad

$$f(x) = \frac{\lambda}{\Gamma(\alpha)} (\lambda x)^{\alpha-1} e^{-\lambda x}, \quad \text{para } x > 0,$$

donde

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx, \quad \text{para } \alpha > 0.$$

Usaremos la notación $X \sim \Gamma(\alpha, \lambda)$ para esta distribución.

Si α es entero y sumamos α variables exponenciales independientes con parámetro λ , esta suma X_α tiene distribución Gamma con parámetros α y λ . Los parámetros de esta distribución son

$$E[X_\alpha] = \frac{\alpha}{\lambda} \quad \text{Var}[X_\alpha] = \frac{\alpha}{\lambda^2}.$$

Distribución Beta

La densidad beta con parámetros $\alpha > 0$ y $\beta > 0$ es

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

Los parámetros de esta distribución son

$$E[X] = \frac{\alpha}{\alpha + \beta} \quad \text{Var}[X] = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}.$$

Distribución Normal Bivariada

Sean $\sigma_X > 0$, $\sigma_Y > 0$, μ_X , μ_Y y ρ con $-1 < \rho < 1$ números reales. Para x e y reales definimos la forma cuadrática

$$Q(x, y) = \frac{1}{1 - \rho^2} \left\{ \left(\frac{x - \mu_X}{\sigma_X} \right)^2 - 2\rho \left(\frac{x - \mu_X}{\sigma_X} \right) \left(\frac{y - \mu_Y}{\sigma_Y} \right) + \left(\frac{y - \mu_Y}{\sigma_Y} \right)^2 \right\}$$

Definimos la distribución normal o gaussiana conjunta para las variables X e Y por la función de densidad

$$\phi_{X,Y}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2}Q(x, y) \right\},$$

para $-\infty < x, y < \infty$. Los momentos de la distribución son

$$E[X] = \mu_X, \quad E[Y] = \mu_Y, \quad \text{Var}[X] = \sigma_X^2, \quad \text{Var}[Y] = \sigma_Y^2,$$

y

$$\text{Cov}[X, Y] = E[(X - \mu_X)(Y - \mu_Y)] = \rho\sigma_X\sigma_Y.$$

Si definimos la matriz de varianzas y covarianzas del vector (X, Y) por

$$\Sigma = \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}$$

entonces podemos escribir la densidad anterior como

$$\phi_{\mathbf{X}}(\mathbf{x}) = \frac{1}{2\pi} \frac{1}{\sqrt{\det \Sigma}} \exp \left\{ -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\}.$$

donde $\mathbf{X} = (X, Y)$, $\mathbf{x} = (x, y)$, $\det \Sigma$ es el determinante de Σ , Σ^{-1} la matriz inversa y \mathbf{x}' es el vector traspuesto de \mathbf{x} . Esta expresión para la densidad se puede generalizar al caso de vectores gaussianos de dimensión n .

1.9. Probabilidad y Esperanza Condicional

1.9.1. El Caso Discreto

Definición 1.3 Sean X, Y variables aleatorias discretas. La función o densidad de probabilidad condicional $p_{X|Y}(x|y)$ de X dado $Y = y$ se define por

$$p_{X|Y}(x|y) = \frac{P(X = x, Y = y)}{P(Y = y)} \quad \text{si } P(Y = y) > 0,$$

y no está definida, o se le asigna un valor arbitrario, si $P(Y = y) = 0$.

En términos de la densidad conjunta $p_{X,Y}(x, y)$ y de la densidad marginal de Y , $p_Y(y) = \sum_x p_{X,Y}(x, y)$, la definición es

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x, y)}{p_Y(y)}, \quad \text{si } p_Y(y) > 0.$$

Observamos que $p_{X|Y}(x|y)$ es una densidad de probabilidad en x para cada y fijo:

$$p_{X|Y}(x|y) \geq 0, \quad \sum_x p_{X|Y}(x|y) = 1, \quad \text{para todo } y.$$

Por lo tanto, podemos definir la función de distribución condicional de X dado que $Y = y$ como la f.d. asociada a la función de probabilidad $p_{X|Y}(x|y)$ (siempre que $p_Y(y) > 0$):

$$F_{X|Y}(x|y) = \sum_{z \leq x} p_{X|Y}(z|y) = \frac{1}{p_Y(y)} \sum_{z \leq x} p_{X,Y}(z, y).$$

La ley de la probabilidad total es

$$P(X = x) = \sum_y P(X = x|Y = y)P(Y = y) = \sum_y p_{X|Y}(x|y)p_Y(y).$$

Ejemplo 1.11

Supongamos que X tiene distribución binomial de parámetros p y N , donde N a su vez tiene distribución de Poisson con media λ . ¿Cuál es la distribución de X ?

Tenemos que

$$p_{X|N}(k|n) = \binom{n}{k} p^k (1-p)^{n-k}, \quad \text{para } k = 0, 1, \dots, n$$

$$p_N(n) = \frac{\lambda^n e^{-\lambda}}{n!}, \quad \text{para } n = 0, 1, \dots$$

Usando la ley de la probabilidad total

$$\begin{aligned} P(X = k) &= \sum_{n=0}^{\infty} p_{X|N}(k|n)p_N(n) = \sum_{n=k}^{\infty} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \frac{\lambda^n e^{-\lambda}}{n!} \\ &= \frac{\lambda^k e^{-\lambda} p^k}{k!} \sum_{n=k}^{\infty} \frac{[\lambda(1-p)]^{n-k}}{(n-k)!} = \frac{(\lambda p)^k e^{-\lambda}}{k!} e^{\lambda(1-p)} \\ &= \frac{(\lambda p)^k e^{-\lambda p}}{k!} \end{aligned}$$

para $k = 0, 1, \dots$, es decir, $X \sim \text{Pois}(\lambda p)$. ▲

Sea g una función tal que $E[|g(X)|] < \infty$. Definimos la esperanza condicional de $g(X)$ dado $Y = y$ por la fórmula

$$E[g(X)|Y = y] = \sum_x g(x)p_{X|Y}(x|y) \quad \text{si } p_Y(y) > 0,$$

y la esperanza condicional no está definida para valores y tales que $p_Y(y) = 0$. La ley de la probabilidad total para esperanzas condicionales es

$$E[g(X)] = \sum_y E[g(X)|Y = y]p_Y(y).$$

La esperanza condicional $E[g(X)|Y = y]$ es una función de la variable real y , que denotaremos $\varphi(y)$. Si evaluamos esta función φ en la variable aleatoria Y obtenemos una nueva variable aleatoria $\varphi(Y)$, que denotamos $E[g(X)|Y]$:

$$E[g(X)|Y](\omega) = E[g(X)|Y = Y(\omega)].$$

Podemos ahora escribir la ley de la probabilidad total como

$$E[g(X)] = E[E[g(X)|Y]]$$

Ejemplo 1.12

Consideremos un dado tetrahedral con 4 resultados posibles: 1, 2, 3 y 4 y probabilidades respectivas $p(i) = p_i$ para $i = 1, \dots, 4$. Lanzamos el dado dos veces y definimos X como el producto de los resultados y Y como su suma. A continuación presentamos una descripción de los resultados posibles del experimento y de los valores de las variables X e Y .

$\Omega :$	(1, 1)	(1, 2)	(1, 3)	(1, 4)	$X :$	1	2	3	4	$Y :$	2	3	4	5
	(2, 1)	(2, 2)	(2, 3)	(2, 4)		2	4	6	8		3	4	5	6
	(3, 1)	(3, 2)	(3, 3)	(3, 4)		3	6	9	12		4	5	6	7
	(4, 1)	(4, 2)	(4, 3)	(4, 4)		4	8	12	16		5	6	7	8

Calculemos ahora la probabilidad condicional $p_{X|Y}(x|y)$ para algunos valores de y . Por ejemplo, si $Y = 2$, el único resultado posible es (1, 1) y el valor de X en este caso es 1. Por lo tanto

$$p_{X|Y}(x|2) = \begin{cases} 1 & \text{si } x = 1, \\ 0 & \text{en otro caso.} \end{cases}$$

Algo similar ocurre cuando $y = 3, 7$ u 8 : Para cada uno de estos valores de la variable Y hay un sólo valor de la variable X , que por lo tanto ocurre condicionalmente con probabilidad 1.

Para los otros valores de y la situación es distinta, pues hay varios valores posibles de x . Veamos, como ejemplo, el caso $y = 5$; tenemos dos valores posibles de X : 4, que corresponde a los eventos elementales (4, 1) y (1, 4), y 6, que corresponde a los eventos elementales (3, 2) y (2, 3). Por lo tanto,

$$p_{X|Y}(4|5) = \frac{P(X = 4, Y = 5)}{P(Y = 5)} = \frac{p_1 p_4}{p_1 p_4 + p_2 p_3},$$

$$p_{X|Y}(6|5) = \frac{P(X = 6, Y = 5)}{P(Y = 5)} = \frac{p_2 p_3}{p_1 p_4 + p_2 p_3}.$$

De manera similar se calculan los otros valores de la función de probabilidad condicional, que son para $Y = 4$:

$$p_{X|Y}(3|4) = \frac{2p_1 p_3}{2p_1 p_3 + p_2^2}, \quad p_{X|Y}(4|4) = \frac{p_2^2}{2p_1 p_3 + p_2^2},$$

para $Y = 6$:

$$p_{X|Y}(8|6) = \frac{2p_2p_4}{2p_2p_4 + p_3^2}, \quad p_{X|Y}(9|6) = \frac{p_3^2}{2p_2p_4 + p_3^2}.$$

En consecuencia vemos que para cada valor de la variable Y tenemos una función de probabilidad sobre los posibles valores de X . Veamos ahora los distintos valores de la esperanza condicional,

$$\begin{aligned} E[X|Y = 2] &= 1, & E[X|Y = 3] &= 2 \\ E[X|Y = 4] &= 3 \frac{2p_1p_3}{2p_1p_3 + p_2^2} + 4 \frac{p_2^2}{2p_1p_3 + p_2^2} \\ E[X|Y = 5] &= 4 \frac{p_1p_4}{p_1p_4 + p_2p_3} + 6 \frac{p_2p_3}{p_1p_4 + p_2p_3} \\ E[X|Y = 6] &= 8 \frac{2p_2p_4}{2p_2p_4 + p_3^2} + 9 \frac{p_3^2}{2p_2p_4 + p_3^2} \\ E[X|Y = 7] &= 12, & E[X|Y = 8] &= 16. \end{aligned}$$

Para el caso particular en el cual el dado es simétrico y todos los valores tienen la misma probabilidad los valores de las tres esperanzas centrales en la expresión anterior son

$$E[X|Y = 4] = \frac{10}{3}; \quad E[X|Y = 5] = 5; \quad E[X|Y = 6] = \frac{25}{3}.$$

Por lo tanto, $E[X|Y]$ es una función de los valores de Y , y como Y es una variable aleatoria, también lo es $E[X|Y]$. La siguiente tabla muestra los valores de Y , los valores asociados de $E[X|Y]$ y las probabilidades correspondientes, y representa una descripción de la variable aleatoria $E[X|Y]$.

y	$E[X Y = y]$	$P(Y = y)$
2	1	1/16
3	2	1/8
4	10/3	3/16
5	5	1/4
6	25/3	3/16
7	12	1/8
8	16	1/16

▲

Propiedades.

Como la esperanza condicional de $g(X)$ dado $Y = y$ es la esperanza respecto a la densidad de probabilidad condicional $p_{X|Y}(x|y)$, las esperanzas condicionales se comportan en muchos aspectos como esperanzas ordinarias.

Suponemos que X e Y tienen distribución conjunta, $c \in \mathbb{R}$, g es una función tal que $E[|g(X)|] < \infty$, h es una función acotada y ν es una función en \mathbb{R}^2 tal que $E[|\nu(X, Y)|] < \infty$.

- $E[c_1g_1(X_1) + c_2g_2(X_2)|Y = y] = c_1 E[g_1(X_1)|Y = y] + c_2 E[g_2(X_2)|Y = y]$.
- Si $g \geq 0$ entonces $E[g(X)|Y = y] \geq 0$.
- $E[\nu(X, Y)|Y = y] = E[\nu(X, y)|Y = y]$.
- $E[g(X)|Y = y] = E[g(X)]$ si X e Y son v.a.i.
- $E[g(X)h(Y)|Y = y] = h(y) E[g(X)|Y = y]$.

$$6. E[g(X)h(Y)] = \sum_y h(y) E[g(X)|Y = y]p_Y(y) = E[h(Y) E[g(X)|Y]].$$

Como consecuencia de 1, 5 y 6 obtenemos

$$7. E[c|Y = y] = c,$$

$$8. E[h(Y)|Y = y] = h(y),$$

$$9. E[g(X)] = \sum_y E[g(X)|Y = y]p_Y(y) = E[E[g(X)|Y]].$$

Ejemplo 1.13

Usando la propiedad 9 podemos obtener la esperanza de X en el ejemplo anterior:

$$E[X] = E[E[X|Y]] = 1 \times \frac{1}{16} + 2 \times \frac{1}{8} + \dots + 16 \times \frac{1}{16} = 6.25$$

y hemos hallado la esperanza de X sin haber usado su función de distribución. ▲

1.9.2. El Caso Continuo

Sean X, Y v.a. con distribución conjunta continua de densidad $f_{X,Y}(x, y)$. Definimos la densidad condicional $f_{X|Y}(x|y)$ para la variable X dado que $Y = y$ por la fórmula

$$f_{X|Y}(x|y) = \frac{f_{XY}(x, y)}{f_Y(y)} \quad \text{si } f_Y(y) > 0,$$

y no está definida si $f_Y(y) = 0$. La f.d. condicional para X dado $Y = y$ se define por

$$F_{X|Y}(x|y) = \int_{-\infty}^x \frac{f_{XY}(s, y)}{f_Y(y)} ds \quad \text{si } f_Y(y) > 0.$$

La densidad condicional tiene las propiedades que uno esperaría. En particular

$$P(a < X < b, c < Y < d) = \int_c^d \left(\int_a^b f_{X|Y}(x|y) dx \right) f_Y(y) dy$$

y haciendo $c = -\infty, d = \infty$ obtenemos la ley de probabilidad total

$$P(a < X < b) = \int_{-\infty}^{\infty} \left(\int_a^b f_{X|Y}(x|y) dx \right) f_Y(y) dy.$$

Ejemplo 1.14

Sea (X, Y) un vector gaussiano de dimensión 2, de densidad

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{x^2}{\sigma_X^2} - 2\rho\frac{xy}{\sigma_X\sigma_Y} + \frac{y^2}{\sigma_Y^2}\right)\right\} \quad (1.23)$$

La variable X es gaussiana, centrada, de varianza σ_X^2 y densidad

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_X} \exp\left\{-\frac{x^2}{2\sigma_X^2}\right\}. \quad (1.24)$$

La densidad condicional de Y dado que $X = x$ es, por lo tanto,

$$f_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi}\sigma_Y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2\sigma_Y^2(1-\rho^2)}\left(y - \frac{\sigma_Y}{\sigma_X}\rho x\right)^2\right\}$$

que es una densidad gaussiana de media $\rho x \sigma_Y / \sigma_X$ y varianza $\sigma_Y^2(1-\rho^2)$. ▲

Si g es una función para la cual $E[|g(X)|] < \infty$, la esperanza condicional de $g(X)$ dado que $Y = y$ se define como

$$E[g(X)|Y = y] = \int g(x)f_{X|Y}(x|y) dx \quad \text{si } f_Y(y) > 0.$$

Estas esperanzas condicionales satisfacen también las propiedades 1-5 que listamos anteriormente. La propiedad 6 es en este caso,

$$E[g(X)h(Y)] = E[h(Y) E[g(X)|Y]] = \int h(y) E[g(X)|Y = y]f_Y(y) dy$$

válida para cualquier h acotada y suponiendo $E[|g(X)|] < \infty$. Cuando $h \equiv 1$ obtenemos

$$E[g(X)] = E[E[g(X)|Y]] = \int E[g(X)|Y = y]f_Y(y) dy.$$

Ejemplo 1.15

Si (X, Y) es un vector gaussiano bidimensional cuya densidad está dada por (1.23), entonces

$$E[Y|X = x] = \int_{-\infty}^{\infty} yf_{Y|X}(y|x) dy$$

es la esperanza condicional de Y dado que $X = x$. A partir de (1.24) vemos que la densidad condicional es gaussiana de media $\rho x \sigma_Y / \sigma_X$, de donde

$$E[Y|X] = \frac{\sigma_Y}{\sigma_X} \rho X.$$

▲

Podemos reunir los casos discreto y continuo en una sola expresión:

$$E[g(X)h(Y)] = E[h(Y) E[g(X)|Y]] = \int h(y) E[g(X)|Y = y]dF_Y(y)$$

y

$$E[g(X)] = E[E[g(X)|Y]] = \int E[g(X)|Y = y]dF_Y(y).$$

1.9.3. El Caso Mixto

Hasta ahora hemos considerado dos casos para el vector (X, Y) : ambas variables discretas o ambas continuas. En esta sección consideraremos los dos casos mixtos posibles. Comenzamos por el caso continuo-discreto. En ambos casos supondremos, sin pérdida de generalidad, que la variable discreta toma valores en los enteros positivos.

Caso 1: Continuo-Discreto

Sean X y N v.a. con distribución conjunta donde N toma valores $0, 1, 2, \dots$. La función de distribución condicional $F_{X|N}(x|n)$ de X dado que $N = n$ es

$$F_{X|N}(x|n) = \frac{P(X \leq x, N = n)}{P(N = n)} \quad \text{si } P(N = n) > 0,$$

y la función de distribución condicional no está definida para otros valores de n . Es sencillo verificar que $F_{X|N}(x|n)$ es una f.d. en x para cada valor fijo de n para el cual esté definida.

Supongamos que X es continua y $F_{X|N}(x|n)$ es diferenciable en x para todo n con $P(N = n) > 0$. Definimos la densidad condicional de X dado $N = n$ por

$$f_{X|N}(x|n) = \frac{d}{dx} F_{X|N}(x|n).$$

De nuevo, $f_{X|N}(x|n)$ es una densidad en x para los n para los cuales está definida y tiene las propiedades que uno esperaría, por ejemplo

$$P(a \leq X \leq b, N = n) = \int_a^b f_{X|N}(x|n) p_N(n) dx, \quad \text{para } a < b.$$

Usando la ley de la probabilidad total obtenemos la densidad marginal de X ,

$$f_X(x) = \sum_n f_{X|N}(x|n) p_N(n).$$

Supongamos que g es una función para la cual $E[|g(X)|] < \infty$. La esperanza condicional de $g(X)$ dado que $N = n$ se define por

$$E[g(X)|N = n] = \int g(x) f_{X|N}(x|n) dx.$$

Esta esperanza condicional satisface las propiedades anteriores y en este caso la ley de la probabilidad total es

$$E[g(X)] = \sum_n E[g(X)|N = n] p_N(n) = E[E[g(X)|N]].$$

Caso 2: Discreto-Continuo

Consideremos ahora un vector (N, X) . Supongamos que X tiene una distribución continua de densidad $f_X(x)$ y, dado el valor x de X , N es discreta con función de probabilidad $p_{N|X}(n|x)$ para $n \geq 0$. Podemos pensar que X es un parámetro (aleatorio) de la distribución de N , y una vez conocido el valor de este parámetro la distribución de N está completamente determinada.

La función de probabilidad condicional de N dado X es

$$p_{N|X}(n|x) = P(N = n|X = x) = \frac{f_{N,X}(n, x)}{f_X(x)}$$

siempre que $f_X(x) > 0$, donde $f_{N,X}(n, x)$ es la densidad de probabilidad conjunta del vector (N, X) . La función de distribución condicional correspondiente es

$$F_{N|X}(n, x) = \frac{1}{f_X(x)} \sum_{k=0}^n P(N = k|X = x) = \frac{1}{f_X(x)} \sum_{k=0}^n p_{N|X}(k|x)$$

Ejemplo 1.16

Suponemos que $X \sim \text{Bin}(p, N)$ con $p \sim \mathcal{U}[0, 1]$. ¿Cuál es la distribución de X ?

$$\begin{aligned} P(X = k) &= \int_{\mathbb{R}} P(X = k|p = \xi) f_p(\xi) d\xi \\ &= \int_0^1 \frac{N!}{k!(N-k)!} \xi^k (1-\xi)^{N-k} d\xi \\ &= \frac{N!}{k!(N-k)!} \frac{1}{(N+1)!} = \frac{1}{N+1}, \quad k = 0, \dots, N. \end{aligned}$$

es decir, X tiene distribución uniforme en los enteros $0, 1, \dots, N$. ▲

Ejemplo 1.17

Sea $Y \sim \text{Exp}(\theta)$ y dado $Y = y$, X tiene distribución de Poisson de media y . Queremos hallar la ley de X .

Usando la ley de probabilidad total

$$\begin{aligned} P(X = k) &= \int_0^\infty P(X = k|Y = y) f_Y(y) dy \\ &= \int_0^\infty \frac{y^k e^{-y}}{k!} \theta e^{-\theta y} dy \\ &= \frac{\theta}{k!} \int_0^\infty y^k e^{-(1+\theta)y} dy \\ &= \frac{\theta}{k!(1+\theta)^{k+1}} \int_0^\infty u^k e^{-u} du \\ &= \frac{\theta}{(1+\theta)^{k+1}}, \quad k = 0, 1, \dots \end{aligned}$$

▲

1.9.4. Sumas Aleatorias

Con frecuencia encontramos sumas de la forma $T = X_1 + \dots + X_N$, donde el número de sumandos es una variable aleatoria. Consideremos una sucesión X_1, X_2, \dots de v.a.i.i.d. y sea N una v.a. discreta, independiente de X_1, X_2, \dots con densidad $p_N(n) = P(N = n)$, $n = 0, 1, \dots$. Definimos la suma aleatoria T como

$$T = \begin{cases} 0 & \text{si } N = 0, \\ X_1 + \dots + X_N & \text{si } N > 0. \end{cases}$$

Ejemplos 1.18

- Colas: N representa el número de clientes, X_i es el tiempo de atención de cada cliente, T es el tiempo total de atención.
- Seguros: N representa el número de reclamos en un período de tiempo dado, X_i es el monto de cada reclamo y T es el monto total de los reclamos en el período.
- Población: N representa el número de plantas, X_i es el número de semillas de cada planta, T es el total de semillas.
- Biometría: N es el tamaño de la población, X_i es el peso de cada ejemplar y T representa el peso total de la muestra.

Momentos de una Suma Aleatoria

Supongamos que X_k y N tienen momentos finitos

$$\begin{aligned} E[X_k] &= \mu, & \text{Var}[X_k] &= \sigma^2, \\ E[N] &= \nu, & \text{Var}[N] &= \tau^2. \end{aligned}$$

y queremos determinar media y varianza de $T = X_1 + \dots + X_N$. Veamos que

$$E[T] = \mu\nu, \quad \text{Var}[T] = \nu\sigma^2 + \mu^2\tau^2.$$

Tenemos

$$\begin{aligned}
 E[T] &= \sum_{n=0}^{\infty} E[T|N = n]p_N(n) = \sum_{n=0}^{\infty} E[X_1 + \cdots + X_N|N = n]p_N(n) \\
 &= \sum_{n=0}^{\infty} E[X_1 + \cdots + X_n|N = n]p_N(n) = \sum_{n=0}^{\infty} E[X_1 + \cdots + X_n]p_N(n) \\
 &= \sum_{n=0}^{\infty} n\mu p_N(n) = \mu\nu.
 \end{aligned}$$

Para determinar la varianza comenzamos por

$$\begin{aligned}
 \text{Var}[T] &= E[(T - \mu\nu)^2] = E[(T - N\mu + N\mu - \nu\mu)^2] \\
 &= E[(T - N\mu)^2] + E[\mu^2(N - \nu)^2] + 2E[\mu(T - N\mu)(N - \nu)].
 \end{aligned}$$

Calculemos cada uno de estos sumandos por separado, el primero es

$$\begin{aligned}
 E[(T - N\mu)^2] &= \sum_{n=0}^{\infty} E[(T - N\mu)^2|N = n]p_N(n) \\
 &= \sum_{n=1}^{\infty} E[(X_1 + \cdots + X_n - n\mu)^2|N = n]p_N(n) \\
 &= \sum_{n=1}^{\infty} E[(X_1 + \cdots + X_n - n\mu)^2]p_N(n) \\
 &= \sigma^2 \sum_{n=1}^{\infty} np_N(n) = \nu\sigma^2.
 \end{aligned}$$

Para el segundo tenemos

$$E[\mu^2(N - \nu)^2] = \mu^2 E[(N - \nu)^2] = \mu^2\tau^2$$

y finalmente el tercero es

$$\begin{aligned}
 E[\mu(T - N\mu)(N - \mu)] &= \mu \sum_{n=0}^{\infty} E[(T - n\mu)(n - \nu)|N = n]p_N(n) \\
 &= \mu \sum_{n=0}^{\infty} (n - \nu) E[(T - n\mu)|N = n]p_N(n) \\
 &= 0.
 \end{aligned}$$

La suma de estos tres términos demuestra el resultado.

Distribución de una Suma Aleatoria

Supongamos que los sumandos X_1, X_2, \dots son v.a.i. continuas con densidad de probabilidad $f(x)$. Para $n \geq 1$ fijo, la densidad de la suma $X_1 + \cdots + X_n$ es la n -ésima convolución de la densidad $f(x)$, que denotaremos por $f^{(n)}(x)$ y definiremos recursivamente por

$$\begin{aligned}
 f^{(1)}(x) &= f(x), \\
 f^{(n)}(x) &= \int f^{(n-1)}(x-u)f(u) du \quad \text{para } n > 1.
 \end{aligned}$$

Como N y X_1, X_2, \dots son independientes, $f^{(n)}(x)$ es también la densidad condicional de $T = X_1 + \dots + X_N$ dado que $N = n \geq 1$.

Supongamos que $P(N = 0) = 0$, es decir, que la suma aleatoria siempre tiene al menos un sumando. Por la ley de la probabilidad total, T es continua y tiene densidad marginal

$$f_T(x) = \sum_{n=1}^{\infty} f^{(n)}(x) p_N(n).$$

Observación 1.1 Si N puede valer 0 con probabilidad positiva entonces $T = X_1 + \dots + X_N$ es una v.a. mixta, es decir, tiene componentes discreta y continua. Si suponemos que X_1, X_2, \dots son continuas con densidad $f(x)$, entonces

$$P(T = 0) = P(N = 0) = p_N(0)$$

mientras que para $0 < a < b$ ó $a < b < 0$,

$$P(a < T < b) = \int_a^b \left(\sum_{n=1}^{\infty} f^{(n)}(x) p_N(n) \right) dx$$

▲

Ejemplo 1.19 (Suma Geométrica de Variables Exponenciales)

Supongamos que

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

$$p_N(n) = \beta(1 - \beta)^{n-1} \quad n = 1, 2, \dots$$

Comenzamos por hallar la convolución de las densidades exponenciales

$$\begin{aligned} f^{(2)}(x) &= \int f(x-u)f(u) du = \int \mathbf{1}_{\{x-u \geq 0\}}(u) \lambda e^{-\lambda(x-u)} \mathbf{1}_{\{u \geq 0\}}(u) \lambda e^{-\lambda u} du \\ &= \lambda^2 e^{-\lambda x} \int_0^x du = x \lambda^2 e^{-\lambda x} \end{aligned}$$

para $x \geq 0$. La siguiente convolución es

$$\begin{aligned} f^{(3)}(x) &= \int f^{(2)}(x-u)f(u) du = \int \mathbf{1}_{\{x-u \geq 0\}}(u) \lambda^2 (x-u) e^{-\lambda(x-u)} \mathbf{1}_{\{u \geq 0\}}(u) \lambda e^{-\lambda u} du \\ &= \lambda^3 e^{-\lambda x} \int_0^x (x-u) du = \frac{x^2}{2} \lambda^3 e^{-\lambda x} \end{aligned}$$

para $x \geq 0$. Procediendo inductivamente obtenemos que

$$f^{(n)}(x) = \frac{x^{n-1}}{(n-1)!} \lambda^n e^{-\lambda x}$$

La densidad de $T = X_1 + \dots + X_N$ es

$$\begin{aligned} f_T(t) &= \sum_{n=1}^{\infty} f^{(n)}(t) p_N(n) = \sum_{n=1}^{\infty} \frac{\lambda^n}{(n-1)!} t^{n-1} e^{-\lambda t} \beta (1 - \beta)^{n-1} \\ &= \lambda \beta e^{-\lambda t} \sum_{n=1}^{\infty} \frac{(\lambda(1 - \beta)t)^{n-1}}{(n-1)!} = \lambda \beta e^{-\lambda t} e^{\lambda(1-\beta)t} \\ &= \lambda \beta e^{-\lambda \beta t} \end{aligned}$$

para $t \geq 0$, y por lo tanto $T \sim \text{Exp}(\lambda\beta)$. ▲

1.10. Funciones Generadoras de Probabilidad

Consideremos una v.a. ξ con valores enteros positivos y distribución de probabilidad

$$P(\xi = k) = p_k, \quad k = 0, 1, \dots$$

La función generadora de probabilidad (f.g.p.) $\phi(s)$ asociada a la v.a. ξ (o equivalentemente a su distribución (p_k)) se define por

$$\phi(s) = E[s^\xi] = \sum_{k=0}^{\infty} s^k p_k, \quad 0 \leq s \leq 1. \quad (1.25)$$

A partir de la definición es inmediato que si ϕ es una f.g.p. entonces

$$\phi(1) = \sum_{k=0}^{\infty} p_k = 1.$$

Resultados Fundamentales:

1. La relación entre funciones de probabilidad y funciones generadoras es 1-1. Es posible obtener las probabilidades (p_k) a partir de ϕ usando la siguiente fórmula

$$p_k = \frac{1}{k!} \left. \frac{d^k \phi(s)}{ds^k} \right|_{s=0}. \quad (1.26)$$

Por ejemplo,

$$\phi(s) = p_0 + p_1 s + p_2 s^2 + \dots \Rightarrow p_0 = \phi(0)$$

$$\frac{d\phi(s)}{ds} = p_1 + 2p_2 s + 3p_3 s^2 + \dots \Rightarrow p_1 = \left. \frac{d\phi(s)}{ds} \right|_{s=0}$$

2. Si ξ_1, \dots, ξ_n son v.a.i. con funciones generadoras $\phi_1(s), \phi_2(s), \dots, \phi_n(s)$ respectivamente, la f. g. p. de su suma $X = \xi_1 + \xi_2 + \dots + \xi_n$ es el producto de las funciones generadoras respectivas

$$\phi_X(s) = \phi_1(s)\phi_2(s) \cdots \phi_n(s). \quad (1.27)$$

3. Los momentos de una variable que toma valores en los enteros no-negativos se pueden obtener derivando la función generadora:

$$\frac{d\phi(s)}{ds} = p_1 + 2p_2 s + 3p_3 s^2 + \dots,$$

por lo tanto

$$\left. \frac{d\phi(s)}{ds} \right|_{s=1} = p_1 + 2p_2 + 3p_3 + \dots = E[\xi]. \quad (1.28)$$

Para la segunda derivada tenemos

$$\frac{d^2\phi(s)}{ds^2} = 2p_2 + 3 \cdot 2p_3 s + 4 \cdot 3p_4 s^2 + \dots,$$

evaluando en $s = 1$,

$$\begin{aligned} \left. \frac{d^2 \phi(s)}{ds^2} \right|_{s=1} &= 2p_2 + 3 \cdot 2p_3 + 4 \cdot 3p_4 \cdots \\ &= \sum_{k=2}^{\infty} k(k-1)p_k \\ &= E[\xi(\xi-1)] = E[\xi^2] - E[\xi] \end{aligned} \quad (1.29)$$

de modo que

$$E[\xi^2] = \left. \frac{d^2 \phi(s)}{ds^2} \right|_{s=1} + E[\xi] = \left. \frac{d^2 \phi(s)}{ds^2} \right|_{s=1} + \left. \frac{d\phi(s)}{ds} \right|_{s=1},$$

y en consecuencia

$$\text{Var}[\xi] = E[\xi^2] - (E[\xi])^2 = \left. \frac{d^2 \phi(s)}{ds^2} \right|_{s=1} + \left. \frac{d\phi(s)}{ds} \right|_{s=1} - \left(\left. \frac{d\phi(s)}{ds} \right|_{s=1} \right)^2.$$

Ejemplo 1.20

Supongamos que $\xi \sim \mathcal{Pois}(\lambda)$:

$$p_k = P(\xi = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, \dots$$

Su función generadora de probabilidad es

$$\begin{aligned} \phi(s) = E[s^\xi] &= \sum_{k=0}^{\infty} s^k \frac{\lambda^k}{k!} e^{-\lambda} \\ &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(s\lambda)^k}{k!} = e^{-\lambda} e^{\lambda s} \\ &= e^{-\lambda(1-s)} \end{aligned}$$

Entonces,

$$\left. \frac{d\phi(s)}{ds} \right|_{s=1} = \lambda e^{-\lambda(1-s)}, \quad \left. \frac{d\phi(s)}{ds} \right|_{s=1} = \lambda \quad (1.30)$$

$$\left. \frac{d^2 \phi(s)}{ds^2} \right|_{s=1} = \lambda^2 e^{-\lambda(1-s)}, \quad \left. \frac{d^2 \phi(s)}{ds^2} \right|_{s=1} = \lambda^2 \quad (1.31)$$

y obtenemos

$$E[\xi] = \lambda, \quad \text{Var}(\xi) = \lambda^2 + \lambda - (\lambda)^2 = \lambda.$$

▲

1.10.1. Funciones Generadoras de Probabilidad y Sumas de V. A. I.

Sean ξ, η v.a.i. con valores $0, 1, 2, \dots$ y con funciones generadoras de probabilidad

$$\phi_\xi(s) = E[s^\xi], \quad \phi_\eta(s) = E[s^\eta], \quad |s| < 1,$$

entonces la f.g.p. de la suma $\xi + \eta$ es

$$\phi_{\xi+\eta}(s) = E[s^{\xi+\eta}] = E[s^\xi s^\eta] = E[s^\xi] E[s^\eta] = \phi_\xi(s) \phi_\eta(s) \quad (1.32)$$

El recíproco también es cierto, si $\phi_{\xi+\eta}(s) = \phi_\xi(s)\phi_\eta(s)$ entonces las variables ξ y η son independientes.

Como consecuencia, si $\xi_1, \xi_2, \dots, \xi_m$ son v.a.i.i.d. con valores en $\{0, 1, 2, \dots\}$ y f.g.p. $\phi(s) = E[s^\xi]$ entonces

$$E[s^{\xi_1 + \dots + \xi_m}] = \phi^m(s) \quad (1.33)$$

¿Qué ocurre si el número de sumandos es aleatorio?

Proposición 1.3 *Sea N una v.a. con valores enteros no-negativos e independiente de ξ_1, ξ_2, \dots con f.g.p. $g_N(s) = E[s^N]$ y consideremos la suma*

$$X = \xi_1 + \dots + \xi_N.$$

Sea $h_X(s) = E[s^X]$ la f.g.p. de X . Entonces

$$h_X(s) = g_N(\phi(s)). \quad (1.34)$$

Demostración.

$$\begin{aligned} h_X(s) &= \sum_{k=0}^{\infty} P(X = k) s^k \\ &= \sum_{k=0}^{\infty} \left(\sum_{n=0}^{\infty} P(X = k | N = n) P(N = n) \right) s^k \\ &= \sum_{k=0}^{\infty} \left(\sum_{n=0}^{\infty} P(\xi_1 + \dots + \xi_n = k | N = n) P(N = n) \right) s^k \\ &= \sum_{k=0}^{\infty} \left(\sum_{n=0}^{\infty} P(\xi_1 + \dots + \xi_n = k) P(N = n) \right) s^k \\ &= \sum_{n=0}^{\infty} \left(\sum_{k=0}^{\infty} P(\xi_1 + \dots + \xi_n = k) s^k \right) P(N = n) \\ &= \sum_{n=0}^{\infty} \phi^n(s) P(N = n) = g_N(\phi(s)) \end{aligned}$$

■

Ejemplo 1.21

Sea N una variable aleatoria con distribución de Poisson de parámetro λ . Dado el valor de N , realizamos N experimentos de Bernoulli con probabilidad de éxito p y llamamos X al número de éxitos. En este caso ξ_i tiene distribución de Bernoulli y su f.g.p. es

$$\phi_\xi(s) = E[s^\xi] = sp + q$$

mientras que $N \sim \text{Pois}(\lambda)$ con f.g.p.

$$g_N(s) = E[s^N] = e^{-\lambda(1-s)}$$

según vimos en el ejemplo 1.20. Por la proposición anterior obtenemos que la f.g.p. de X es

$$h_X(s) = g_N(\phi_\xi(s)) = g_N(q + sp) = \exp \left\{ -\lambda(1 - q - sp) \right\} = \exp \left\{ -\lambda p(1 - s) \right\}$$

que es la f.g.p. de una distribución de Poisson de parámetro λp .

▲

1.11. Funciones Generadoras de Momentos.

Dada una variable aleatoria X , o su función de distribución F , vamos a definir otra función generadora, como

$$M_X(t) = E(e^{tX}).$$

siempre que este valor esperado exista.

Notemos que cuando el recorrido de X son los enteros no-negativos, $M_X(t) = \phi_X(e^t)$. Si X está acotada, M_X está bien definida para todo t real; en cambio, si X no está acotada, es posible que el dominio de M no sea el conjunto de todos los reales. En todo caso, p siempre está definida en cero, y $M(0) = 1$.

Si la función M está definida en un entorno de $t = 0$, entonces la serie

$$M_X(t) = E(e^{tX}) = E\left(1 + \sum_{n=1}^{\infty} \frac{t^n X^n}{n!}\right) = 1 + \sum_{n=1}^{\infty} \frac{t^n}{n!} E(X^n)$$

es convergente y en consecuencia se puede derivar término a término. Obtenemos

$$M'_X(0) = E(X); \quad M''_X(0) = E(X^2) \quad \text{y en general } M_X^{(n)}(0) = E(X^n).$$

Es por esta última propiedad que esta función se conoce como *función generadora de momentos* (f.g.m.).

Ejemplos 1.22

1. Si $X \sim \text{Bin}(n, p)$ veamos que $M(t) = (pe^t + 1 - p)^n$: Un cálculo directo muestra que

$$M(t) = \sum_{j=0}^n e^{jt} \binom{n}{j} p^j (1-p)^{n-j} = (pe^t + 1 - p)^n,$$

2. Si $X \sim \text{Exp}(\lambda)$, es decir, si $P(X \leq x) = 1 - e^{-\lambda x}$, para $x \geq 0$, entonces $M(t) = \lambda/(\lambda - t)$ para $t \leq \lambda$.

El resultado se obtiene a partir del cálculo

$$M(t) = \int_0^{\infty} \lambda e^{-\lambda x} e^{tx} dx = \lambda \left. \frac{e^{(t-\lambda)x}}{t-\lambda} \right|_0^{\infty} = \frac{\lambda}{\lambda - t}.$$

Observamos que en este caso, $M(t)$ no está definida si $t \geq \lambda$.

3. Si $X \sim \mathcal{N}(0, 1)$, es decir, si $P(X \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-x^2/2} dx$, entonces $M(t) = e^{t^2/2}$.

Calculemos

$$M(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-t)^2} e^{t^2/2} dx = e^{t^2/2}$$

ya que $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-t)^2} dx = 1$ puesto que el integrando es la densidad de una variable aleatoria con distribución $\mathcal{N}(t, 1)$

Observación 1.2 Por la forma en la cual hemos definido la función generadora de momentos, cuando las f.g.m. de dos variables aleatorias X_1, X_2 coinciden para todos los valores de t en un entorno de $t = 0$, entonces las distribuciones de probabilidad de X_1 y X_2 deben ser idénticas. Este resultado lo enunciamos en el próximo teorema, sin demostración

Teorema 1.5 Si X tiene función generadora de momentos $M(t)$ que está definida en un entorno $(-a, a)$ de 0, entonces $M(t)$ caracteriza a la distribución de X , es decir, si otra variable Y tiene la misma función generadora de momentos, las distribuciones de X e Y coinciden.

La función generadora de momentos resulta particularmente útil cuando consideramos sucesiones de variables aleatorias, como lo muestra el siguiente teorema que enunciamos sin demostración.

Teorema 1.6 (de Continuidad) Sea $F_n(x)$, $n \geq 1$ una sucesión de f.d. con funciones generadoras de momento respectivas $M_n(t)$, $n \geq 1$, que están definidas para $|t| < b$. Supongamos que cuando $n \rightarrow \infty$, $M_n(t) \rightarrow M(t)$ para $|t| \leq a < b$, donde $M(t)$ es la función generadora de momentos de la distribución $F(x)$. Entonces $F_n(x) \rightarrow F(x)$ cuando $n \rightarrow \infty$ para todo punto x en el cual F es continua.

Veamos una aplicación del teorema anterior para demostrar el Teorema de de Moivre y Laplace.

Teorema 1.7 (de Moivre-Laplace) Sea $S_n \sim \text{Bin}(n, p)$ para $n \geq 1$ y $q = 1 - p$. Definimos

$$T_n = \frac{S_n - np}{(npq)^{1/2}}$$

Entonces para todo $x \in \mathbb{R}$,

$$P(T_n \leq x) \rightarrow \Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

Demostración. Recordemos que S_n es la suma de n v.a.i. con distribución de Bernoulli de parámetro p : $S_n = \sum_{i=1}^n X_i$. Usamos esto para calcular la función generadora de momentos de T_n .

$$\begin{aligned} E(e^{tT_n}) &= E \left[\exp \left(\frac{t(S_n - np)}{(npq)^{1/2}} \right) \right] = E \left[\exp \left(\frac{t(\sum_{i=1}^n (X_i - p))}{(npq)^{1/2}} \right) \right] \\ &= E \left[\prod_{i=1}^n \exp \left(\frac{t(X_i - p)}{(npq)^{1/2}} \right) \right] = \prod_{i=1}^n E \left[\exp \left(\frac{t(X_i - p)}{(npq)^{1/2}} \right) \right] \\ &= \left(E \left[\exp \left(\frac{t(X_1 - p)}{(npq)^{1/2}} \right) \right] \right)^n \\ &= \left(p \exp \left(\frac{t(1-p)}{(npq)^{1/2}} \right) + q \exp \left(\frac{-pt}{(npq)^{1/2}} \right) \right)^n. \end{aligned} \quad (1.35)$$

Ahora hacemos un desarrollo de Taylor para las dos exponenciales que aparecen en esta última expresión para obtener

$$p \exp \left(\frac{t(1-p)}{(npq)^{1/2}} \right) = p \left(1 + \frac{qt}{(npq)^{1/2}} + \frac{q^2 t^2}{2npq} + \frac{C_1 q^3 t^3}{3!(npq)^{3/2}} \right) \quad (1.36)$$

$$q \exp \left(\frac{-pt}{(npq)^{1/2}} \right) = q \left(1 - \frac{pt}{(npq)^{1/2}} + \frac{p^2 t^2}{2npq} + \frac{C_2 p^3 t^3}{3!(npq)^{3/2}} \right). \quad (1.37)$$

La suma de estas dos expresiones nos da $1 + \frac{t^2}{2n} + O(n^{-3/2})$ y sustituyendo en (1.35) obtenemos

$$E(e^{tT_n}) = \left(1 + \frac{t^2}{2n} + O(n^{-3/2}) \right)^n \rightarrow e^{t^2/2}$$

que es la f.g.m. de la distribución normal típica. ■

1.12. Simulación de Variables Aleatorias

Los generadores de números aleatorios simulan valores de la distribución $\mathcal{U}[0, 1]$, pero con frecuencia nos interesa simular valores de otras distribuciones. Vamos a estudiar en esta sección dos métodos para generar valores a partir de una función de distribución F .

1.12.1. Método de la Distribución Inversa

Este método se basa en el siguiente resultado:

Proposición 1.4 *Sea X una variable aleatoria con función de distribución F_X y sea g una función continua y estrictamente creciente. Definimos $Y = g(X)$ y sea F_Y la función de distribución de esta variable. Entonces*

$$F_Y(y) = F_X(g^{-1}(y)). \quad (1.38)$$

Demostración. Como g es estrictamente creciente los eventos $\{X \leq g^{-1}(y)\}$ y $\{g(X) \leq y\}$ son iguales. Por lo tanto,

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \leq g^{-1}(y)) = F_X(g^{-1}(y))$$

■

Si g es estrictamente decreciente entonces $F_Y(y) = 1 - F_X(g^{-1}(y))$.

Corolario 1.3 *Sea F una función de distribución continua y estrictamente creciente para los x tales que $0 < F(x) < 1$ y sea $U \sim \mathcal{U}[0, 1]$. Entonces la variable $Z = F^{-1}(U)$ tiene distribución F .*

Demostración. La función de distribución de U es $F_U(u) = u$ para $u \in [0, 1]$. Entonces

$$F_Z(z) = F_U(F(z)) = F(z) \quad (1.39)$$

de modo que Z tiene función de distribución F .

■

Observación 1.3 El resultado anterior es cierto en general si utilizamos la *inversa generalizada* F^{\leftarrow} de la función F cuando esta no sea invertible, que se define por la siguiente expresión:

$$F^{\leftarrow}(y) = \inf\{x : F(x) \geq y\}$$

Por lo tanto, para cualquier función de distribución F , la variable aleatoria $Z = F^{\leftarrow}(U)$ tiene función de distribución F . Para ver que esto es cierto observamos que, a partir de la definición, es fácil demostrar que

$$F^{\leftarrow}(y) \leq t \Leftrightarrow y \leq F(t); \quad F^{\leftarrow}(y) > t \Leftrightarrow y > F(t).$$

Usando esto obtenemos

$$F_Z(z) = P(Z \leq z) = P(F^{\leftarrow}(U) \leq z) = P(U \leq F(z)) = F(z).$$

■

El Corolario 1.3 y la Observación 1.3 nos dan un método para simular una variable aleatoria con función de distribución F : Generamos el valor u de una variable uniforme en $[0, 1]$ y evaluamos la inversa generalizada en u : $F^{\leftarrow}(u)$. Sin embargo, dependiendo de la naturaleza de la función de distribución F , es posible que la inversa generalizada tenga una expresión complicada o incluso no sea posible escribirla en términos de funciones elementales, como ocurre en el caso de las variables Gaussianas. Por esta razón hay métodos particulares que resultan más eficientes en muchos casos.

Ejemplos 1.23

1. **Variables Discretas.** Si queremos simular una variable aleatoria finita X con valores x_1, \dots, x_n y probabilidades respectivas p_1, \dots, p_n , podemos dividir el intervalo $[0, 1]$ en subintervalos usando las probabilidades p_i :

$$[0, p_1); \quad [p_1, p_1 + p_2); \quad [p_1 + p_2, p_1 + p_2 + p_3); \quad \dots \quad \left[\sum_{j < n} p_j, 1 \right].$$

Ahora generamos una variable U con distribución uniforme en $[0, 1]$ y si el valor cae en el i -ésimo intervalo le asignamos a X el valor x_i . Como la probabilidad de que U caiga en el intervalo i es igual a la longitud del intervalo, que es p_i , vemos que

$$P(X = x_i) = p_i, \quad \text{para } 1 \leq i \leq n.$$

Esta es una implementación del método de la distribución inversa. Desde el punto de vista computacional es conveniente ordenar los valores según el tamaño de las p_i , colocando estas probabilidades de mayor a menor, porque para identificar el intervalo en cual cae U tenemos que comparar con p_1 , luego con $p_1 + p_2$, y así sucesivamente hasta obtener el primer valor mayor que U . Ordenar las probabilidades hace que se maximice la probabilidad de que U esté en los primeros intervalos, y esto reduce el número de comparaciones que hay que hacer en promedio para obtener el valor de X .

Este método también funciona para variables discretas con una cantidad infinita de valores. La misma observación sobre el ordenamiento de los valores de las probabilidades es válida.

2. **Distribución de Bernoulli.** Un caso particular sencillo es el de la distribución de Bernoulli con probabilidad de éxito p . Para generar un valor de la variable X con esta distribución, generamos U y si $U < p$, $X = 1$ y si no, $X = 0$.
3. **Distribución Uniforme Discreta.** Sea X una variable aleatoria que toma valores $\{x_1, x_2, \dots, x_n\}$ con igual probabilidad. Para simular esta distribución generamos un número aleatorio $U \in (0, 1]$, dividimos el intervalo $[0, 1]$ en n intervalos iguales y le asignamos a la variables el valor x_k si

$$\frac{k-1}{n} < U \leq \frac{k}{n},$$

es decir, el valor de la variable es x_k con $k = \lceil Un \rceil$, donde $\lceil a \rceil$ es la función *techo* y representa el menor entero que es mayor o igual a a .

4. **Variables Continuas.** Si X es una variable continua con función de distribución F invertible, para simular X basta generar una variable uniforme U y poner $X = F^{-1}(U)$. Esto es consecuencia del corolario 1.3. Por ejemplo, si queremos simular una v.a. X con función de distribución $F(x) = x^n$ para $0 < x < 1$, observamos que F es invertible y su inversa es $F^{-1}(u) = u^{1/n}$. Por lo tanto basta generar una variables uniforme U y poner $X = U^{1/n}$.
5. **Distribución Uniforme Continua.** Si queremos simular la distribución $\mathcal{U}[a, b]$ generamos U uniforme en $[0, 1]$ y usamos la transformación $u \mapsto a + u(b - a)$.
6. **Distribución Exponencial.** Si $X \sim \mathcal{Exp}(\lambda)$ su f.d. está dada por $F(x) = 1 - e^{-\lambda x}$. La inversa de esta función es

$$F^{-1}(u) = -\frac{1}{\lambda} \log(1 - u).$$

Por lo tanto para generar X podemos generar una uniforme U y ponemos $X = -\ln(1 - U)/\lambda$. Observamos ahora que si U tiene distribución uniforme en $(0, 1)$, $1 - U$ también. Por lo tanto, para simular esta distribución a partir de una variable $U \sim \mathcal{U}(0, 1)$ basta hacer la transformación $-\ln(U)/\lambda$.

1.12.2. Método de Rechazo

Variables Discretas

Supongamos que tenemos un método eficiente para simular una variable Y que tiene función de probabilidad $\{q_j, j \geq 1\}$. Podemos usar este método como base para simular otra variable X con función de probabilidad diferente $\{p_j, j \geq 1\}$, siempre que las dos variables tengan el mismo conjunto de valores posibles o al menos cuando los valores de X sean un subconjunto de los valores de Y . La idea es simular primero la variable Y y luego aceptar este valor para la variable X con probabilidad proporcional a p_Y/q_Y .

Sea c una constante tal que

$$\frac{p_j}{q_j} \leq c \quad \text{para todo } j \text{ tal que } p_j > 0, \quad (1.40)$$

entonces el algoritmo para el método de rechazo es el siguiente,

Algoritmo.

- Paso 1. Simulamos una variable Y con función de probabilidad q_j .
- Paso 2. Generamos una variable uniforme U .
- Paso 3. Si $U < p_Y/cq_Y$, ponemos $X = Y$ y paramos. Si no, regresamos al paso 1.

Veamos que este método efectivamente produce una variable con distribución p_j . Calculemos primero la probabilidad de obtener el valor j en una sola iteración:

$$\begin{aligned} P(Y = j \text{ y este valor sea aceptado}) &= P(Y = j)P(\text{Aceptar}|Y = j) \\ &= q_j P(U < \frac{p_j}{cq_j}) \\ &= q_j \frac{p_j}{cq_j} = \frac{p_j}{c}. \end{aligned}$$

Si sumamos ahora sobre los valores posibles j obtenemos la probabilidad de que el valor de la variable generada sea aceptado:

$$P(\text{Aceptar el valor de } Y) = \sum_j \frac{p_j}{c} = \frac{1}{c},$$

Es decir, cada interacción resulta en un valor que es aceptado con probabilidad $1/c$ y esto ocurre de manera independiente, de modo que la distribución del número de iteraciones necesarias para aceptar un valor es geométrica con parámetro $1/c$. En consecuencia

$$\begin{aligned} P(X = j) &= \sum_n P(j \text{ es aceptado en la iteración } n) \\ &= \sum_n \left(1 - \frac{1}{c}\right)^{n-1} \frac{p_j}{c} = p_j. \end{aligned}$$

Como el número de iteraciones es geométrico con parámetro $1/c$, en promedio es necesario realizar c iteraciones para aceptar un valor. Por lo tanto conviene escoger c lo más pequeño posible, siempre que satisfaga (1.40).

Ejemplo 1.24

Supongamos que queremos generar una variable aleatoria con la siguiente distribución: $P(X = j) = p_j$ para $j = 1, 2, 3, 4$ y $p_1 = 0.20, p_2 = 0.15, p_3 = 0.25, p_4 = 0.4$ usando el método de rechazo. Vamos a usar una variable Y con distribución uniforme sobre los valores 1, 2, 3, 4 y por lo tanto podemos tomar

$$c = \max\left\{\frac{p_j}{q_j} : 1 \leq j \leq 4\right\} = \frac{0.4}{0.25} = 1.6$$

y utilizar el algoritmo descrito anteriormente. En este caso en promedio hacemos 1.6 iteraciones por cada valor aceptado para la variable que queremos generar.

Variables Continuas

Este método funciona exactamente igual que en el caso discreto. Supongamos que tenemos una manera eficiente de generar una variable aleatoria con densidad $g(x)$ y queremos generar otra variable que tiene densidad $f(x)$ con el mismo conjunto de valores posibles. Podemos hacer esto generando Y con distribución g y luego aceptando este valor con probabilidad proporcional a $f(Y)/g(Y)$.

Sea c una constante tal que

$$\frac{f(y)}{g(y)} \leq c \quad \text{para todo } y,$$

entonces tenemos el siguiente algoritmo para generar una variable con densidad f .

Algoritmo.

- Paso 1. Generamos Y con densidad g .
- Paso 2. Generamos un número aleatorio U .
- Paso 3. Si $U \leq \frac{f(Y)}{cg(Y)}$ ponemos $X = Y$ y paramos. Si no, volvemos al paso 1.

Al igual que en caso discreto tenemos el siguiente resultado que justifica el método y que presentamos sin demostración.

Teorema

- (i) La variable generada con el método del rechazo tiene densidad f .
- (ii) El número de iteraciones necesarias en el algoritmo es una variable geométrica con media c .

Ejemplo 1.25

Vamos a usar el método de rechazo para generar una variable aleatoria con densidad

$$f(x) = 20x(1-x)^3, \quad 0 < x < 1.$$

Como esta variable aleatoria está concentrada en el intervalo $(0, 1)$, usaremos el método de rechazo con la distribución uniforme

$$g(x) = 1, \quad 0 < x < 1.$$

Para determinar la menor constante c que satisface $f(x)/g(x) < c$ para todo $x \in (0, 1)$ calculamos el máximo de

$$\frac{f(x)}{g(x)} = 20x(1-x)^3.$$

Derivando esta expresión e igualando a cero obtenemos la ecuación

$$20[(1-x)^3 - 3x(1-x)^2] = 0$$

con soluciones 1 y $1/4$. Esta última solución corresponde al máximo y por lo tanto

$$\frac{f(1/4)}{g(1/4)} = 20 \frac{1}{4} \left(\frac{3}{4}\right)^3 = \frac{135}{64} \equiv c.$$

En consecuencia

$$\frac{f(x)}{cg(x)} = \frac{256}{27} x(1-x)^3$$

y el algoritmo es

Algoritmo.

- Paso 1. Generamos dos números aleatorios U_1 y U_2 .
- Paso 2. Si $U_2 \leq 256U_1(1 - U_1)^3/27$ ponemos $X = U_1$ y paramos. Si no, volvemos al paso 1.

En promedio, el paso 1 se realiza $c = \frac{256}{27} \approx 2.11$ veces por cada número generado.

1.12.3. Métodos Particulares

Distribución Binomial

Una manera sencilla de simular una variable con distribución binomial de parámetros n y p es generar n variables de Bernoulli con probabilidad de éxito p y sumarlas. Esto resulta un poco pesado si n es grande, pero en este caso podemos usar el Teorema Central del Límite, (teorema 1.7).

Otra posibilidad es usar el método de la transformada inversa junto con la siguiente relación iterativa para la distribución binomial:

$$\frac{P(S_n = i + 1)}{P(S_n = i)} = \frac{n!i!(n - i)!}{(i + 1)!(n - i - 1)!n!} \frac{p^{i+1}(1 - p)^{n-i-1}}{p^i(1 - p)^{n-i}} = \frac{n - i}{i + 1} \frac{p}{1 - p},$$

es decir,

$$P(S_n = i + 1) = \frac{n - i}{i + 1} \frac{p}{1 - p} P(S_n = i).$$

En consecuencia, generamos una variable uniforme U y comparamos con $P(X = 0) = (1 - p)^n$. Si U es menor que este valor ponemos $X = 0$, en caso contrario multiplicamos $P(X = 0)$ por $pn/(1 - p)$ para obtener $P(X = 1)$ y comparamos. Si U es menor que este valor ponemos $X = 1$, en caso contrario repetimos el procedimiento hasta conseguir el valor de X . El algoritmo se puede describir como sigue:

- Paso 1: Generamos una variable uniforme U .
- Paso 2: Ponemos $a = p/(1 - p)$; $b = (1 - p)^n$; $c = b$; $i = 0$.
- Paso 3: Si $U < c$ ponemos $X = i$ y paramos.
- Paso 4: $b = ab(n - i)/(i + 1)$; $c = c + b$; $i = i + 1$.
- Paso 5: Vamos al paso 3.

Distribución de Poisson

Al igual que para la distribución binomial, tenemos una relación recursiva para la función de probabilidad que permite aplicar el método de la transformada inversa para generar la distribución de Poisson:

$$P(X = i + 1) = \frac{\lambda}{i + 1} P(X = i),$$

que es sencilla de demostrar. El algoritmo es el siguiente:

- Paso 1: Generamos una variable uniforme U .
- Paso 2: Ponemos $a = e^{-\lambda}$; $b = a$; $i = 0$.
- Paso 3: Si $U < b$ ponemos $X = i$ y paramos.
- Paso 4: $a = \lambda a/(i + 1)$; $b = b + a$; $i = i + 1$.
- Paso 5: Vamos al paso 3.

Distribución Geométrica

Una manera de generar variables con distribución geométrica es generar una sucesión de variables de Bernoulli hasta obtener el primer éxito, es decir, generamos una sucesión de números aleatorios en $[0, 1]$ hasta obtener el primero que sea menor que p . Sin embargo, si p es pequeño esto puede ser lento (toma en promedio $1/p$ pasos). Para evitar esto podemos seguir el método alternativo que describimos a continuación. Sea X una v.a. con distribución geométrica de parámetro p , $0 < p < 1$ y sea U un número aleatorio en $[0, 1]$. Definimos Y como el menor entero que satisface la desigualdad $1 - q^Y \geq U$. Entonces

$$\begin{aligned} P(Y = j) &= P(1 - q^j \geq U > 1 - q^{j-1}) \\ &= q^{j-1} - q^j = q^{j-1}(1 - q) = q^{j-1}p, \end{aligned}$$

de modo que Y también tiene una distribución geométrica de parámetro p . Por lo tanto, para generar Y basta resolver la ecuación que la define, es decir,

$$Y = \left\lceil \frac{\log(1 - u)}{\log q} \right\rceil$$

pero como $1 - u$ y u tienen la misma distribución, podemos usar

$$Y = \left\lceil \frac{\log(u)}{\log q} \right\rceil.$$

Distribución Binomial Negativa

Observamos que una variable con distribución binomial negativa de parámetros k y p es la suma de k variables geométricas con parámetro p : una por cada éxito en la sucesión de ensayos. Esta observación es útil para generar variables con esta distribución: si u_j , $j = 1, \dots, k$ son números aleatorios en $[0, 1]$, la siguiente expresión produce el valor de una variable con distribución binomial negativa:

$$\sum_{j=1}^k \left\lceil \frac{\log(u_j)}{\log q} \right\rceil.$$

Distribución Normal

La función de distribución normal Φ no se puede escribir en términos de funciones simples, y lo mismo ocurre con su inversa, lo que dificulta la aplicación del método de la transformada inversa. Sin embargo existen otros métodos y uno de los más populares es el de Box-Muller, también conocido como el método polar.

Aún cuando la justificación del método no es complicada, requiere algunos conceptos que no hemos introducido, así que vamos a describir el método sin demostrar que efectivamente lo que obtenemos es el valor de una variable normal. El algoritmo es el siguiente:

- Paso 1: Generamos variables uniformes U_1 y U_2 .
- Paso 2: Ponemos $V_1 = 2U_1 - 1$; $V_2 = 2U_2 - 1$; $S = V_1^2 + V_2^2$.
- Paso 3: Si $S > 1$ regresamos al paso 1.
- Paso 4: X y Y son variables normales típicas independientes:

$$X = \sqrt{\frac{-2 \log S}{S}} V_1, \quad Y = \sqrt{\frac{-2 \log S}{S}} V_2.$$

1.12.4. Generación de Variables Aleatorias en R

El lenguaje R tiene incorporadas una serie de rutinas para generar variables aleatorias. La sintaxis precisa de la instrucción correspondiente depende de la distribución, pero todas tienen el formato común `rdist`, donde *dist* designa la distribución; por ejemplo, para generar valores a partir de la distribución normal usamos `rnorm`. Según la distribución, puede ser necesario especificar uno o varios parámetros. La tabla que presentamos a continuación presenta las distribuciones más comunes, los parámetros requeridos y sus valores por defecto. `n` representa siempre el tamaño de la muestra.

Distribución	Función en R
Binomial	<code>rbinom(n, size, prob)</code>
Poisson	<code>rpois(n, lambda)</code>
Geométrica	<code>rgeom(n, prob)</code>
Hipergeométrica	<code>rhyper(nm, m, n, k)</code>
Binomial Negativa	<code>rnbinom(n, size, prob)</code>
Multinomial	<code>rmultinom(n, size, prob)</code>
Uniforme	<code>runif(n, min=0, max=1)</code>
Exponencial	<code>rexp(n, rate=1)</code>
Gaussiana	<code>rnorm(n, mean=0, sd=1)</code>
Gamma	<code>rgamma(n, shape, scale=1)</code>
Weibull	<code>rweibull(n, shape, scale=1)</code>
Cauchy	<code>rcauchy(n, location=0, scale=1)</code>
Beta	<code>rbeta(n, shape1, shape2)</code>
t	<code>rt(n, df)</code>
Fisher	<code>rf(n, df1, df2)</code>
χ^2	<code>rchisq(n, df)</code>
Logística	<code>rlogis(n, location=0, scale=1)</code>
Lognormal	<code>rlnorm(n, meanlog=0, sdlog=1)</code>

Además, R tiene la función `sample` que permite obtener muestras con o sin reposición de conjuntos finitos de valores. La sintaxis es

```
sample(x, size, replace = FALSE, prob = NULL)
```

donde

- `x` es el conjunto a partir del cual queremos obtener la muestra, escrito como un vector,
- `size` es el tamaño de la muestra,
- `replace` permite indicar si se permiten repeticiones (`replace = TRUE`) o no y finalmente
- `prob` es un vector de probabilidades si se desea hacer un muestreo pesado y no uniforme.

1.13. Convergencia de Variables Aleatorias

Hay varios modos de convergencia en la Teoría de Probabilidades. Vamos a considerar algunos de ellos a continuación. Sea $X_n, n \geq 1$ una sucesión de variables aleatorias definidas en un espacio de probabilidad común (Ω, \mathcal{F}, P) y sea X otra variable definida sobre este mismo espacio.

Definición 1.4 La sucesión X_n converge *puntualmente* a X si para todo $\omega \in \Omega$ se cumple que

$$\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega).$$

Notación: $X_n \rightarrow X$.

Definición 1.5 La sucesión X_n converge *casi seguramente* o *con probabilidad 1* a X si existe un conjunto nulo $N \in \mathcal{F}$ tal que para todo $\omega \notin N$ se cumple que

$$\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega).$$

Notación: $X_n \rightarrow X$ c.s. o c.p.1, o también $X_n \xrightarrow{\text{c.s.}} X$.

Definición 1.6 La sucesión X_n converge *en probabilidad* a X si dado cualquier $\varepsilon > 0$ se tiene que

$$\lim_{n \rightarrow \infty} P(|X_n - X| > \varepsilon) = 0.$$

Notación: $X_n \xrightarrow{P} X$.

Definición 1.7 La sucesión X_n converge *en L^p* , $1 \leq p < \infty$, a X si $E[|X_n|^p] < \infty$ y

$$\lim_{n \rightarrow \infty} E[|X_n - X|^p] = 0.$$

Notación: $X_n \xrightarrow{L^p} X$ o también $X_n \rightarrow X$ en L^p .

Definición 1.8 La sucesión X_n converge *en distribución* a X

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x), \quad \text{para todo } x \in \mathcal{C}(F_X),$$

donde $\mathcal{C}(F_X)$ es el conjunto de puntos de continuidad de F_X .

Notación: $X_n \xrightarrow{D} X$. También usaremos la notación $X_n \xrightarrow{D} F_X$

Observación 1.4

1. Cuando consideramos la convergencia c.s. consideramos para cada $\omega \in \Omega$, si la sucesión de números reales $X_n(\omega)$ converge al número real $X(\omega)$. Si esto ocurre fuera de un conjunto de ω de medida 0, decimos que hay convergencia c.s.
2. La convergencia en L^2 se conoce usualmente como convergencia en media cuadrática.
3. En la definición de convergencia en distribución, las variables sólo aparecen a través de sus funciones de distribución. Por lo tanto las variables no tienen que estar definidas en un mismo espacio de probabilidad.
4. Es posible demostrar que una función de distribución tiene a lo sumo una cantidad numerable de discontinuidades. Como consecuencia $\mathcal{C}(F_X)$ es la recta real, excepto, posiblemente, por un conjunto numerable de puntos.
5. Es posible demostrar que en cualquiera de estos modos de convergencia el límite es (esencialmente) único.

▲

Ejemplo 1.26

Sea $X_n \sim \Gamma(n, n)$. Veamos que $X_n \xrightarrow{P} 1$ cuando $n \rightarrow \infty$.

Observamos que $E[X_n] = 1$ mientras que $\text{Var}[X_n] = 1/n$. Usando la desigualdad de Chebyshev obtenemos que para todo $\varepsilon > 0$,

$$P(|X_n - X| > \varepsilon) \leq \frac{1}{n\varepsilon^2} \rightarrow 0 \quad \text{cuando } n \rightarrow \infty.$$

▲

Ejemplo 1.27

Sean X_1, X_2, \dots v.a.i. con densidad común

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

y sea $Y_n = n^{-1/\alpha} \max_{1 \leq k \leq n} X_k$, $n \geq 1$. Demuestre que Y_n converge en distribución y determine la distribución límite.

Para resolver este problema vamos a calcular la f.d. común:

$$F(x) = \int_1^x \alpha x^{-\alpha-1} dy = 1 - x^{-\alpha}$$

siempre que $x > 1$ y vale 0 si no. Por lo tanto, para cualquier $x > 1$,

$$\begin{aligned} F_{Y_n}(x) &= P(\max_{1 \leq k \leq n} X_k \leq xn^{1/\alpha}) = (F(xn^{1/\alpha}))^n \\ &= \left(1 - \frac{1}{nx^\alpha}\right)^n \rightarrow e^{-x^{-\alpha}} \quad \text{cuando } n \rightarrow \infty. \end{aligned}$$

▲

Ejemplo 1.28 (La Ley de los Grandes Números)

Esta es una versión débil de la LGN. Sean X_1, X_2, \dots v.a.i.i.d. con media μ y varianza finita σ^2 y pongamos $S_n = X_1 + \dots + X_n$, $n \geq 1$. La Ley (Débil) de los Grandes Números dice que para todo $\varepsilon > 0$,

$$P\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) \rightarrow 0 \quad \text{cuando } n \rightarrow \infty,$$

es decir

$$\frac{S_n}{n} \xrightarrow{P} \mu \quad \text{cuando } n \rightarrow \infty.$$

La prueba de esta proposición sigue de la desigualdad de Chebyshev:

$$P\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon^2} \rightarrow 0 \quad \text{cuando } n \rightarrow \infty.$$

▲

Ejemplo 1.29 (Aproximación de Poisson)

Sea $X_n \sim \text{Bin}(n, \frac{\lambda}{n})$, entonces

$$X_n \xrightarrow{D} \text{Pois}(\lambda)$$

Vemos esto

$$\begin{aligned} P(X_n = k) &= \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} = \frac{n(n-1)\cdots(n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{n(n-1)\cdots(n-k+1)}{n^k} \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{\left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right)}{\left(1 - \frac{\lambda}{n}\right)^k} \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \end{aligned}$$

Si ahora hacemos $n \rightarrow \infty$ la primera fracción tiende a 1 porque k y λ están fijos, mientras que

$$\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}$$

Por lo tanto

$$\lim_{n \rightarrow \infty} P(X_n = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

▲

1.13.1. Relación entre los Distintos Tipos de Convergencia

En esta sección nos planteamos investigar la relación entre los distintos tipos de convergencia que hemos definido, y en particular exploramos la posibilidad de poder ordenar estos conceptos.

I. Convergencia c.s. implica Convergencia en Probabilidad.

Es posible demostrar que $X_n \xrightarrow{c.s.} X$ cuando $n \rightarrow \infty$ sí y sólo sí para todo $\varepsilon > 0$ y δ , $0 < \delta < 1$, existe n_0 tal que, para todo $n > n_0$

$$P\left(\bigcap_{m>n} \{|X_m - X| < \varepsilon\}\right) > 1 - \delta \quad (1.41)$$

o equivalentemente

$$P\left(\bigcup_{m>n} \{|X_m - X| > \varepsilon\}\right) < \delta.$$

Como, para $m > n$,

$$\{|X_m - X| > \varepsilon\} \subset \bigcup_{m>n} \{|X_m - X| > \varepsilon\},$$

la sucesión también converge en probabilidad. El siguiente ejemplo muestra que el recíproco es falso.

Ejemplo 1.30

Sean X_1, X_2, \dots v.a.i. tales que

$$P(X_n = 1) = 1 - \frac{1}{n} \quad \text{y} \quad P(X_n = n) = \frac{1}{n}, \quad n \geq 1.$$

Claramente,

$$P(|X_n - 1| > \varepsilon) = P(X_n = n) = \frac{1}{n} \rightarrow 0, \quad \text{cuando } n \rightarrow \infty,$$

para todo $\varepsilon > 0$, es decir,

$$X_n \xrightarrow{P} 1 \quad \text{cuando } n \rightarrow \infty.$$

Veamos ahora que X_n no converge c.s. a 1 cuando $n \rightarrow \infty$. Para todo $\varepsilon > 0$, $\delta \in (0, 1)$ y $N > n$ tenemos

$$\begin{aligned}
 P\left(\bigcap_{m>n} \{|X_m - X| < \varepsilon\}\right) &= P\left(\lim_N \bigcap_{m=n+1}^N \{|X_m - X| < \varepsilon\}\right) \\
 &= \lim_N P\left(\bigcap_{m=n+1}^N \{|X_m - X| < \varepsilon\}\right) \\
 &= \lim_N \prod_{m=n+1}^N P(|X_m - 1| < \varepsilon) \\
 &= \lim_N \prod_{m=n+1}^N P(X_m = 1) = \lim_N \prod_{m=n+1}^N \left(1 - \frac{1}{m}\right) \\
 &= \lim_N \prod_{m=n+1}^N \frac{m-1}{m} = \lim_N \frac{n}{N} = 0,
 \end{aligned}$$

para cualquier n . Esto muestra que no existe n_0 para el cual (1.41) valga, y por lo tanto X_n no converge c.s. a 1 cuando $n \rightarrow \infty$. ▲

II. Convergencia en L^p implica Convergencia en Probabilidad

Usando la desigualdad de Markov, para $\varepsilon > 0$ fijo

$$P(|X_n - X| > \varepsilon) \leq \frac{1}{\varepsilon^p} E[|X_n - X|^p] \rightarrow 0$$

cuando $n \rightarrow \infty$, lo que muestra la conclusión.

En este caso el recíproco tampoco es cierto. Para empezar, $E[|X_n - X|]$ no tiene por qué existir, pero aun si existe puede ocurrir que haya convergencia en probabilidad sin que haya convergencia en L^p .

Ejemplo 1.31

Sea $\alpha > 0$ y sea X_1, X_2, \dots v.a. tales que

$$P(X_n = 1) = 1 - \frac{1}{n^\alpha} \quad \text{y} \quad P(X_n = n) = \frac{1}{n^\alpha}, \quad n \geq 1.$$

Como

$$P(|X_n - 1| > \varepsilon) = P(X_n = n) = \frac{1}{n^\alpha} \rightarrow 0, \quad \text{cuando } n \rightarrow \infty,$$

para todo $\varepsilon > 0$, tenemos que

$$X_n \xrightarrow{P} 1 \quad \text{cuando } n \rightarrow \infty.$$

Por otro lado

$$E[|X_n - 1|^p] = 0^p \cdot \left(1 - \frac{1}{n^\alpha}\right) + |n - 1|^p \frac{1}{n^\alpha} = \frac{(n-1)^p}{n^\alpha},$$

de donde obtenemos que

$$E[|X_n - 1|^p] \rightarrow \begin{cases} 0, & \text{para } p < \alpha, \\ 1, & \text{para } p = \alpha, \\ +\infty, & \text{para } p > \alpha, \end{cases} \quad (1.42)$$

Esto muestra que $X_n \xrightarrow{L^p} 1$ cuando $n \rightarrow \infty$ si $p < \alpha$ pero X_n no converge en L^p si $p \geq \alpha$. Por lo tanto, convergencia en L^p es más fuerte que convergencia en probabilidad. ▲

Observación 1.5 Si $\alpha = 1$ y las variables son independientes, cuando $n \rightarrow \infty$

$$\begin{aligned} X_n &\xrightarrow{P} 1, \\ X_n &\xrightarrow{\text{c.s.}}, \\ E[X_n] &\rightarrow 2, \\ X_n &\xrightarrow{L^p} 1 \quad \text{para } 0 < p < 1, \\ X_n &\xrightarrow{L^p} \quad \text{para } p \geq 1. \end{aligned}$$

▲

Observación 1.6 Si $\alpha = 2$ y las variables son independientes, cuando $n \rightarrow \infty$

$$\begin{aligned} X_n &\xrightarrow{P} 1, \\ X_n &\xrightarrow{\text{c.s.}} 1, \\ E[X_n] &\rightarrow 1, \quad \text{y} \quad \text{Var}[X_n] \rightarrow 1 \\ X_n &\xrightarrow{L^p} 1 \quad \text{para } 0 < p < 2, \\ X_n &\xrightarrow{L^p} \quad \text{para } p \geq 2. \end{aligned}$$

▲

III. Convergencia en L^p y Convergencia c.s. son Independientes

Ninguna de las dos implica la otra, y esto lo podemos ver de las observaciones anteriores. En el primer caso, para $0 < p < 1$ hay convergencia en L^p mientras que no hay convergencia c.s. En el segundo hay convergencia c.s. pero no hay convergencia en L^p para $p \geq 2$.

IV. Convergencia en Probabilidad implica Convergencia en Distribución

Sea $\varepsilon > 0$, entonces

$$\begin{aligned} F_{X_n}(x) &= P(X_n \leq x) \\ &= P(\{X_n \leq x\} \cap \{|X_n - X| \leq \varepsilon\}) + P(\{X_n \leq x\} \cap \{|X_n - X| > \varepsilon\}) \\ &\leq P(\{X \leq x + \varepsilon\} \cap \{|X_n - X| \leq \varepsilon\}) + P(|X_n - X| > \varepsilon) \\ &\leq P(X \leq x + \varepsilon) + P(|X_n - X| > \varepsilon) \end{aligned}$$

es decir,

$$F_{X_n}(x) \leq F_X(x + \varepsilon) + P(|X_n - X| > \varepsilon). \quad (1.43)$$

De manera similar se demuestra que

$$F_X(x - \varepsilon) \leq F_{X_n}(x) + P(|X_n - X| > \varepsilon). \quad (1.44)$$

Como $X_n \xrightarrow{P} X$ cuando $n \rightarrow \infty$ obtenemos, haciendo $n \rightarrow \infty$ en (1.43) y (1.44),

$$F_X(x - \varepsilon) \leq \liminf_{n \rightarrow \infty} F_{X_n}(x) \leq \limsup_{n \rightarrow \infty} F_{X_n}(x) \leq F_X(x + \varepsilon)$$

Esta relación es válida para todo x y todo $\varepsilon > 0$. Para demostrar la convergencia en distribución suponemos que $x \in \mathcal{C}(F_X)$ y hacemos $\varepsilon \rightarrow 0$. Obtenemos

$$F_X(x) \leq \liminf_{n \rightarrow \infty} F_{X_n}(x) \leq \limsup_{n \rightarrow \infty} F_{X_n}(x) \leq F_X(x),$$

por lo tanto

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x),$$

y como esto vale para cualquier $x \in \mathcal{C}(F_X)$ obtenemos la convergencia en distribución. \blacktriangle

Observación 1.7 Observamos que si F_X tiene un salto en x , sólo podemos concluir que

$$F_X(x-) \leq \liminf_{n \rightarrow \infty} F_{X_n}(x) \leq \limsup_{n \rightarrow \infty} F_{X_n}(x) \leq F_X(x),$$

y $F_X(x) - F_X(x-)$ es el tamaño del salto. Esto explica por qué sólo se toman en cuenta los puntos de continuidad en la definición de convergencia en distribución.

Como mencionamos anteriormente, la convergencia en distribución no requiere que las variables estén definidas en un mismo espacio de probabilidad, y por lo tanto es más débil que los otros modos de convergencia. El siguiente ejemplo muestra que aun cuando las distribuciones conjuntas existan, existen variables que convergen sólo en distribución.

Ejemplo 1.32

Sea X una variable con distribución simétrica, continua y no-degenerada y definimos X_1, X_2, \dots por $X_{2n} = X$ y $X_{2n-1} = -X$, $n = 1, 2, \dots$. Como $X_n \stackrel{D}{=} X$ para todo n , tenemos, en particular, $X_n \stackrel{D}{\rightarrow} X$ cuando $n \rightarrow \infty$. Por otro lado, como X tiene distribución no-degenerada existe $a > 0$ tal que $P(|X| > a) > 0$ (¿por qué?). En consecuencia, para todo $\varepsilon > 0$, $0 < \varepsilon < 2a$,

$$P(|X_n - X| > \varepsilon) = \begin{cases} 0, & \text{para } n \text{ par,} \\ P(|X| > \frac{\varepsilon}{2}) > 0, & \text{para } n \text{ impar.} \end{cases}$$

Esto muestra que X_n no puede converger en probabilidad a X cuando $n \rightarrow \infty$, y en consecuencia tampoco c.s. o en L^p . \blacktriangle

Podemos resumir todo lo anterior en el siguiente teorema.

Teorema 1.8 Sean X y X_1, X_2, \dots variables aleatorias, entonces, cuando $n \rightarrow \infty$,

$$\begin{array}{c} X_n \xrightarrow{\text{c.s.}} X \implies X_n \xrightarrow{P} X \implies X_n \xrightarrow{D} X \\ \uparrow \\ X_n \xrightarrow{L^p} X \end{array}$$

Ninguna de las implicaciones se puede invertir.

Capítulo 2

Cadenas de Markov

2.1. Introducción

Sea $T \subset \mathbb{R}$ y (Ω, \mathcal{F}, P) un espacio de probabilidad. Un *proceso aleatorio* es una función

$$X : T \times \Omega \rightarrow \mathbb{R}$$

tal que para cada $t \in T$, $X(t, \cdot)$ es una variable aleatoria.

Si fijamos $\omega \in \Omega$ obtenemos una función $X(\cdot, \omega) : T \rightarrow \mathbb{R}$ que se conoce como una *trayectoria* del proceso.

En general interpretamos el parámetro t como el tiempo aunque también se pueden considerar procesos con índices en espacios más generales. En este curso T será un subconjunto de \mathbb{R} . Los casos más comunes serán

- T discreto (Procesos a tiempo discreto): $T = \mathbb{N}$, $T = \{0, 1, 2, \dots\}$, $T = \mathbb{Z}$.
- T continuo (Procesos a tiempo continuo): $T = [0, 1]$, $T = [0, \infty)$, $T = \mathbb{R}$.

En cuanto a los valores del proceso llamaremos \mathcal{E} al *espacio de estados* y consideraremos también dos casos:

- Valores discretos, por ejemplo $\mathcal{E} = \{0, 1, 2, \dots\}$, $\mathcal{E} = \mathbb{N}$ o $\mathcal{E} = \mathbb{Z}$
- Valores continuos, por ejemplo $\mathcal{E} = [0, \infty)$, $\mathcal{E} = \mathbb{R}$, etc.

2.2. Definiciones

Hablando informalmente, un proceso de Markov es un proceso aleatorio con la propiedad de que dado el valor actual del proceso X_t , los valores futuros X_s para $s > t$ son independientes de los valores pasados X_u para $u < t$. Es decir, que si tenemos la información del estado presente del proceso, saber cómo llegó al estado actual no afecta las probabilidades de pasar a otro estado en el futuro. En el caso discreto la definición precisa es la siguiente.

Definición 2.1 Una *Cadena de Markov* a tiempo discreto es una sucesión de variables aleatorias X_n , $n \geq 1$ que toman valores en un conjunto finito o numerable \mathcal{E} , conocido como espacio de estados, y que satisface la siguiente propiedad

$$P(X_{n+1} = j | X_0 = i_0, \dots, X_{n-1} = i_{n-1}, X_n = i_n) = P(X_{n+1} = j | X_n = i_n) \quad (2.1)$$

para todo n y cualesquiera estados i_0, i_1, \dots, i_n, j en \mathcal{E} . La propiedad (2.1) se conoce como la *propiedad de Markov*.

Resulta cómodo designar los estados de la cadena usando los enteros no-negativos $\{0, 1, 2, \dots\}$ y diremos que X_n está en el estado i si $X_n = i$.

La probabilidad de que X_{n+1} esté en el estado j dado que X_n está en el estado i es la *probabilidad de transición* en un paso de i a j y la denotaremos P_{ij}^{n+1} :

$$P_{ij}^{n+1} = P(X_{n+1} = j | X_n = i). \quad (2.2)$$

En general, las probabilidades de transición dependen no sólo de los estados sino también del instante en el cual se efectúa la transición. Cuando estas probabilidades son independientes del tiempo (o sea, de n) decimos que la cadena tiene probabilidades de transición *estacionarias* u *homogéneas en el tiempo*. En este caso $P_{ij}^{n+1} = P_{ij}$ no depende de n y P_{ij} es la probabilidad de que la cadena pase del estado i al estado j en un paso. A continuación sólo consideraremos cadenas con probabilidades de transición estacionarias.

Podemos colocar las probabilidades de transición en una matriz

$$P = \begin{pmatrix} P_{00} & P_{01} & P_{02} & P_{03} & \cdots \\ P_{10} & P_{11} & P_{12} & P_{13} & \cdots \\ P_{20} & P_{21} & P_{22} & P_{23} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \\ P_{i0} & P_{i1} & P_{i2} & P_{i3} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{pmatrix}$$

que será finita o infinita según el tamaño de \mathcal{E} . P se conoce como la *matriz de transición* o la *matriz de probabilidades de transición* de la cadena. La i -ésima fila de P para $i = 0, 1, \dots$ es la distribución condicional de X_{n+1} dado que $X_n = i$. Si el número de estados es finito, digamos k entonces P es una matriz cuadrada cuya dimensión es $k \times k$. Es inmediato que

$$P_{ij} = P(X_{n+1} = j | X_n = i) \geq 0, \quad \text{para } i, j = 0, 1, 2, \dots \quad (2.3)$$

$$\sum_{j=0}^{\infty} P_{ij} = \sum_{j=0}^{\infty} P(X_{n+1} = j | X_n = i) = 1, \quad \text{para } i = 0, 1, 2, \dots \quad (2.4)$$

de modo que cada fila de la matriz representa una distribución de probabilidad. Una matriz con esta propiedad se llama una *matriz estocástica* o *de Markov*.

Ejemplo 2.1 (Línea telefónica)

Consideremos una línea telefónica que puede tener dos estados, libre (0) u ocupada (1), y para simplificar vamos a considerar su comportamiento en los intervalos de tiempo de la forma $[n, n+1)$. Para cualquiera de estos períodos la probabilidad de que llegue una llamada es $\alpha \in [0, 1]$. Si una nueva llamada llega cuando la línea está ocupada, ésta no se registra, mientras que si la línea está libre, se toma la llamada y la línea pasa a estar ocupada. La probabilidad de que la línea se desocupe es $\beta \in [0, 1]$. En cada intervalo de tiempo puede llegar una llamada o se puede desocupar la línea, pero no ambas cosas.

Esta situación se puede modelar por una cadena de Markov con espacio de estados $\mathcal{E} = \{0, 1\}$. Las probabilidades de transición están dadas por

$$\begin{aligned} P(X_{n+1} = 0 | X_n = 0) &= 1 - \alpha = P_{00} \\ P(X_{n+1} = 1 | X_n = 0) &= \alpha = P_{01} \\ P(X_{n+1} = 1 | X_n = 1) &= 1 - \beta = P_{11} \\ P(X_{n+1} = 0 | X_n = 1) &= \beta = P_{10} \end{aligned}$$

Por lo tanto la matriz de transición es

$$P = \begin{pmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{pmatrix}.$$

Un caso un poco más complicado es el siguiente. Supongamos ahora que si la línea esta ocupada y llega una llamada, ésta se guarda y permanece en espera hasta que la línea se desocupa. Pero si hay una llamada en espera no se registran las siguientes llamadas. En cada período entra a lo sumo una llamada a la cola. Cuando en un período se desocupa la línea, se identifica el fin del período con el momento de colgar, de modo tal que después de haber colgado ya no se registran más llamadas en ese período. Como en el ejemplo anterior $\alpha \in [0, 1]$ denota la probabilidad de que llegue una llamada y $\beta \in [0, 1]$ la probabilidad de que la línea se desocupe. Suponemos además que en un mismo período no puede ocurrir que el teléfono esté ocupado con llamada en espera, cuelgue, entre la llamada que estaba en espera y entre otra llamada en espera.

Para este caso consideramos un espacio de estados con tres elementos: 0 denota que la línea está libre, 1 cuando la línea está ocupada y no hay llamada en espera y 2 cuando la línea está ocupada y hay una llamada en espera. Para simplificar las expresiones vamos a usar la notación $\alpha' = 1 - \alpha$, $\beta' = 1 - \beta$. Las probabilidades de transición para $n \in \mathbb{N}$ son

$$\begin{aligned} P(X_{n+1} = 0|X_n = 0) &= \alpha', & P(X_{n+1} = 1|X_n = 0) &= \alpha, & P(X_{n+1} = 2|X_n = 0) &= 0, \\ P(X_{n+1} = 0|X_n = 1) &= \beta, & P(X_{n+1} = 1|X_n = 1) &= \alpha'\beta', & P(X_{n+1} = 2|X_n = 1) &= \alpha\beta', \\ P(X_{n+1} = 0|X_n = 2) &= 0, & P(X_{n+1} = 1|X_n = 2) &= \beta, & P(X_{n+1} = 2|X_n = 2) &= \beta'. \end{aligned}$$

▲

Ejemplo 2.2 (Paseo al Azar o Caminata Aleatoria)

Sean $\{Y_i, i \in \mathbb{N}\}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas, definidas sobre un espacio de probabilidad $(\Omega, \mathcal{F}; P)$ y que toman valores en los enteros $\mathcal{E} = \mathbb{Z}$; denotaremos por p_Y la distribución de Y_i , es decir $p_Y(x) = P(Y_i = x), x \in \mathbb{Z}$. Consideremos la sucesión

$$X_n = \sum_{i=0}^n Y_i, \quad n \in \mathbb{N}.$$

Veamos que esta sucesión es una cadena de Markov con espacio de estados \mathbb{Z} y determinemos sus probabilidades de transición.

Es evidente que el espacio de estados del proceso X es \mathbb{Z} ya que la suma de dos enteros es un entero. Para demostrar la propiedad de Markov bastará con probar que para todo $n \in \mathbb{N}$ y para cualesquiera $x_0, \dots, x_n, x_{n+1} \in \mathbb{Z}$ se tiene que

$$P(X_{n+1} = x_{n+1}|X_n = x_n, \dots, X_0 = x_0) = P(X_{n+1} = x_{n+1}|X_n = x_n). \quad (2.5)$$

Para ver esto comenzamos por calcular

$$P(X_k = x_k, \dots, X_1 = x_1, X_0 = x_0), \quad x_0, x_1, \dots, x_k \in \mathbb{Z} \text{ y } k \in \mathbb{N}.$$

Teniendo en cuenta que $X_n = \sum_{i=0}^n Y_i$, $n \in \mathbb{N}$ se obtiene que

$$\begin{aligned} P(X_k = x_k, \dots, X_1 = x_1, X_0 = x_0) &= P(Y_0 = x_0, Y_1 = x_1 - x_0, \dots, Y_k = x_k - x_{k-1}) \\ &= p_Y(x_0)p_Y(x_1 - x_0) \cdots p_Y(x_k - x_{k-1}). \end{aligned}$$

Usando este cálculo es inmediato que

$$P(X_{n+1} = x_{n+1}|X_n = x_n, \dots, X_0 = x_0) = p_Y(x_{n+1} - x_n).$$

Un cálculo similar muestra que

$$P(X_{n+1} = x_{n+1} | X_n = x_n) = \frac{P(Y_{n+1} = x_{n+1} - x_n, X_n = x_n)}{P(X_n = x_n)} = p_Y(x_{n+1} - x_n) \quad (2.6)$$

y por lo tanto (2.5) se satisface. Las probabilidades de transición están dadas por la ecuación (2.6). \blacktriangle

Una cadena de Markov está completamente determinada si se especifican su matriz de transición y la distribución de probabilidad del estado inicial X_0 . Veamos esto: Sea $P(X_0 = x_i) = \pi(x_i)$ para $i \geq 1$. Es suficiente mostrar como se calculan las probabilidades

$$P(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) \quad (2.7)$$

ya que cualquier probabilidad que involucre a X_{j_1}, \dots, X_{j_k} , $j_1 < j_2 < \dots < j_k$ se puede obtener usando la Ley de la Probabilidad Total y sumando términos como (2.7). Tenemos

$$\begin{aligned} P(X_0 = x_0, \dots, X_n = x_n) &= P(X_n = x_n | X_0 = x_0, \dots, X_{n-1} = x_{n-1}) \\ &\quad \times P(X_0 = x_0, \dots, X_{n-1} = x_{n-1}), \end{aligned}$$

pero

$$\begin{aligned} P(X_n = x_n | X_0 = x_0, \dots, X_{n-1} = x_{n-1}) &= P(X_n = x_n | X_{n-1} = x_{n-1}) \\ &= P_{x_{n-1}x_n} \end{aligned}$$

y sustituyendo en la ecuación anterior

$$\begin{aligned} P(X_0 = x_0, \dots, X_n = x_n) &= P_{x_{n-1}x_n} P(X_0 = x_0, \dots, X_{n-1} = x_{n-1}) \\ &= P_{x_{n-1}x_n} P_{x_{n-2}x_{n-1}} P(X_0 = x_0, \dots, X_{n-2} = x_{n-2}) \\ &= P_{x_{n-1}x_n} P_{x_{n-2}x_{n-1}} \cdots P_{x_0x_1} \pi(x_0). \end{aligned} \quad (2.8)$$

■

Un cálculo similar muestra que (2.1) es equivalente a

$$\begin{aligned} P(X_{n+1} = y_1, \dots, X_{n+m} = y_m | X_0 = x_0, \dots, X_{n-1} = x_{n-1}, X_n = x_n) \\ = P(X_{n+1} = y_1, \dots, X_{n+m} = y_m | X_n = x_n). \end{aligned}$$

2.2.1. Consecuencias de la Propiedad de Markov

En esta sección veremos algunas consecuencias de la propiedad de Markov (2.1).

Proposición 2.1 *La propiedad de Markov es equivalente a la siguiente condición: Para todo $n \in \mathbb{N}$ y $x \in \mathcal{E}$, condicionalmente a $X_n = x$, la distribución de X_{n+1} es $(P_{xy}, y \in \mathcal{E})$ y es independiente de X_0, \dots, X_{n-1} .*

Demostración. Comenzamos por suponer cierta la propiedad de Markov. Para probar la propiedad del enunciado basta demostrar que para todo $n \in \mathbb{N}$ y $x_0, x_1, \dots, x_n, x_{n+1}$ en \mathcal{E} se tiene que

$$\begin{aligned} P(X_0 = x_0, \dots, X_{n-1} = x_{n-1}, X_{n+1} = x_{n+1} | X_n = x_n) \\ = P(X_0 = x_0, \dots, X_{n-1} = x_{n-1} | X_n = x_n) P_{x_n, x_{n+1}}. \end{aligned} \quad (2.9)$$

En efecto, el lado izquierdo (LI) de la ecuación (2.9) se puede escribir como

$$\begin{aligned}
 LI &= \frac{P(X_0 = x_0, \dots, X_{n-1} = x_{n-1}, X_{n+1} = x_{n+1}, X_n = x_n)}{P(X_n = x_n)} \\
 &= P(X_{n+1} = x_{n+1} | X_0 = x_0, \dots, X_{n-1} = x_{n-1}, X_n = x_n) \\
 &\quad \times \frac{P(X_0 = x_0, \dots, X_{n-1} = x_{n-1}, X_n = x_n)}{P(X_n = x_n)} \\
 &= P_{x_n, x_{n+1}} P(X_0 = x_0, \dots, X_{n-1} = x_{n-1} | X_n = x_n),
 \end{aligned}$$

la última igualdad es consecuencia de la propiedad de Markov y de la homogeneidad de la cadena. Ahora demostraremos que la afirmación en la proposición implica la propiedad de Markov. Para ello basta demostrar que la probabilidad

$$P(X_{n+1} = x_{n+1} | X_0 = x_0, \dots, X_{n-1} = x_{n-1}, X_n = x_n),$$

no depende de x_0, \dots, x_{n-1} . En efecto, el resultado se deduce de la siguiente serie de igualdades:

$$\begin{aligned}
 &P(X_{n+1} = x_{n+1} | X_n = x_n, X_{n-1} = x_{n-1}, \dots, X_0 = x_0) \\
 &= \frac{P(X_{n+1} = x_{n+1}, X_n = x_n, X_{n-1} = x_{n-1}, \dots, X_0 = x_0)}{P(X_0 = x_0, \dots, X_{n-1} = x_{n-1}, X_n = x_n)} \\
 &= \frac{P(X_{n+1} = x_{n+1}, X_{n-1} = x_{n-1}, \dots, X_0 = x_0 | X_n = x_n)}{P(X_0 = x_0, \dots, X_{n-1} = x_{n-1} | X_n = x_n)} \\
 &= \frac{P(X_0 = x_0, \dots, X_{n-1} = x_{n-1} | X_n = x_n) \times P_{x_n, x_{n+1}}}{P(X_0 = x_0, \dots, X_{n-1} = x_{n-1} | X_n = x_n)} \\
 &= P_{x_n, x_{n+1}}.
 \end{aligned}$$

■

El siguiente resultado refuerza la idea de que si conocemos el presente, el pasado no tiene ninguna influencia en el comportamiento futuro de una cadena de Markov.

Proposición 2.2 Sean X una cadena de Markov (π, P) , con espacio de estados \mathcal{E} , $x, y, z \in \mathcal{E}$ y $0 \leq m \leq n-1$, $n, m \in \mathbb{N}$. Se tiene que

$$P(X_{n+1} = y | X_n = x, X_m = z) = P(X_{n+1} = y | X_n = x) = P_{x,y}.$$

Demostración. Para simplificar la notación haremos la prueba solamente en el caso en que $m = 0$, la prueba del caso general es muy similar. Usaremos el hecho que Ω se puede escribir como la unión disjunta de los siguientes conjuntos:

$$\Omega = \bigcup_{x_1, \dots, x_{n-1} \in \mathcal{E}} \{X_1 = x_1, \dots, X_{n-1} = x_{n-1}\},$$

y en particular $\Omega = \{X_i \in \mathcal{E}\}$, para calcular la probabilidad $P(X_{n+1} = y | X_n = x, X_0 = z)$. Se tienen las siguientes igualdades:

$$\begin{aligned}
P(X_{n+1} = y | X_n = x, X_0 = z) &= \frac{P(X_{n+1} = y, X_n = x, X_{n-1} \in \mathcal{E}, \dots, X_1 \in \mathcal{E}, X_0 = z)}{P(X_n = x, X_0 = z)} \\
&= \sum_{x_1, \dots, x_{n-1} \in \mathcal{E}} \frac{P(X_{n+1} = y, X_n = x, X_{n-1} = x_{n-1}, \dots, X_1 = x_1, X_0 = z)}{P(X_n = x, X_0 = z)} \\
&= \sum_{x_1, \dots, x_{n-1} \in \mathcal{E}} P(X_{n+1} = y | X_n = x, X_{n-1} = x_{n-1}, \dots, X_1 = x_1, X_0 = z) \\
&\quad \times \frac{P(X_n = x, X_{n-1} = x_{n-1}, \dots, X_1 = x_1, X_0 = z)}{P(X_n = x, X_0 = z)} \\
&= \sum_{x_1, \dots, x_{n-1} \in \mathcal{E}} P(X_{n+1} = y | X_n = x) \\
&\quad \times \frac{P(X_n = x, X_{n-1} = x_{n-1}, \dots, X_1 = x_1, X_0 = z)}{P(X_n = x, X_0 = z)} \\
&= P_{x,y} \sum_{x_1, \dots, x_{n-1} \in \mathcal{E}} \frac{P(X_n = x, X_{n-1} = x_{n-1}, \dots, X_1 = x_1, X_0 = z)}{P(X_n = x, X_0 = z)} \\
&= P_{x,y} \frac{P(X_n = x, X_{n-1} \in \mathcal{E}, \dots, X_1 \in \mathcal{E}, X_0 = z)}{P(X_n = x, X_0 = z)} \\
&= P_{x,y},
\end{aligned}$$

para justificar estas igualdades utilizamos el hecho de que la probabilidad de la unión numerable de conjuntos disjuntos es igual a la suma de las probabilidades de dichos conjuntos y la propiedad de Markov. ■

De manera análoga se puede demostrar que al condicionar con respecto a cualquier evento del pasado la propiedad de Markov sigue siendo válida en el sentido siguiente.

Proposición 2.3 Sean $y, x \in \mathcal{E}$, $n \in \mathbb{N}$ y $A_0, A_1, \dots, A_{n-1} \subseteq \mathcal{E}$. Se tiene que

$$P(X_{n+1} = y | X_n = x, X_{n-1} \in A_{n-1}, \dots, X_0 \in A_0) = P(X_{n+1} = y | X_n = x). \quad (2.10)$$

Otra forma de probar este resultado es utilizando la forma equivalente de la propiedad de Markov que fue enunciada en la Proposición 2.1. En efecto, basta con darse cuenta que el lado izquierdo de (2.10) puede escribirse como sigue

$$\begin{aligned}
P(X_{n+1} = y | X_n = x, X_{n-1} \in A_{n-1}, \dots, X_0 \in A_0) &= \frac{P(X_{n+1} = y, X_{n-1} \in A_{n-1}, \dots, X_0 \in A_0 | X_n = x)}{P(X_{n-1} \in A_{n-1}, \dots, X_0 \in A_0 | X_n = x)} \\
&= P(X_{n+1} = y | X_n = x),
\end{aligned}$$

donde la última igualdad es consecuencia de la ecuación (2.9).

Finalmente, la idea de que una cadena es homogénea en el tiempo será reforzada por el siguiente resultado, que dice que la ley de la cadena observada a partir del instante m es la misma que la de aquella observada al tiempo $n = 0$.

Proposición 2.4 Sean $m, k \in \mathbb{N}$ y $x_0, x_1, \dots, x_m, \dots, x_{m+k} \in \mathcal{E}$, se tiene que

$$\begin{aligned}
P(X_{m+1} = x_{m+1}, \dots, X_{m+k} = x_{m+k} | X_m = x_m, \dots, X_0 = x_0) \\
= P(X_1 = x_{m+1}, \dots, X_k = x_{m+k} | X_0 = x_m)
\end{aligned}$$

Demostración. Usando la ecuación (2.8) se tiene que

$$\begin{aligned}
& P(X_{m+1} = x_{m+1}, \dots, X_{m+k} = x_{m+k} | X_m = x_m, \dots, X_0 = x_0) \\
&= \frac{P(X_{m+1} = x_{m+1}, \dots, X_{m+k} = x_{m+k}, X_m = x_m, \dots, X_0 = x_0)}{P(X_m = x_m, \dots, X_0 = x_0)} \\
&= \frac{\pi(x_0) P_{x_0, x_1} \cdots P_{x_{m-1}, x_m} P_{x_m, x_{m+1}} \cdots P_{x_{m+k-1}, x_{m+k}}}{\pi(x_0) P_{x_0, x_1} \cdots P_{x_{m-1}, x_m}} \\
&= P_{x_m, x_{m+1}} \cdots P_{x_{m+k-1}, x_{m+k}} \\
&= P(X_1 = x_{m+1}, \dots, X_k = x_{m+k} | X_0 = x_m)
\end{aligned}$$

2.2.2. Ejemplos

Ejemplo 2.3

Sea $\xi_i, i \geq 1$ v.a.i.i.d. con valores sobre los enteros positivos: $P(\xi_i = j) = p_j$ para $j \geq 0$. Construiremos varios ejemplos con base en esta sucesión.

a) El primer ejemplo es la sucesión (ξ_i) con ξ_0 fijo. La matriz de transición es

$$P = \begin{pmatrix} p_0 & p_1 & p_2 & p_3 & \cdots \\ p_0 & p_1 & p_2 & p_3 & \cdots \\ p_0 & p_1 & p_2 & p_3 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

El hecho de que todas las filas sea idénticas refleja la independencia de las variables.

b) Sea $S_n = \xi_1 + \cdots + \xi_n, n = 1, 2, \dots$ y $S_0 = 0$ por definición. Este proceso es una cadena de Markov:

$$\begin{aligned}
P(S_{n+1} = j | S_1 = i_1, \dots, S_n = i_n) &= P(S_n + \xi_{n+1} = j | S_1 = i_1, \dots, S_n = i_n) \\
&= P(i_n + \xi_{n+1} = j | S_1 = i_1, \dots, S_n = i_n) \\
&= P(i_n + \xi_{n+1} = j | S_n = i_n) \\
&= P(S_{n+1} = j | S_n = i_n).
\end{aligned}$$

Por otro lado tenemos

$$\begin{aligned}
P(S_{n+1} = j | S_n = i) &= P(S_n + \xi_{n+1} = j | S_n = i) \\
&= P(\xi_{n+1} = j - i | S_n = i) \\
&= \begin{cases} p_{j-i}, & \text{para } j \geq i, \\ 0, & \text{para } j < i, \end{cases}
\end{aligned}$$

y la matriz de transición es

$$P = \begin{pmatrix} p_0 & p_1 & p_2 & p_3 & \cdots \\ 0 & p_0 & p_1 & p_2 & \cdots \\ 0 & 0 & p_0 & p_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

c) Si las variables ξ_i pueden tomar valores positivos y negativos entonces las sumas S_n toman valores en

\mathbb{Z} y la matriz de transición es

$$P = \begin{pmatrix} \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \cdots & p_{-1} & p_0 & p_1 & p_2 & p_3 & p_4 & \cdots & \cdots \\ \cdots & p_{-2} & p_{-1} & p_0 & p_1 & p_2 & p_3 & \cdots & \cdots \\ \cdots & p_{-3} & p_{-2} & p_{-1} & p_0 & p_1 & p_2 & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

(d) Máximos Sucesivos.

Sea $M_n = \max\{\xi_1, \dots, \xi_n\}$, para $n = 1, 2, \dots$ con $M_0 = 0$. El proceso M_n es una cadena de Markov y la relación

$$M_{n+1} = \max\{M_n, \xi_{n+1}\}$$

nos permite obtener la matriz de transición

$$P = \begin{pmatrix} Q_0 & p_1 & p_2 & p_3 & \cdots \\ 0 & Q_1 & p_2 & p_3 & \cdots \\ 0 & 0 & Q_2 & p_3 & \cdots \\ 0 & 0 & 0 & Q_3 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

donde $Q_k = p_0 + p_1 + \dots + p_k$ para $k \geq 0$.

▲

Ejemplo 2.4 (El Paseo al Azar Simple)

Consideremos una sucesión de juegos de azar en los que en cada juego ganamos \$1 con probabilidad $p = 0.4$ y perdemos \$1 con probabilidad $1 - p = 0.6$. Supongamos que decidimos dejar de jugar si nuestro capital llega a N o si nos arruinamos. El capital inicial es X_0 y X_n es nuestro capital al cabo de n juegos. Sea ξ_i el resultado del i -ésimo juego:

$$\xi_i = \begin{cases} +1, & \text{con probabilidad } p, \\ -1, & \text{con probabilidad } q, \end{cases}$$

entonces

$$X_n = X_0 + \xi_1 + \dots + \xi_n$$

y estamos en la situación del ejemplo anterior,

$$\begin{aligned} P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_1 = i_1) &= P(X_n + \xi_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_1 = i_1) \\ &= P(X_n + \xi_{n+1} = j | X_n = i) = P(\xi_{n+1} = j - i | X_n = i) \\ &= P(\xi_{n+1} = j - i) = \begin{cases} 0.4, & \text{si } j = i + 1, \\ 0.6, & \text{si } j = i - 1, \\ 0, & \text{en otro caso.} \end{cases} \end{aligned}$$

La matriz de transición en este caso es

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0.6 & 0 & 0.4 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0.6 & 0 & 0.4 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0.6 & 0 & 0.4 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 \end{pmatrix}$$

Con mayor generalidad podemos pensar que una partícula describe el paseo al azar cuyos estados son un conjunto de enteros finito o infinito, por ejemplo $a, a + 1, \dots, b - 1, b$. Si la partícula se encuentra en el estado i , en una transición puede quedarse en i o pasar a los estados $i + 1$ o $i - 1$. Supongamos que las probabilidades de transición son estacionarias y llamémoslas r, p y q , respectivamente, con $r + p + q = 1$. Hay dos casos especiales en los extremos de la cadena. Si ponemos $a = 0$ y $b = N$ entonces

$$\begin{aligned} P(X_n = 0 | X_{n-1} = 0) &= r_0, & P(X_n = N | X_{n-1} = N) &= r_N, \\ P(X_n = 1 | X_{n-1} = 0) &= p_0, & P(X_n = N - 1 | X_{n-1} = N) &= q_N, \\ P(X_n = -1 | X_{n-1} = 0) &= 0, & P(X_n = N + 1 | X_{n-1} = N) &= 0, \end{aligned}$$

y la matriz de transición es

$$P = \begin{pmatrix} r_0 & p_0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ q & r & p & 0 & \cdots & 0 & 0 & 0 \\ 0 & q & r & p & \cdots & 0 & 0 & 0 \\ 0 & 0 & q & r & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & q & r & p \\ 0 & 0 & 0 & 0 & \cdots & 0 & q_N & r_N \end{pmatrix}$$

El paseo al azar simétrico corresponde al caso $r = 0, p = q = 1/2$ y representa una aproximación discreta de un Movimiento Browniano. Si $p_0 = 0, r_0 = 1$ decimos que el estado 0 es *absorbente* o que 0 es una *barrera absorbente*. Si en cambio $p_0 = 1, r_0 = 0$, al llegar al estado 0 la partícula regresa inmediatamente al estado 1. Decimos en este caso que 0 es un *estado reflector* o que es una *barrera reflectora*. Algo similar ocurre para el estado N . Si $0 < p_0, q_0 < 1$ el estado 0 es un *reflector parcial* o una *barrera parcialmente reflectora*. ▲

Ejemplo 2.5 (El Modelo de Ehrenfest)

Este modelo, propuesto inicialmente por Paul y Tatiana Ehrenfest, representa una descripción matemática simplificada del proceso de difusión de gases o líquidos a través de una membrana. El modelo consiste de dos cajas A y B que contienen un total de N bolas. Seleccionamos al azar una de las N bolas y la colocamos en la otra caja. Sea X_n el número de bolas en la caja A después de la n -ésima transición; X_n es una cadena de Markov:

$$P(X_{n+1} = i + 1 | X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = \frac{N - i}{N},$$

ya que para aumentar el número de bolas en A hay que escoger una de las bolas en B . Similarmente,

$$P(X_{n+1} = i - 1 | X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = \frac{i}{N}.$$

Resumiendo, las probabilidades de transición son

$$P_{ii+1} = \frac{N - i}{N}, \quad P_{ii-1} = \frac{i}{N}.$$

Para el caso $N = 5$, por ejemplo, la matriz de transición es

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1/5 & 0 & 4/5 & 0 & 0 & 0 \\ 0 & 2/5 & 0 & 3/5 & 0 & 0 \\ 0 & 0 & 3/5 & 0 & 2/5 & 0 \\ 0 & 0 & 0 & 4/5 & 0 & 1/5 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

▲

Ejemplo 2.6 (Modelo de Inventario)

Una tienda de aparatos electrónicos vende un sistema de juegos de video y opera bajo el siguiente esquema: Si al final del día el número de unidades disponibles es 1 ó 0, se ordenan nuevas unidades para llevar el total a 5. Para simplificar supondremos que la nueva mercancía llega antes de que la tienda abra al día siguiente. Sea X_n el número de unidades disponibles al final del n -ésimo día y supongamos que el número de clientes que quieren comprar un juego en un día es 0, 1, 2, ó 3 con probabilidades 0.3; 0.4; 0.2 y 0.1 respectivamente. Tenemos entonces la siguiente matrix de transición

$$P = \begin{pmatrix} 0 & 0 & 0.1 & 0.2 & 0.4 & 0.3 \\ 0 & 0 & 0.1 & 0.2 & 0.4 & 0.3 \\ 0.3 & 0.4 & 0.3 & 0 & 0 & 0 \\ 0.1 & 0.2 & 0.4 & 0.3 & 0 & 0 \\ 0 & 0.1 & 0.2 & 0.4 & 0.3 & 0 \\ 0 & 0 & 0.1 & 0.2 & 0.4 & 0.3 \end{pmatrix}$$

Esta cadena es un ejemplo de una política de control de inventarios (s, S) con $s = 1$ y $S = 5$: cuando el stock disponible cae a s o por debajo de s , se ordena suficiente mercancía para llevar el stock a $S = 5$. Sea D_n la demanda en el n -ésimo día. Tenemos

$$X_{n+1} = \begin{cases} (X_n - D_{n+1})^+ & \text{si } X_n > s, \\ (S - D_{n+1})^+ & \text{si } X_n \leq s. \end{cases} \quad (2.11)$$

La descripción general de este esquema de inventario es la siguiente: se tiene un inventario de cierto producto con el fin de satisfacer la demanda. Suponemos que el inventario se repone al final de períodos que etiquetamos $n = 0, 1, 2, \dots$ y suponemos que la demanda total durante un período n es una v.a. X_n cuya distribución es independiente del período (es decir, es estacionaria):

$$P(D_n = k) = p_k, \quad k = 0, 1, 2, \dots$$

donde $p_k \geq 0$, $\sum_k p_k = 1$.

El nivel del inventario se verifica al final de cada período y la política (s, S) de reposición ($s < S$) estipula que si el nivel de inventario no está por encima de s , se ordena una cantidad suficiente para llevar el inventario a S . Si, en cambio, el inventario disponible es mayor que s , no se produce una orden. Llamemos X_n al inventario disponible al final del n -ésimo período.

Hay dos situaciones posibles cuando la demanda excede al inventario:

1. La demanda no satisfecha se pierde.

En este caso el nivel del inventario nunca puede ser negativo y vale la relación (2.11).

2. La demanda no satisfecha en un período se satisface inmediatamente después de renovar el inventario.

En este caso el nivel del inventario puede ser negativo y satisface

$$X_{n+1} = \begin{cases} X_n - D_{n+1} & \text{si } X_n > s, \\ S - D_{n+1} & \text{si } X_n \leq s. \end{cases}$$

La sucesión $(X_n)_{n \geq 1}$ es una cadena de Markov con probabilidades de transición

$$P_{ij} = P(X_{n+1} = j | X_n = i) = \begin{cases} P(D_{n+1} = i - j) & \text{si } s < i < S, \\ P(D_{n+1} = S - j) & \text{en otro caso.} \end{cases}$$

▲

Ejemplo 2.7 (Rachas)

Realizamos una sucesión de juegos en idénticas condiciones con probabilidad de éxito (E) p y de fracaso (F) $q = 1 - p$. Decimos que ocurre una racha de longitud k en el juego n si han ocurrido k éxitos sucesivos en el instante n luego de un fracaso en el instante $n - k$

$$\begin{array}{ccccccccc} \underline{F} & \underline{E} & \underline{E} & \underline{E} & \dots & \underline{E} & & & \\ n-k & n-k+1 & n-k+2 & n-k+3 & & n & & & \end{array}$$

Para estudiar este proceso como una cadena de Markov definimos los siguientes estados: Si el ensayo resulta en fracaso, el estado es 0. Si resulta en éxito, el estado es el número de éxitos que han ocurrido en sucesión. Por lo tanto, desde cualquier estado i puede haber una transición al estado 0 (si hay un fracaso en el próximo juego) con probabilidad $1 - p$, mientras que si hay un éxito la racha continua y la transición es de i a $i + 1$. La matriz de transición es

$$P = \begin{pmatrix} q & p & 0 & 0 & 0 & \dots \\ q & 0 & p & 0 & 0 & \dots \\ q & 0 & 0 & p & 0 & \dots \\ q & 0 & 0 & 0 & p & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

▲

Ejemplo 2.8

Sean X_0 una variable aleatoria que toma valores en \mathcal{E} , $\{Y_n : \Omega \rightarrow S, n \in \mathbb{N}\}$ una sucesión de variables aleatorias independientes entre sí y de X_0 , con valores en un conjunto S y $F : \mathcal{E} \times S \rightarrow \mathcal{E}$. En general, cualquier fenómeno descrito por una relación en recurrencia aleatoria de la forma

$$X_{n+1} = F(X_n, Y_{n+1}), \quad n \in \mathbb{N},$$

es una cadena de Markov. Verifiquemos la propiedad de Markov

$$P(X_{n+1} = y | X_0 = x_0, \dots, X_{n-1} = x_{n-1}, X_n = x) = P(X_{n+1} = y | X_n = x)$$

para $n \in \mathbb{N}$ y $x_0, \dots, x_{n-1}, x, y \in \mathcal{E}$. Para ello observemos que existe una sucesión de funciones deterministas $g_n : \mathcal{E} \times S^n \rightarrow \mathcal{E}$, tal que $X_n = g_n(X_0, Y_1, \dots, Y_n)$. En efecto, estas pueden ser definidas por la recurrencia $g_0(x) = x, x \in \mathcal{E}$, y para $x \in \mathcal{E}$ y $z_1, \dots, z_n \in S$

$$g_n(x, z_1, \dots, z_n) = F(g_{n-1}(x, z_1, \dots, z_{n-1}), z_n), \quad n \geq 1.$$

Usando esto se tiene que

$$\begin{aligned} & P(X_{n+1} = y | X_0 = x_0, \dots, X_{n-1} = x_{n-1}, X_n = x) \\ &= \frac{P(X_0 = x_0, \dots, X_{n-1} = x_{n-1}, X_n = x, X_{n+1} = y)}{P(X_0 = x_0, \dots, X_{n-1} = x_{n-1}, X_n = x)} \\ &= \frac{P(X_0 = x_0, g_1(x_0, Y_1) = x_1, \dots, g_n(x_0, \tilde{x}_1, \dots, \tilde{x}_{n-1}, Y_n) = x, F(x, Y_{n+1}) = y)}{P(X_0 = x_0, g_1(x_0, Y_1) = x_1, \dots, g_n(x_0, \tilde{x}_1, \dots, \tilde{x}_{n-1}, Y_n) = x)} \\ &= \frac{P(X_0 = x_0, g_1(x_0, Y_1) = x_1, \dots, g_n(x_0, \tilde{x}_1, \dots, \tilde{x}_{n-1}, Y_n) = x) P(F(x, Y_{n+1}) = y)}{P(X_0 = x_0, g_1(x_0, Y_1) = x_1, \dots, g_n(x_0, \tilde{x}_1, \dots, \tilde{x}_{n-1}, Y_n) = x)} \\ &= P(F(x, Y_{n+1}) = y), \end{aligned}$$

donde \tilde{x}_i representa los valores tales que $F(x_{i-1}, \tilde{x}_i) = x_{i+1}$. Por otro lado, usando la independencia se tiene que

$$\begin{aligned} P(X_{n+1} = y | X_n = x) &= \frac{P(F(X_n, Y_{n+1}) = y, X_n = x)}{P(X_n = x)} \\ &= \frac{P(F(x, Y_{n+1}) = y, X_n = x)}{P(X_n = x)} \\ &= \frac{P(F(x, Y_{n+1}) = y)P(X_n = x)}{P(X_n = x)} \\ &= P(F(x, Y_{n+1}) = y). \end{aligned}$$

▲

2.3. Matrices de Transición

Una herramienta fundamental en el estudio de las cadenas de Markov lo constituyen las matrices de transición en n pasos: $P^{(n)} = (P_{ij}^{(n)})$, donde $P_{ij}^{(n)}$ denota la probabilidad de que el proceso pase del estado i al estado j en n pasos:

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

Recordamos que estamos trabajando con procesos cuyas matrices de transición son estacionarias.

Teorema 2.1 (Ecuaciones de Chapman-Kolmogorov) Si $P = (P_{ij})$ es la matriz de transición (en un paso) de una cadena de Markov, entonces

$$P_{ij}^{(n)} = \sum_{k \in \mathcal{E}} P_{ik}^{(r)} P_{kj}^{(s)} \quad (2.12)$$

para cualquier par fijo de enteros no-negativos r y s que satisfagan $r + s = n$, donde definimos

$$P_{ij}^{(0)} = \begin{cases} 1, & i = j, \\ 0, & i \neq j \end{cases}$$

Demostración.

Para calcular $P_{ij}^{(n)}$ hacemos una partición según los valores posibles de la cadena en el instante r :

$$\begin{aligned} P_{ij}^{(n)} &= P(X_n = j | X_0 = i) \\ &= \sum_k P(X_n = j, X_r = k | X_0 = i) \\ &= \sum_k \frac{P(X_n = j, X_r = k, X_0 = i)}{P(X_0 = i)} \\ &= \sum_k \frac{P(X_n = j, X_r = k, X_0 = i)}{P(X_r = k, X_0 = i)} \frac{P(X_r = k, X_0 = i)}{P(X_0 = i)} \\ &= \sum_k P(X_n = j | X_r = k, X_0 = i) P(X_r = k | X_0 = i) \\ &= \sum_k P(X_n = j | X_r = k) P(X_r = k | X_0 = i) \\ &= \sum_k P_{ik}^{(r)} P_{kj}^{(s)}. \end{aligned}$$

■

La relación (2.12) representa la multiplicación de las matrices $P^{(r)}$ y $P^{(s)}$, de modo que $P^{(n)}$ es simplemente la n -ésima potencia de P . Por lo tanto $P_{ij}^{(n)}$ es el elemento ij de la n -ésima potencia de la matriz P .

Si la probabilidad de que el proceso esté inicialmente en j es $\pi(j)$:

$$P(X_0 = j) = \pi(j)$$

entonces la probabilidad de que el proceso esté en el estado k en el instante n es

$$P(X_n = k) = P_k^{(n)} = \sum_j \pi(j) P_{jk}^{(n)}.$$

Notación. $P_i(A) = P(A|X_0 = i)$ y $E_i[\cdot] = E[\cdot | X_0 = i]$.

En general no es sencillo calcular las matrices de transición a n pasos. En el próximo ejemplo presentamos un caso particular en el cual esto se puede hacer explícitamente.

Ejemplo 2.9 (Cadena con dos estados)

Consideremos una cadena con dos estados, 0 y 1, y matriz de transición

$$P = \begin{pmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{pmatrix}$$

Si $\alpha = 1 - \beta$ las filas coinciden y se trata de v.a.i.i.d. con $P(X_n = 0) = \beta$, $P(X_n = 1) = \alpha$.

Para una cadena de estas dimensiones podemos calcular explícitamente la matriz de transición en n pasos. Vamos a demostrar que:

$$P^n = \frac{1}{\alpha + \beta} \begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix} + \frac{(1 - \alpha - \beta)^n}{\alpha + \beta} \begin{pmatrix} \alpha & -\alpha \\ -\beta & \beta \end{pmatrix} \quad (2.13)$$

Usando la notación

$$A = \begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix} \quad B = \begin{pmatrix} \alpha & -\alpha \\ -\beta & \beta \end{pmatrix}$$

tenemos

$$P^n = \frac{1}{\alpha + \beta} (A + (1 - \alpha - \beta)^n B).$$

Comenzamos por calcular los siguientes productos

$$AP = \begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix} \begin{pmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{pmatrix} = A$$

$$\begin{aligned} BP &= \begin{pmatrix} \alpha & -\alpha \\ -\beta & \beta \end{pmatrix} \begin{pmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{pmatrix} \\ &= \begin{pmatrix} \alpha - \alpha^2 - \alpha\beta & -\alpha + \alpha^2 + \alpha\beta \\ -\beta + \alpha\beta + \beta^2 & \beta - \alpha\beta - \beta^2 \end{pmatrix} \\ &= (1 - \alpha - \beta)B \end{aligned}$$

Demostraremos (2.13) por inducción. Para $n = 1$ tenemos

$$\begin{aligned} & \frac{1}{\alpha + \beta} \begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix} + \frac{(1 - \alpha - \beta)}{\alpha + \beta} \begin{pmatrix} \alpha & -\alpha \\ -\beta & \beta \end{pmatrix} \\ &= \frac{1}{\alpha + \beta} \begin{pmatrix} \beta + \alpha(1 - \alpha - \beta) & \alpha - \alpha(1 - \alpha - \beta) \\ \beta - \beta(1 - \alpha - \beta) & \alpha + \beta(1 - \alpha - \beta) \end{pmatrix} \\ &= \frac{1}{\alpha + \beta} \begin{pmatrix} \beta + \alpha - \alpha^2 - \alpha\beta & \alpha^2 + \alpha\beta \\ \alpha\beta + \beta^2 & \alpha + \beta - \alpha\beta - \beta^2 \end{pmatrix} = P \end{aligned}$$

Para completar la prueba por inducción supongamos cierta la fórmula para n , entonces

$$\begin{aligned} P^{n+1} &= P^n P = \frac{1}{\alpha + \beta} (A + (1 - \alpha - \beta)^n B) P \\ &= \frac{1}{\alpha + \beta} (A + (1 - \alpha - \beta)^{n+1} B) \end{aligned}$$

Observamos que $|1 - \alpha - \beta| < 1$ cuando $0 < \alpha, \beta < 1$ y, por lo tanto $(1 - \alpha - \beta)^n \rightarrow 0$ cuando $n \rightarrow \infty$ y

$$\lim_{n \rightarrow \infty} P^n = \begin{pmatrix} \frac{\beta}{\alpha + \beta} & \frac{\alpha}{\alpha + \beta} \\ \frac{\beta}{\alpha + \beta} & \frac{\alpha}{\alpha + \beta} \end{pmatrix}$$

▲

En otros casos podemos utilizar el siguiente resultado de álgebra lineal, que enunciamos sin demostración.

Teorema 2.2 *Sea A una matriz de transición de $n \times n$ con n valores propios distintos $\lambda_1, \lambda_2, \dots, \lambda_n$. Sean v_1, v_2, \dots, v_n vectores columna de \mathbb{R}^n tales que v_i es un vector propio correspondiente a λ_i para $i = 1, 2, \dots, n$. Sea C la matriz $n \times n$ que tiene a v_i como i -ésimo vector columna. Entonces C es invertible y $C^{-1}AC = D$, con D la matriz cuyas entradas están dadas por $d_{i,i} = \lambda_i$ y $d_{i,j} = 0$ si $i \neq j$, $i, j \in \{1, 2, \dots, n\}$. Además, la k -ésima potencia de A , esta dada por*

$$A^{(k)} = CD^{(k)}C^{-1},$$

y las entradas de $D^{(k)}$ están dadas por $d_{i,j}^{(k)} = d_{i,j}^k$ para $i, j \in \{1, 2, \dots, n\}$.

Para cualquier vector $x \in \mathbb{R}^n$ se tiene que

$$A^{(k)}x = r_1\lambda_1^k v_1 + r_2\lambda_2^k v_2 + \dots + r_n\lambda_n^k v_n,$$

donde $(r_i, i \in \{1, \dots, n\})$ están dados por

$$C^{-1}x = (r_1, r_2, \dots, r_n)^t$$

Ejemplo 2.10

Sea $X = \{X_n, n \geq 0\}$ una cadena con matriz de transición

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

El objetivo de este ejercicio es encontrar la forma general de la entrada $(1, 1)$ de la matriz P^n , es decir $P_{1,1}^{(n)}$, para toda $n \geq 0$. Usando el teorema 2.2 calculemos la ecuación característica de P .

$$\begin{aligned} 0 &= \det\{P - xI\} \\ &= -x(1/2 - x)^2 + 1/4 \\ &= -(x/4 - x^2 + x^3 - 1/4) \\ &= -(x - 1)(x^2 + 1/4) \\ &= -(x - 1)(x - i/2)(x + i/2). \end{aligned}$$

Los valores propios de P son todos distintos $\lambda_1 = 1, \lambda_2 = i/2, \lambda_3 = -i/2$. Se tiene entonces que:

$$P = C \begin{pmatrix} 1 & 0 & 0 \\ 0 & i/2 & 0 \\ 0 & 0 & -i/2 \end{pmatrix} C^{-1} \quad \text{y} \quad P^n = C \begin{pmatrix} 1 & 0 & 0 \\ 0 & (i/2)^n & 0 \\ 0 & 0 & (-i/2)^n \end{pmatrix} C^{-1},$$

y por lo tanto que para algunos números complejos a, b, c

$$P_{11}^{(n)} = a + b(i/2)^n + c(-i/2)^n.$$

Recordemos que $\pm i = e^{\pm i\pi/2}$ y

$$\left(\pm \frac{i}{2}\right)^n = \left(\frac{1}{2}\right)^n e^{\pm in\pi/2} = \left(\frac{1}{2}\right)^n (\cos(n\pi/2) \pm i \operatorname{sen}(n\pi/2)).$$

Usando esto podemos escribir a $P_{11}^{(n)}$ como

$$P_{11}^{(n)} = \alpha + \beta \left(\frac{1}{2}\right)^n \cos(n\pi/2) + \gamma \left(\frac{1}{2}\right)^n \operatorname{sen}(n\pi/2), \quad n \geq 0.$$

donde $\alpha = a, \beta = b + c, \gamma = i(b - c)$. Para terminar determinemos a α, β, γ , (los cuales serán a fortiori números reales dado que $P_{11}^{(n)}$ es un número real). Basta con resolver el siguiente sistema de ecuaciones:

$$\begin{aligned} 1 &= P_{11}^{(0)} = \alpha + \beta \\ 0 &= P_{11}^{(1)} = \alpha + \frac{1}{2}\gamma \\ 0 &= P_{11}^{(2)} = \alpha - \frac{1}{4}\beta, \end{aligned}$$

para ver que $\alpha = 1/5, \beta = 4/5, \gamma = -2/5$. De donde vemos que

$$P_{11}^{(n)} = \frac{1}{5} + \frac{4}{5} \left(\frac{1}{2}\right)^n \cos(n\pi/2) - \frac{2}{5} \left(\frac{1}{2}\right)^n \operatorname{sen}(n\pi/2), \quad n \geq 0.$$

El mismo método puede ser empleado para calcular el resto de las entradas de la matriz P^n . ▲

2.4. Clasificación de los Estados

Definición 2.2 Sea \mathcal{E} el espacio de estados de una cadena de Markov y $A \subset \mathcal{E}$. El *tiempo de llegada* a A se define como

$$T_A = \min\{n \geq 1 : X_n \in A\} \tag{2.14}$$

si $X_n \in A$ para algún n y $T_A = \infty$ si $X_n \notin A$ para todo $n > 0$. Es decir, es el primer instante luego del inicio de la cadena, en el que la cadena visita al conjunto A . Si $A = \{a\}$ para algún $a \in \mathcal{E}$ escribimos T_a .

Una relación importante asociada a los tiempos de llegada es la siguiente:

$$P_{ij}^{(n)} = \sum_{m=1}^n P_i(T_j = m)P_{jj}^{(n-m)}, \quad n \geq 1. \quad (2.15)$$

Veamos cómo se demuestra esta relación. Descomponemos el evento de interés según el instante de la primera visita al estado j :

$$\begin{aligned} P_{ij}^{(n)} &= \sum_{m=1}^n P_i(X_n = j | T_j = m) P_i(T_j = m) \\ &= \sum_{m=1}^n P(X_n = j | X_m = j, X_m \neq j, \dots, X_1 \neq j, X_0 = i) P_i(T_j = m) \\ &= \sum_{m=1}^n P(X_n = j | X_m = j) P_i(T_j = m) \\ &= \sum_{m=1}^n P_i(T_j = m) P_{jj}^{(n-m)}. \end{aligned}$$

Observamos que

$$P_i(T_j = 1) = P_i(X_1 = j) = P_{ij}$$

y además

$$P_i(T_j = 2) = \sum_{k \neq j} P_i(X_1 = k, X_2 = j) = \sum_{k \neq j} P_{ik} P_{kj}.$$

Para valores mayores de n tenemos

$$P_i(T_j = n + 1) = \sum_{k \neq j} P_{ik} P_k(T_j = n), \quad n \geq 1. \quad (2.16)$$

Definición 2.3 Definimos

$$\rho_{ij} = P_i(T_j < \infty) = P(T_j < \infty | X_0 = i), \quad (2.17)$$

la probabilidad de que una cadena que comienza en i visite el estado j . En particular, ρ_{jj} es la probabilidad de que una cadena que comienza en j , regrese a j .

Observamos que

$$\rho_{ij} = P_i(T_j < \infty) = \sum_{m=1}^{\infty} P_i(T_j = m). \quad (2.18)$$

Definición 2.4 Decimos que un estado j es *recurrente* si $\rho_{jj} = 1$ y *transitorio* si $\rho_{jj} < 1$.

Si j es recurrente y la cadena comienza en j , entonces regresa a j con probabilidad 1. Si, en cambio, j es transitorio, hay una probabilidad positiva e igual a $1 - \rho_{jj}$ de que si la cadena comienza en j , nunca regrese a ese estado. Si j es un estado absorbente, $P_j(T_j = 1) = 1$ y por lo tanto $\rho_{jj} = 1$, de modo que un estado absorbente es necesariamente recurrente.

Ejemplo 2.11 (Paseo al Azar con N=4)

La matriz de transición es

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

Los estados 0 y 4 son absorbentes, y por lo tanto son recurrentes. Veamos que los otros estados, 1, 2 y 3, son transitorios.

Si estamos en 1 y la cadena pasa a 0, nunca regresará a 1, de modo que la probabilidad de nunca regresar a 1 es

$$P_1(T_1 = \infty) = P(T_1 = \infty | X_0 = 1) \geq P_{10} = q > 0.$$

De manera similar, comenzando en 2, la cadena puede ir a 1 y luego a 0, de modo que

$$P_2(T_2 = \infty) = P(T_2 = \infty | X_0 = 2) \geq P_{21}P_{10} = q^2 > 0.$$

Finalmente, si comenzamos en 3 observamos que la cadena puede ir inmediatamente a 4 y no regresar nunca con probabilidad 0.4:

$$P_3(T_3 = \infty) = P(T_3 = \infty | X_0 = 3) \geq P_{34} = p > 0.$$

▲

Sea $\mathbf{1}_j(x)$ la función indicadora del estado j , definida por

$$\mathbf{1}_j(x) = \begin{cases} 1, & \text{si } x = j, \\ 0, & \text{si } x \neq j. \end{cases}$$

Sea $N(j)$ el número de veces que la cadena visita el estado j :

$$N(j) = \sum_{n=1}^{\infty} \mathbf{1}_j(X_n). \quad (2.19)$$

Como el evento $\{N(j) \geq 1\}$ equivale al evento $\{T_j < \infty\}$, tenemos que

$$P_i(N(j) \geq 1) = P_i(T_j < \infty) = \rho_{ij}. \quad (2.20)$$

Proposición 2.5 *La probabilidad de que una cadena que comienza en i visite el estado j por primera vez en el instante m y que la próxima visita ocurra n unidades de tiempo después es*

$$P_i(T_j = m)P_j(T_j = n). \quad (2.21)$$

Demostración. Tenemos

$$\begin{aligned} & P(X_{n+m} = j, X_{n+m-1} \neq j, \dots, X_{m+1} \neq j, X_m = j, X_{m-1} \neq j, \dots, X_1 \neq j | X_0 = i) \\ &= P(X_{n+m} = j, X_{n+m-1} \neq j, \dots, X_{m+1} \neq j | X_m = j, X_{m-1} \neq j, \dots, X_1 \neq j, X_0 = i) \\ &\quad \times P(X_m = j, X_{m-1} \neq j, \dots, X_1 \neq j, | X_0 = i) \\ &= P(X_{n+m} = j, X_{n+m-1} \neq j, \dots, X_{m+1} \neq j | X_m = j)P(X_m = j, X_{m-1} \neq j, \dots, X_1 \neq j | X_0 = i) \\ &= P_j(T_j = n)P_i(T_j = m). \end{aligned}$$

■

Usando (2.21) y (2.18)

$$\begin{aligned} P_i(N(j) \geq 2) &= \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} P_i(T_j = m)P_j(T_j = n) \\ &= \left(\sum_{m=1}^{\infty} P_i(T_j = m) \right) \left(\sum_{n=1}^{\infty} P_j(T_j = n) \right) \\ &= \rho_{ij}\rho_{jj}. \end{aligned}$$

De manera similar se demuestra que

$$P_i(N(j) \geq m) = \rho_{ij}\rho_{jj}^{m-1}, \quad m \geq 1. \quad (2.22)$$

Como

$$P_i(N(j) = m) = P_i(N(j) \geq m) - P_i(N(j) \geq m+1)$$

obtenemos que

$$P_i(N(j) = m) = \rho_{ij}\rho_{jj}^{m-1}(1 - \rho_{jj}), \quad m \geq 1, \quad (2.23)$$

y además

$$P_i(N(j) = 0) = 1 - P_i(N(j) \geq 1) = 1 - \rho_{ij}. \quad (2.24)$$

Recordemos que la notación $E_i[X]$ indica la esperanza de la variable aleatoria X dado que la cadena de Markov comienza en i . Entonces

$$E_i[\mathbf{1}_j(X(n))] = P_i(X_n = j) = P_{ij}^{(n)}. \quad (2.25)$$

Obtenemos a partir de (2.19) y (2.25) que

$$E_i[N(j)] = E_i\left[\sum_{n=1}^{\infty} \mathbf{1}_j(X_n)\right] = \sum_{n=1}^{\infty} E_i[\mathbf{1}_j(X_n)] = \sum_{n=1}^{\infty} P_{ij}^{(n)}.$$

En la expresión anterior podemos intercambiar la esperanza y la serie usando el Teorema de Tonelli ya que los sumandos son todos no-negativos. Llamemos

$$G(i, j) = E_i[N(j)] = \sum_{n=1}^{\infty} P_{ij}^{(n)}, \quad (2.26)$$

$G(i, j)$ denota el valor esperado del número de visitas a j de una cadena de Markov que comienza en i .

Teorema 2.3 a) Sea j un estado transitorio, entonces $P_i(N(j) < \infty) = 1$ y

$$G(i, j) = \frac{\rho_{ij}}{1 - \rho_{jj}}, \quad i \in \mathcal{E}, \quad (2.27)$$

que es finita para todo $i \in \mathcal{E}$.

b) Sea j un estado recurrente, entonces $P_j(N(j) = \infty) = 1$ y $G(j, j) = \infty$. Además

$$P_i(N(j) = \infty) = P_i(T_j < \infty) = \rho_{ij}, \quad i \in \mathcal{E}. \quad (2.28)$$

Si $\rho_{ij} = 0$ entonces $G(i, j) = 0$ mientras que si $\rho_{ij} > 0$, $G(i, j) = \infty$.

Demostración. a) Sea j un estado transitorio, como $0 \leq \rho_{jj} < 1$, sigue de (2.22) que

$$P_i(N(j) = \infty) = \lim_{m \rightarrow \infty} P_i(N(j) \geq m) = \lim_{m \rightarrow \infty} \rho_{ij}\rho_{jj}^{m-1} = 0.$$

Usando ahora (2.23)

$$G(i, j) = E_i[N(j)] = \sum_{m=1}^{\infty} mP_i(N(j) = m) = \sum_{m=1}^{\infty} m\rho_{ij}\rho_{jj}^{m-1}(1 - \rho_{jj}).$$

Por otro lado, tenemos el siguiente resultado para series de potencias

$$\sum_{m=1}^{\infty} mt^{m-1} = \frac{1}{(1-t)^2}, \quad \text{para } |t| < 1,$$

y usandolo obtenemos que

$$G(i, j) = \frac{\rho_{ij}(1 - \rho_{jj})}{(1 - \rho_{jj})^2} = \frac{\rho_{ij}}{1 - \rho_{jj}} < \infty.$$

b) Supongamos que j es recurrente, entonces $\rho_{jj} = 1$ y de (2.22) sigue que

$$P_i(N(j) = \infty) = \lim_{m \rightarrow \infty} P_i(N(j) \geq m) = \lim_{m \rightarrow \infty} \rho_{ij} = \rho_{ij}$$

y en particular, $P_j(N(j) = \infty) = 1$. Si una v.a. no negativa tiene probabilidad positiva de ser infinita, su valor esperado es infinito. Por lo tanto

$$G(j, j) = E_j[N(j)] = \infty$$

Si $\rho_{ij} = 0$ entonces $P_i(T_j = m) = 0$ para todo $m \in \mathbb{N}$ y por (2.15) obtenemos que $P_{ij}^n = 0$, $n \geq 1$. Por lo tanto $G(i, j) = 0$ en este caso. Si $\rho_{ij} > 0$ entonces $P_i(N(j) = \infty) = \rho_{ij} > 0$ y en consecuencia

$$G(i, j) = E_i[N(j)] = \infty.$$

■

Observación 2.1

1. Sea j un estado transitorio, como $\sum_{n=1}^{\infty} P_{ij}^{(n)} = G(i, j) < \infty$, $i \in \mathcal{E}$, vemos que

$$\lim_{n \rightarrow \infty} P_{ij}^{(n)} = 0, \quad i \in \mathcal{E}. \quad (2.29)$$

2. Una cadena de Markov con espacio de estados finito debe tener al menos un estado recurrente: Si \mathcal{E} es finito y todos los estados son transitorios, por (2.29),

$$0 = \sum_{j \in \mathcal{E}} \lim_{n \rightarrow \infty} P_{ij}^{(n)} = \lim_{n \rightarrow \infty} \sum_{j \in \mathcal{E}} P_{ij}^{(n)} = \lim_{n \rightarrow \infty} P_i(X_n \in \mathcal{E}) = 1,$$

lo cual es una contradicción.

2.5. Descomposición del Espacio de Estados

Decimos que desde el estado i se llega o se accede al estado j si $P_{ij}^{(n)} > 0$ para algún $n \geq 0$. Es fácil ver que esto es cierto para $i \neq j$ sí y sólo sí $\rho_{ij} > 0$. Por lo tanto desde i se accede a j si hay una probabilidad positiva de que en un número finito de pasos, se pueda llegar al estado j partiendo del estado i . Notación: $i \rightarrow j$.

Si $i \rightarrow j$ y $j \rightarrow i$ decimos que los estados *se comunican* y escribimos $i \leftrightarrow j$. Si dos estados no se comunican, o bien $P_{ij}^{(n)} = 0$, $\forall n \geq 0$, o bien $P_{ji}^{(n)} = 0$, $\forall n \geq 0$, o ambos. La comunicación es una relación de equivalencia:

a) $i \leftrightarrow i : P_{ij}^{(0)} = \delta_{ij}$.

b) $i \leftrightarrow j \Leftrightarrow j \leftrightarrow i$

c) Si $i \leftrightarrow j$ y $j \leftrightarrow k$ entonces $i \leftrightarrow k$:

$$i \leftrightarrow j \Rightarrow \exists n \text{ tal que } P_{ij}^{(n)} > 0, \quad j \leftrightarrow k \Rightarrow \exists m \text{ tal que } P_{jk}^{(m)} > 0,$$

y usando ahora las ecuaciones de Chapman-Kolmogorov

$$P_{ik}^{(n+m)} = \sum_r P_{ir}^{(n)} P_{rk}^{(m)} \geq P_{ij}^{(n)} P_{jk}^{(m)} > 0.$$

Un argumento similar muestra que existe s tal que $P_{ki}^{(s)} > 0$.

Esta relación de equivalencia divide el espacio de estados \mathcal{E} en clases de equivalencia que también llamaremos *clases de comunicación*. Los estados de una clase de equivalencia son aquellos que se comunican entre sí.

Puede ocurrir que partiendo de una clase de equivalencia, la cadena entre en otra. Si esto ocurre, claramente la cadena no puede regresar a la clase inicial, pues si lo hiciera las dos clases se comunicarían y formarían una sola clase.

Definición 2.5 Decimos que la cadena es *irreducible* si hay una sola clase de equivalencia, es decir, si todos los estados se comunican entre sí.

Teorema 2.4 (a) Si $i \rightarrow j$ pero $j \not\rightarrow i$ entonces i es transitorio.

(b) Si i es recurrente e $i \rightarrow j$ entonces j es recurrente y $\rho_{ij} = \rho_{ji} = 1$.

Demostración. Supongamos que $i \rightarrow j$ y sea

$$\kappa = \min\{k : P_{ij}^{(k)} > 0\} \tag{2.30}$$

el menor número de pasos necesarios para ir de i a j . Como $P_{ij}^{(\kappa)} > 0$ existe una sucesión de estados $j_1, j_2, \dots, j_{\kappa-1}$ tal que

$$P_{ij_1} P_{j_1 j_2} \cdots P_{j_{\kappa-1} j} > 0$$

Como κ es el mínimo, todos los $j_r \neq i$, $1 \leq r < \kappa$, pues en caso contrario habría una sucesión más corta, y tenemos

$$P_i(T_i = \infty) \geq P_{ij_1} P_{j_1 j_2} \cdots P_{j_{\kappa-1} j} (1 - \rho_{ji}). \tag{2.31}$$

Si $j \not\rightarrow i$ tenemos $\rho_{ji} = 0$ y por lo tanto $P_i(T_i = \infty) > 0$, es decir $\rho_{ii} = P_i(T_i < \infty) < 1$, lo que implica que i es transitorio. Esto demuestra (a).

Supongamos ahora que i es recurrente, entonces el lado izquierdo de (2.31) es 0, de modo que si $\rho_{ji} < 1$ tendríamos una contradicción. Por lo tanto $\rho_{ji} = 1$ y $j \rightarrow i$. Para completar la demostración de (b) falta ver que j es recurrente y que $\rho_{ij} = 1$.

Como $\rho_{ji} = 1$ existe un entero positivo N tal que $P_{ji}^{(N)} > 0$ y tenemos

$$\begin{aligned} P_{jj}^{(N+n+\kappa)} &= P_j(X_{N+n+\kappa} = j) \\ &\geq P_j(X_N = i, X_{N+n} = i, X_{N+n+\kappa} = j) \\ &= P_{ji}^{(N)} P_{ii}^{(n)} P_{ij}^{(\kappa)}. \end{aligned}$$

Por lo tanto

$$\begin{aligned}
 G(j, j) &= \sum_{m=1}^{\infty} P_{jj}^{(m)} \geq \sum_{m=N+n+\kappa}^{\infty} P_{jj}^{(m)} \\
 &= \sum_{n=1}^{\infty} P_{jj}^{(N+n+\kappa)} \geq P_{ji}^{(N)} P_{ij}^{(\kappa)} \sum_{n=1}^{\infty} P_{ii}^{(n)} \\
 &= P_{ji}^{(N)} P_{ij}^{\kappa} G(i, i) = \infty
 \end{aligned}$$

y vemos que j también es recurrente.

Ahora, como j es recurrente y $j \rightarrow i$, por lo que hemos demostrado vemos que $\rho_{ij} = 1$. ■

Corolario 2.1 *Sea C una clase de comunicación. Entonces todos los elementos de C son recurrentes o todos son transitorios.*

Ejemplo 2.12

Consideremos una cadena con matriz de transición

$$P = \begin{pmatrix} 0.3 & 0.7 & 0 \\ 0.2 & 0.4 & 0.4 \\ 0.1 & 0.6 & 0.3 \end{pmatrix}$$

Veamos que todos los estados son recurrentes. En primer lugar observemos que no importa en cual estado estemos, siempre hay una probabilidad positiva de ir al estado 1 en el paso siguiente. Esta probabilidad es de, al menos, 0.1 y en consecuencia la probabilidad de no visitar el estado 1 en el paso siguiente es, a lo sumo, 0.9. Por lo tanto, la probabilidad de no visitar el estado 1 en los próximos n pasos es, a lo sumo, $(0.9)^n$ y obtenemos

$$P_1(T_1 > n) = P(T_1 > n | X_0 = 1) \leq (0.9)^n \rightarrow 0, \quad (n \rightarrow \infty).$$

Por lo tanto

$$\begin{aligned}
 P_1(T_1 < \infty) &= P_1(\cup_{n=1}^{\infty} \{T_1 = n\}) = 1 - P_1((\cup_{n=1}^{\infty} \{T_1 = n\})^c) \\
 &= 1 - P_1(\cap_{n=1}^{\infty} \{T_1 = n\}^c) \\
 &= 1 - \lim_{k \rightarrow \infty} P_1(\cap_{n=1}^k \{T_1 = n\}^c) \\
 &= 1 - \lim_{k \rightarrow \infty} P_1(T_1 > k) = 1.
 \end{aligned}$$

Un argumento similar funciona para el estado 2, sólo que ahora la probabilidad de hacer una transición desde cualquier estado al 2 es, al menos, 0.4.

Este argumento no funciona para el estado 3 ya que $P_{13} = 0$. Sin embargo, si efectuamos el producto de P por si misma, P^2 , obtenemos las probabilidades de transición a dos pasos:

$$P^2 = \begin{pmatrix} 0.23 & 0.49 & 0.28 \\ 0.18 & 0.54 & 0.28 \\ 0.18 & 0.49 & 0.33 \end{pmatrix}$$

y vemos que para cualquier j , la probabilidad de pasar de j a 3 en dos pasos es mayor o igual a 0.28:

$$P_j(T_3 = 2) = P(X_{n+2} = 3 | X_n = j) \geq 0.28.$$

Si consideramos la cadena en los instantes pares $2, 4, \dots, 2k, \dots$ obtenemos que la probabilidad de no visitar el estado 3 antes del instante $2k$ es

$$P_3(T_3 > 2k) \leq (0.72)^k \rightarrow 0 \quad (k \rightarrow \infty)$$

de modo que el estado 3 también es recurrente. ▲

Ejemplo 2.13

Consideremos una cadena de Markov con $\mathcal{E} = \{1, 2, 3, 4, 5, 6, 7\}$ y la siguiente matriz de transición

$$\begin{pmatrix} 0.3 & 0 & 0 & 0 & 0.7 & 0 & 0 \\ 0.1 & 0.2 & 0.3 & 0.4 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0 & 0.5 & 0 \\ 0.6 & 0 & 0 & 0 & 0.4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.2 & 0.8 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

Las transiciones posibles entre estados diferentes se presentan en la figura 2.1. Una gráfica de este tipo se conoce como la gráfica de transiciones de la cadena.

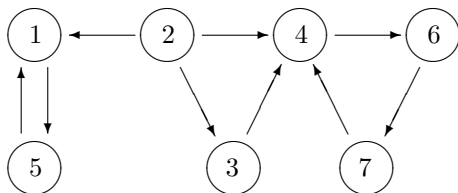


Figura 2.1

Observamos que $2 \rightarrow 4$ pero $4 \not\rightarrow 2$ y algo similar ocurre con 3, de modo que 2 y 3 son estados transitorios. Sin embargo, estos dos estados no se comunican y por lo tanto forman dos clases de equivalencia disjuntas, $\{2\}$ y $\{3\}$. El resto de los estados se separan en dos clases de equivalencia, $\{1, 5\}$ y $\{4, 6, 7\}$. Veremos luego que ambas son recurrentes. ▲

Definición 2.6 Un conjunto no vacío $C \subset \mathcal{E}$ es *cerrado* si no se puede salir de él, es decir, desde ningún estado de C se tiene acceso a estados que estén fuera de C . Esto quiere decir que

$$p_{ij} = 0 \quad \text{si } i \in C, j \notin C.$$

Equivalentemente, C es cerrado sí y sólo sí

$$P_{ij}^{(n)} = 0, \quad i \in C, j \notin C \text{ y } n \geq 1.$$

Si C es cerrado y la cadena comienza en C entonces, con probabilidad 1 se queda en C todo el tiempo. Si a es un estado absorbente entonces $\{a\}$ es cerrado.

Definición 2.7 Un conjunto cerrado es *irreducible* si todos sus estados se comunican.

Ejemplo 2.14

En el ejemplo 2.13 los conjuntos $\{1, 5\}$, $\{1, 4, 5, 6, 7\}$ y $\{1, 3, 4, 5, 6, 7\}$ son cerrados, pero sólo el primero es irreducible.

De los resultados anteriores vemos que si C es cerrado e irreducible entonces todos los estados de C son recurrentes o todos son transitorios. El siguiente resultado es consecuencia del teorema 2.4.

Corolario 2.2 *Sea C un conjunto cerrado e irreducible de estados recurrentes. Entonces, para cualesquiera $i, j \in C$, $\rho_{ij} = 1$, $P_i(N(j) = \infty) = 1$ y $G(i, j) = \infty$.*

Una cadena de Markov irreducible es una cadena en la cual cada estado se comunica consigo mismo y con cualquier otro estado. En una cadena de este tipo o bien todos los estados son recurrentes o bien todos son transitorios.

Vimos anteriormente que si \mathcal{E} es finito, tiene al menos un estado recurrente. El mismo argumento muestra que cualquier conjunto cerrado finito C tiene al menos un estado recurrente. Por lo tanto, todos los estados de C lo son:

Teorema 2.5 *Sea C un conjunto finito, irreducible y cerrado. Entonces todos los estados de C son recurrentes.*

Consideremos una cadena de Markov con espacio de estados finito. Por el teorema 2.5 si la cadena es irreducible entonces debe ser recurrente. Si la cadena no es irreducible, podemos usar el teorema 2.4 para determinar cuáles estados son recurrentes y cuáles transitorios.

Ejemplo 2.15

Determine cuáles estados son recurrentes y cuáles transitorios para la cadena de Markov con la siguiente matriz de transición.

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1/4 & 1/2 & 1/4 & 0 & 0 & 0 \\ 0 & 1/5 & 2/5 & 1/5 & 0 & 1/5 \\ 0 & 0 & 0 & 1/6 & 1/3 & 1/2 \\ 0 & 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1/4 & 0 & 3/4 \end{pmatrix}$$

La siguiente gráfica presenta las transiciones posibles (en un paso) entre estados diferentes para esta cadena

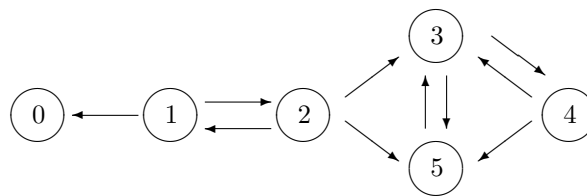


Figura 2.2

Vemos que hay tres clases de equivalencia $\{0\}$; $\{1, 2\}$ y $\{3, 4, 5\}$. La primera clase es recurrente porque 0 es un estado absorbente. La clase $\{1, 2\}$ es transitoria porque es posible salir de ella y no regresar nunca, por ejemplo, pasando de 1 a 0. Finalmente, la tercera clase es recurrente porque es finita, cerrada e irreducible. ▲

Llamemos \mathcal{E}_T a la colección de los estados transitorios de \mathcal{E} y \mathcal{E}_R a la colección de estados recurrentes. En el ejemplo anterior $\mathcal{E}_T = \{1, 2\}$ y $\mathcal{E}_R = \{0, 3, 4, 5\}$. Esta última clase puede descomponerse en dos conjuntos cerrados irreducibles, $C_1 = \{0\}$ y $C_2 = \{3, 4, 5\}$.

Teorema 2.6 *Supongamos que el conjunto \mathcal{E}_R de estados recurrentes no es vacío. Entonces es la unión de una colección finita o numerable de conjuntos cerrados, disjuntos e irreducibles C_1, C_2, \dots .*

Demostración. Escogemos $i \in \mathcal{E}_R$ y sea C el conjunto de todos los estados $j \in \mathcal{E}_R$ tales que $i \rightarrow j$. Como i es recurrente, $\rho_{ii} = 1$ y por lo tanto $i \in C$. Veamos ahora que C es un conjunto cerrado e irreducible.

Supongamos que $j \in C$ y $j \rightarrow k$. Como j es recurrente, por el teorema 2.4 sabemos que k también es recurrente. Por transitividad $i \rightarrow k$ y por lo tanto $k \in C$. Esto muestra que C es cerrado.

Supongamos ahora que j y k están en C . Como i es recurrente y $i \rightarrow j$, por el teorema 2.4 vemos que $j \rightarrow i$. Como $j \rightarrow i \rightarrow k$ vemos que $j \rightarrow k$, de modo que C es irreducible.

Para completar la demostración necesitamos ver que si C y D son dos conjuntos cerrados irreducibles de \mathcal{E}_R , o bien son disjuntos o bien son idénticos. Supongamos que no son disjuntos y sea $i \in C \cap D$. Escojamos $j \in C$ entonces $i \rightarrow j$ porque $i \in C$ y C es irreducible. Como D es cerrado, $i \in D$, $i \rightarrow j$ entonces $j \in D$. Por lo tanto, todo estado de C también está en D . El recíproco también es cierto, de modo que C y D son idénticos. ■

2.6. Estudio de las Transiciones Iniciales

Comenzamos esta sección con un ejemplo para presentar las ideas básicas de esta técnica.

Ejemplo 2.16

Consideremos la cadena de Markov con espacio de estados $\mathcal{E} = \{0, 1, 2\}$ y la siguiente matriz de transición

$$\begin{pmatrix} 1 & 0 & 0 \\ \alpha & \beta & \gamma \\ 0 & 0 & 1 \end{pmatrix}$$

con $\alpha > 0$, $\beta > 0$, $\gamma > 0$, $\alpha + \beta + \gamma = 1$. Si la cadena comienza en 1, permanece en este estado por un tiempo aleatorio y luego pasa al estado 0 o al estado 2, donde se queda para siempre. Nos hacemos las siguientes preguntas:

1. ¿En cuál de los dos estados, 0 ó 2, se queda la cadena?
2. ¿Cuánto tiempo toma, en promedio, alcanzar uno de estos estados?

Sea $A = \{0, 2\}$ el conjunto de los estados absorbentes y llamemos H_A el tiempo que transcurre hasta que la cadena es absorbida por 0 o por 2:

$$H_A = \min\{n \geq 0 : X_n = 0 \text{ ó } X_n = 2\}.$$

La diferencia entre H_A y T_A , que definimos anteriormente está en que H incluye al estado inicial.

Para responder las preguntas anteriores debemos hallar

$$\begin{aligned} u &= P(X_{H_A} = 0 | X_0 = 1), \\ \nu &= E[H_A | X_0 = 1]. \end{aligned}$$

Para hacer el análisis de la primera transición consideramos por separado lo que puede ocurrir en el primer paso:

$$X_1 = 0, \quad X_1 = 1 \quad \text{ó} \quad X_1 = 2,$$

con probabilidades respectivas α, β y γ . Consideremos u ,

- Si $X_1 = 0$ entonces $H_A = 1$ y $X_{H_A} = 0$. Esto ocurre con probabilidad α .
- Si $X_1 = 2$ entonces $H_A = 1$ y $X_{H_A} = 2$. Esto ocurre con probabilidad γ .

• Si $X_1 = 1$, regresamos a las condiciones iniciales y esto ocurre con probabilidad β .
Tenemos, además,

$$P(X_{H_A} = 0|X_1 = 0) = 1; \quad P(X_{H_A} = 0|X_1 = 2) = 0; \quad P(X_{H_A} = 0|X_1 = 1) = u,$$

en consecuencia

$$\begin{aligned} u &= P(X_{H_A} = 0|X_0 = 1) \\ &= \sum_{k=0}^2 P(X_{H_A} = 0|X_0 = 1, X_1 = k)P(X_1 = k|X_0 = 1) \\ &= \sum_{k=0}^2 P(X_{H_A} = 0|X_1 = k)P(X_1 = k|X_0 = 1) \\ &= 1 \cdot \alpha + u\beta + 0 \cdot \gamma = \alpha + u\beta \end{aligned}$$

y obtenemos

$$u = \frac{\alpha}{1 - \beta} = \frac{\alpha}{\alpha + \gamma}$$

que es la probabilidad condicional de ir a 0 dado que el proceso comienza en 1 y termina en A .

Regresemos a la determinación del tiempo medio hasta absorción $H_A \geq 1$. Si $X_1 = 0$ ó $X_1 = 2$, no hacen falta más pasos. Si, en cambio, $X_1 = 1$, el proceso se encuentra en la misma situación del comienzo y necesita en promedio, $\nu = E[H_A|X_0 = 1]$ pasos adicionales para ser absorbido. Tenemos entonces

$$\nu = 1 + \alpha \cdot 0 + \beta \cdot \nu + \gamma \cdot 0 = 1 + \beta\nu$$

de donde

$$\nu = \frac{1}{1 - \beta}.$$

En el ejemplo que estamos estudiando es posible hacer un cálculo directo. Observemos que

$$P(H_A > k|X_0 = 1) = \beta^k \quad \text{para } k = 0, 1, \dots$$

y por lo tanto

$$E[H_A|X_0 = 1] = \sum_{k=0}^{\infty} P(H_A > k|X_0 = 1) = \frac{1}{1 - \beta}.$$

▲

Observación 2.2 Para calcular el valor esperado hemos usado la relación

$$E[X] = \sum_{k=0}^{\infty} P(X > k)$$

que puede demostrarse como sigue,

$$\begin{aligned}
 E[X] &= \sum_{k \geq 0} k p_k \\
 &= p_1 + 2p_2 + 3p_3 + 4p_4 + \cdots \\
 &= p_1 + p_2 + p_3 + p_4 + \cdots \\
 &\quad + p_2 + p_3 + p_4 + \cdots \\
 &\quad \quad + p_3 + p_4 + \cdots \\
 &\quad \quad \quad + p_4 + \cdots \\
 &\quad \quad \quad \quad \vdots \\
 &= P(X \geq 1) + P(X \geq 2) + P(X \geq 3) + \cdots \\
 &= \sum_{k=1}^{\infty} P(X \geq k) = \sum_{k=0}^{\infty} P(X > k).
 \end{aligned}$$

Ejemplo 2.17

Consideremos ahora una cadena con 4 estados y matriz de transición

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ P_{10} & P_{11} & P_{12} & P_{13} \\ P_{20} & P_{21} & P_{22} & P_{23} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Vemos que 0 y 3 son absorbentes mientras que 1 y 2 son transitorios. La probabilidad de que la cadena sea absorbida en el estado 0, por ejemplo, depende ahora del estado transitorio en el cual comenzó la cadena.

Modificamos las definiciones del ejemplo anterior de la siguiente manera

$$\begin{aligned}
 H_A &= \min\{n \geq 0 : X_n = 0 \text{ ó } X_n = 3\}, \\
 u_i &= P(X_{H_A} = 0 | X_0 = i) \quad i = 1, 2, \\
 \nu_i &= E[H_A | X_0 = i], \quad i = 1, 2.
 \end{aligned}$$

Podemos extender las definiciones de u_i y ν_i poniendo $u_0 = 1$, $u_3 = 0$, $\nu_0 = \nu_3 = 0$.

Para el análisis de la primera transición tenemos que considerar los dos posibles estados iniciales $X_0 = 1$ y $X_0 = 2$ separadamente. Si $X_0 = 1$, en el primer paso podemos tener

$$\begin{aligned}
 u_1 &= P(X_{H_A} = 0 | X_0 = 1) \\
 &= \sum_{k=0}^3 P(X_{H_A} = 0 | X_0 = 1, X_1 = k) P(X_1 = k | X_0 = 1) \\
 &= \sum_{k=0}^3 P(X_{H_A} = 0 | X_1 = k) P(X_1 = k | X_0 = 1) \\
 &= P_{10} + u_1 P_{11} + u_2 P_{12}.
 \end{aligned}$$

De manera similar

$$u_2 = P_{20} + u_1 P_{21} + u_2 P_{22}$$

y hay que resolver este sistema de ecuaciones. Veamos un ejemplo concreto, supongamos que

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0.4 & 0.3 & 0.2 & 0.1 \\ 0.1 & 0.3 & 0.3 & 0.3 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Las ecuaciones son

$$\left. \begin{aligned} u_1 &= 0.4 + 0.3u_1 + 0.2u_2 \\ u_2 &= 0.1 + 0.3u_1 + 0.3u_2 \end{aligned} \right\}$$

con soluciones

$$u_1 = \frac{30}{43}, \quad u_2 = \frac{19}{43}.$$

Así, si comenzamos en 2 hay probabilidad $19/43$ de ser absorbido por 0 y $1-19/43=24/43$ de ser absorbido por 3.

El tiempo promedio hasta la absorción también depende del estado de inicio. Las ecuaciones son

$$\left. \begin{aligned} \nu_1 &= 1 + P_{11}\nu_1 + P_{12}\nu_2 \\ \nu_2 &= 1 + P_{21}\nu_1 + P_{22}\nu_2 \end{aligned} \right\}$$

En el ejemplo concreto que estudiamos tenemos

$$\left. \begin{aligned} \nu_1 &= 1 + 0.3\nu_1 + 0.2\nu_2 \\ \nu_2 &= 1 + 0.3\nu_1 + 0.3\nu_2 \end{aligned} \right\}$$

con soluciones $\nu_1 = 90/43$ y $\nu_2 = 100/43$. ▲

En general sea C un conjunto cerrado e irreducible de estados recurrentes y sea

$$\rho_C(i) = P_i(H_C < \infty)$$

la probabilidad de que una cadena de Markov que comienza en i llegue en algún momento a C . Decimos que $\rho_C(i)$ es la probabilidad de que una cadena que comienza en i sea absorbida por el conjunto C .

Es claro que $\rho_C(i) = 1$ si $i \in C$ y $\rho_C(i) = 0$ si i es recurrente y no está en C . Queremos calcular esta probabilidad cuando i es un estado transitorio.

Si sólo hay un número finito de estados transitorios, y en particular si \mathcal{E}_T es finito, siempre es posible calcular $\rho_C(i)$, $i \in \mathcal{E}_T$, resolviendo un sistema de ecuaciones lineales con tantas ecuaciones como incógnitas. Observemos que si $i \in \mathcal{E}_T$, la cadena que comienza en i puede entrar a C sólo si entra en el instante 1 o si permanece en \mathcal{E}_T en 1 y entra a C en algún instante futuro:

$$\rho_C(i) = \sum_{j \in C} P_{ij} + \sum_{k \in \mathcal{E}_T} P_{ik} \rho_C(k), \quad i \in \mathcal{E}_T.$$

Teorema 2.7 *Supongamos que el conjunto de estados transitorios \mathcal{E}_T es finito y sea C un conjunto cerrado e irreducible de estados recurrentes. Entonces el sistema de ecuaciones*

$$f(i) = \sum_{j \in C} P_{ij} + \sum_{j \in \mathcal{E}_T} P_{ij} f(j), \quad i \in \mathcal{E}_T, \quad (2.32)$$

tiene una única solución $f(i) = \rho_C(i)$, $i \in \mathcal{E}_T$.

Demostración

Si (2.32) vale entonces para el estado $j \in \mathcal{E}$ también se tiene que

$$f(j) = \sum_{k \in C} P_{jk} + \sum_{k \in \mathcal{E}_T} P_{jk} f(k), \quad j \in \mathcal{E}_T. \quad (2.33)$$

Sustituyendo (2.33) en (2.32) obtenemos

$$f(i) = \sum_{j \in C} P_{ij} + \sum_{j \in \mathcal{E}_T} \sum_{k \in C} P_{ij} P_{jk} + \sum_{j \in \mathcal{E}_T} \sum_{k \in \mathcal{E}_T} P_{ij} P_{jk} f(k).$$

La suma de los dos primeros términos es $P_i(H_C \leq 2)$ y el tercero es

$$\sum_{k \in \mathcal{E}_T} P_{ik}^{(2)} f(k) = \sum_{j \in \mathcal{E}_T} P_{ij}^{(2)} f(j),$$

de modo que

$$f(i) = P_i(H_C \leq 2) + \sum_{j \in \mathcal{E}_T} P_{ij}^{(2)} f(j).$$

Repitiendo este argumento concluimos que para todo n ,

$$f(i) = P_i(H_C \leq n) + \sum_{j \in \mathcal{E}_T} P_{ij}^{(n)} f(j), \quad i \in \mathcal{E}_T. \quad (2.34)$$

Como los estados $j \in \mathcal{E}_T$ son transitorios,

$$\lim_{n \rightarrow \infty} P_{ij}^{(n)} = 0, \quad i \in \mathcal{E}, \quad j \in \mathcal{E}_T. \quad (2.35)$$

Por hipótesis, \mathcal{E}_T es finito y por lo tanto de (2.35) se sigue que la suma de (2.34) tiende a 0 cuando $n \rightarrow \infty$. En consecuencia, para $i \in \mathcal{E}_T$

$$f(i) = \lim_{n \rightarrow \infty} P_i(H_C \leq n) = P_i(H_C < \infty) = \rho_C(i).$$

■

2.7. Paseo al Azar

Consideramos de nuevo la situación del ejemplo 2.4. Tenemos una sucesión ξ_i , $i \geq 1$ de v.a.i.i.d. que toman valor 1 con probabilidad p y valor -1 con probabilidad $q = 1 - p$ y ponemos $S_n = S_0 + \sum_1^n \xi_i$. S_n es un paseo al azar simple. Vamos a estudiar en esta sección algunas propiedades importantes de este proceso.

Lema 2.1 *Un paseo al azar simple tiene las siguientes propiedades*

- *Propiedad de Markov*

$$P(S_{n+m} = j | S_0, S_1, \dots, S_n) = P(S_{n+m} = j | S_n).$$

- *Homogeneidad espacial*

$$P(S_n = j | S_0 = i) = P(S_n = j + k | S_0 = i + k).$$

- *Homogeneidad Temporal*

$$P(S_n = j | S_0 = i) = P(S_{n+m} = j | S_m = i).$$

Demostración. La propiedad de Markov ya fue demostrada.

- *Homogeneidad espacial.*

Se tiene que el lado izquierdo es igual a $P(\sum_{i=1}^n \xi_i = j - i)$ mientras que el lado derecho es igual a $P(\sum_{i=1}^n \xi_i = j + b - (i + b))$.

- *Homogeneidad temporal.*

Es una consecuencia fácil de la independencia y del hecho que las v. a. ξ_i son idénticamente distribuidas que el lado derecho es igual a:

$$\frac{P\left(S_0 + \sum_{i=1}^{n+m} \xi_i = j, S_0 + \sum_{i=1}^m \xi_i = i\right)}{P\left(S_0 + \sum_{i=1}^m \xi_i = i\right)} = P\left(\sum_{i=m+1}^{m+n} \xi_i = j - i\right) = P\left(\sum_{i=1}^n \xi_i = j - i\right);$$

un cálculo elemental prueba que el lado izquierdo es idéntico a esta cantidad. ■

Proposición 2.6 *Se tiene que para a, b enteros, $n \geq 0$ y $|b - a| \leq n$*

$$P(S_n = b | S_0 = a) = \begin{cases} \binom{n}{(n+b-a)/2} p^{(n+b-a)/2} q^{(n-b+a)/2} & \text{si } (n+b-a)/2 \in \mathbb{Z} \\ 0 & \text{en otro caso.} \end{cases}$$

Demostración. Se tiene que una trayectoria que lleva del punto $(0, a)$ al punto (n, b) tiene r pasos para arriba $(+1)$ y l pasos hacia abajo (-1) . Estos son tales que $r + l = n$ y $r - l = b - a$ (pues $S_n = r(+1) + l(-1) = b - a$). Estas ecuaciones determinan a l y r , lo que implica que $r = (n + b - a)/2$ y $l = (n - b + a)/2$. Cada trayectoria que lleva de a a b en n pasos tiene probabilidad $p^r q^l$, y hay $\binom{n}{(n+b-a)/2}$ trayectorias posibles. El resultado sigue. ■

Ahora vamos a considerar el problema de la ruina de un jugador que apuesta 1 peso en cada juego y tiene probabilidad p de ganar y $q = 1 - p$ de perder. Definamos

$$H_j = \min\{n \geq 0 : X_n = j\}, \quad h(i) = P_i(H_N < H_0).$$

La diferencia entre H_j y T_j que definimos anteriormente está en que H incluye el estado inicial; $h(i)$ es la probabilidad de que el jugador con capital inicial i alcance una fortuna N antes de arruinarse.

Proposición 2.7 *Sea $h(i)$ la probabilidad de que un paseo al azar que parte del estado i llegue al nivel N antes que al nivel 0. Entonces*

$$h(i) = \begin{cases} \frac{\theta^i - 1}{\theta^N - 1} & \text{si } p \neq q \\ \frac{i}{N} & \text{si } p = q = 0.5, \end{cases}$$

donde $\theta = q/p$.

Demostración. Por la definición de H_i tenemos $h(0) = 0$ y $h(N) = 1$. Para calcular $h(i)$ para $1 < i < N$ estudiamos la primera transición y obtenemos

$$h(i) = ph(i+1) + (1-p)h(i-1),$$

y rearreglando

$$p(h(i+1) - h(i)) = (1-p)(h(i) - h(i-1));$$

concluimos que

$$h(i+1) - h(i) = \frac{1-p}{p}(h(i) - h(i-1)). \quad (2.36)$$

Si $p = 1/2$ obtenemos

$$h(i+1) - h(i) = h(i) - h(i-1) = C \quad \text{para } 1 \leq i \leq N.$$

Entonces

$$1 = h(N) - h(0) = \sum_{i=1}^N (h(i) - h(i-1)) = NC$$

de modo que $C = 1/N$. Usando esto y el hecho de que $h(0) = 0$, tenemos

$$h(i) = h(i) - h(0) = \sum_{j=1}^i (h(j) - h(j-1)) = \frac{i}{N},$$

es decir, si $p = 1/2$,

$$P_i(H_N < H_0) = \frac{i}{N} \quad \text{para } 0 \leq i \leq N.$$

Como consecuencia la probabilidad de ruina es

$$P_i(H_0 < H_N) = 1 - \frac{i}{N} = \frac{N-i}{N}.$$

Si $p \neq 1/2$ los detalles son un poco más difíciles. Poniendo $C = h(1) - h(0)$, (2.36) implica que para $i \geq 1$,

$$h(i) - h(i-1) = C \left(\frac{1-p}{p} \right)^{i-1} = C\theta^{i-1},$$

con $\theta = q/p$. Sumando de $i = 1$ a N obtenemos

$$1 = h(N) - h(0) = \sum_{i=1}^N (h(i) - h(i-1)) = C \sum_{i=1}^N \theta^{i-1}$$

Recordemos que si $\theta \neq 1$,

$$\sum_{j=0}^{N-1} \theta^j = \frac{1 - \theta^N}{1 - \theta},$$

y vemos que

$$C = \frac{1 - \theta}{1 - \theta^N}.$$

Usando de nuevo que $h(0) = 0$ y sumando

$$h(j) = h(j) - h(0) = C \sum_{i=1}^j \theta^{i-1} = C \frac{1 - \theta^j}{1 - \theta} = \frac{1 - \theta^j}{1 - \theta^N}$$

Recordando la definición de $h(i)$ obtenemos que cuando $p \neq 1/2$,

$$P_i(H_N < H_0) = \frac{\theta^i - 1}{\theta^N - 1}, \quad P_i(H_0 < H_N) = \frac{\theta^N - \theta^i}{\theta^N - 1}$$

con $\theta = \frac{1-p}{p}$. ■

Ejemplo 2.18

En la siguiente tabla presentamos algunos valores de la probabilidad de ruina en diferentes circunstancias. p es la probabilidad de éxito en cada juego, i es el capital inicial, N el objetivo y P_R la probabilidad de ruina. Las probabilidades 0.4865 y 0.4737 representan la probabilidad de ganar en ruleta cuando se apuesta al color o a par-impar. En el primer caso la ruleta sólo tiene un 0 (ruleta europea) mientras que en el segundo tiene 0 y 00 (ruleta americana). La probabilidad 0.493 corresponde al juego de dados (craps).

$p = 0.474$			$p = 0.486$			$p = 0.493$		
i	N	P_R	i	N	P_R	i	N	P_R
1	10	0.94	1	10	0.92	1	10	0.91
10	100	0.99995	10	100	0.997	10	100	0.98
100	1000	1	100	1000	1	100	1000	1
5	10	0.63	5	10	0.569	5	10	0.535
50	100	0.995	50	100	0.942	50	100	0.8
500	1000	1	500	1000	1	500	1000	0.9999992
9	10	0.153	9	10	0.127	9	10	0.113
90	100	0.647	90	100	0.43	90	100	0.26
900	1000	0.99997	900	1000	0.996	900	1000	0.939

▲

Corolario 2.3 *Se tiene que para todo $i \in \mathbb{N}$*

$$P(H_0 < \infty | S_0 = i) = \begin{cases} 1 & \text{si } q \geq p \\ \left(\frac{q}{p}\right)^i & \text{si } q < p. \end{cases}$$

Demostración. Para $n \geq 1$ sea A_n el evento $A_n = \{H_0 < H_n\}$. Observemos que si $n > i$, $A_n \subseteq A_{n+1}$ puesto que $H_n \leq H_{n+1}$, para todo n . Es decir que la sucesión A_n es una sucesión creciente de eventos y además

$$\{H_0 < \infty\} = \cup_{n \geq 1} \{H_0 < H_n\},$$

y por la continuidad por debajo de la probabilidad se tiene que

$$\lim_{n \rightarrow \infty} P(A_n | S_0 = i) = P(H_0 < \infty | S_0 = i).$$

■

Sean $a, b \in \mathbb{Z}$ y $n \in \mathbb{N}$. Denotaremos por $N_n(a, b)$ al número de trayectorias que van de a a b en n pasos y por $N_n^0(a, b)$ a aquellas que van de a a b en n pasos pasando por 0 al menos una vez. Un resultado fundamental sobre el paseo al azar simple es el principio de reflexión.

Teorema 2.8 (Principio de Reflexión) *Para $a, b > 0$ se tiene que*

$$N_n^0(a, b) = N_n(-a, b).$$

Demostración. En la figura 2.3 vemos que cada trayectoria que lleva de $(0, -a)$ a (n, b) cruza el eje x al menos una vez; denotemos por $(k, 0)$ el punto en el cual esto ocurre por primera vez. Reflejando el segmento de la trayectoria anterior a $(k, 0)$ respecto al eje x se obtiene una trayectoria lleva de $(0, a)$ a (b, n) y que toca al eje x al menos una vez.

Podemos hacer algo similar para las trayectorias que inician en $(0, a)$ y pasan por 0, y obtenemos una trayectoria de $(0, -a)$ a (n, b) . ■

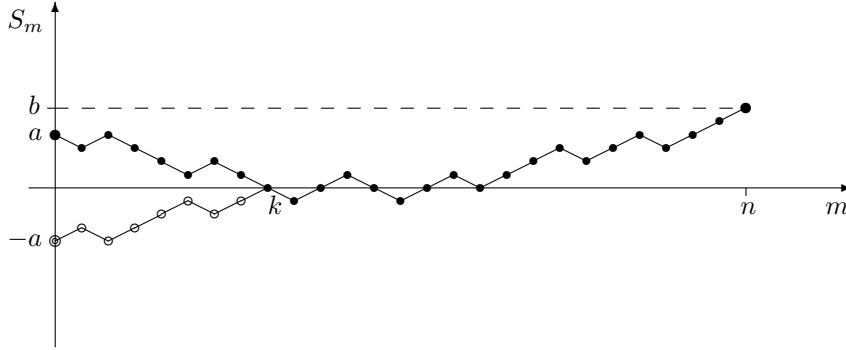


Figura 2.3

Lema 2.2 Para todo $a, b \in \mathbb{Z}$, $n \in \mathbb{N}$ con $|a - b| \leq n$, se tiene que

$$N_n(a, b) = \binom{n}{(n+b-a)/2}$$

siempre que $(n + b - a)/2 \in \mathbb{Z}$.

Demostración. La prueba de este resultado es similar a la prueba de la proposición 2.6. ■

Como consecuencia importante del lema anterior y el principio de reflexión tenemos el siguiente teorema.

Teorema 2.9 (Teorema del Escrutinio (Ballot Theorem)) Si $b > 0$, entonces el número de trayectorias que llevan de $(0, 0)$ a (n, b) y que no visitan el eje x después del primer paso es igual a

$$\frac{b}{n} N_n(0, b).$$

Demostración. Observemos que el primer paso de dichas trayectorias lleva a $(1, 1)$, por lo tanto el número que buscamos calcular está dado por

$$\begin{aligned} N_{n-1}(1, b) - N_{n-1}^0(1, b) &= N_{n-1}(1, b) - N_{n-1}(-1, b) \\ &= \frac{(n-1)!}{\left(\frac{n-b}{2}\right)! \left(\frac{n+b-2}{2}\right)!} - \frac{(n-1)!}{\left(\frac{n-b-2}{2}\right)! \left(\frac{n+b}{2}\right)!} \\ &= \frac{(n-1)!}{\left(\frac{n-b}{2}\right)! \left(\frac{n+b}{2}\right)!} \left(\frac{n+b}{2} - \frac{n-b}{2} \right) \\ &= \frac{b}{n} N_n(0, b). \end{aligned}$$

■

¿Por qué se llama este resultado el teorema del escrutinio? Supongamos que en una elección tenemos dos candidatos A y B , y que A obtiene α votos, B obtiene β votos, y $\alpha > \beta$. ¿Cuál es la probabilidad de que A lleve la ventaja durante todo el escrutinio? Supongamos que $\xi_i = 1$ si el i -ésimo individuo vota por A y $\xi_i = -1$ si vota por B . Supongamos que cualquier combinación de votos es igualmente probable, es decir que cada una tiene probabilidad $1/\binom{\alpha+\beta}{\alpha}$. La trayectoria que deben seguir los escrutinios para que A tenga la mayoría durante toda la jornada de votaciones va del punto $(0, 0)$ al punto $(\alpha + \beta, \alpha - \beta)$ y no regresa al origen. Por lo tanto la probabilidad buscada está dada por

$$\frac{\alpha - \beta}{\alpha + \beta} N_{\alpha+\beta}(0, \alpha - \beta) \frac{1}{\binom{\alpha+\beta}{\alpha}} = \frac{\alpha - \beta}{\alpha + \beta}.$$

Veamos ahora una aplicación del teorema del escrutinio a las caminatas aleatorias.

Teorema 2.10 *Supongamos que $S_0 = 0$, entonces para $n \geq 1, b \neq 0$*

$$P(S_1 S_2 \cdots S_n \neq 0, S_n = b) = \frac{|b|}{n} P(S_n = b),$$

y por lo tanto

$$P(S_1 S_2 \cdots S_n \neq 0) = \frac{E(|S_n|)}{n}.$$

Demostración. Supongamos que $S_0 = 0$ y que $b > 0$. Queremos calcular la probabilidad de que la caminata aleatoria parta de $(0, 0)$ y no visite el eje x en el intervalo $[1, n]$. Una trayectoria de este tipo tiene r pasos hacia arriba y l hacia abajo, y en consecuencia $r = (n + b)/2$ y $l = (n - b)/2$. Cada una de estas trayectoria tiene probabilidad $p^{(n+b)/2} q^{(n-b)/2}$ y por el teorema del escrutinio hay $\frac{b}{n} N_n(0, b)$ trayectorias de este tipo. La probabilidad que queremos calcular es igual a

$$P(S_1 S_2 \cdots S_n \neq 0, S_n = b) = \frac{b}{n} \binom{n}{\frac{n+b}{2}} p^{(n+b)/2} q^{(n-b)/2} = \frac{b}{n} P(S_n = b).$$

Un argumento similar vale para $b < 0$. ■

Otro resultado particularmente interesante concierne el valor máximo que alcanza la caminata aleatoria. Sea $M_n = \max\{S_j, 0 \leq j \leq n\}$, para $n \geq 0$ y supongamos que $S_0 = 0$. En particular se tiene que $M_n \geq 0$. Tenemos el siguiente teorema

Teorema 2.11 *Supongamos que $S_0 = 0$. Entonces para $r \geq 1$,*

$$P(M_n \geq r, S_n = b) = \begin{cases} P(S_n = b) & \text{si } b \geq r \\ \left(\frac{q}{p}\right)^{r-b} P(S_n = 2r - b) & \text{si } b < r. \end{cases}$$

Demostración. Supongamos que $r \geq 1$ y que $b < r$. Sea $N_n^r(0, b)$ el número de trayectorias que van de $(0, 0)$ a (n, b) pasando por r . Cualquiera de estas trayectorias visita el nivel r por lo menos una vez; denotemos por T_r el menor de estos instantes (ver figura 2.4). Reflejando la trayectoria entre T_r y n respecto a la recta $y = r$ se obtiene una trayectoria que va de $(0, 0)$ a $(n, 2r - b)$. Recíprocamente, a una trayectoria de esta forma le aplicamos la transformación inversa y obtenemos una trayectoria que va de $(0, 0)$ a (n, b) pasando por el nivel r , de longitud n . Por lo tanto podemos afirmar que

$$N_n^r(0, b) = N_n(0, 2r - b).$$

Además cada una de estas trayectorias tiene probabilidad

$$p^{(n+b)/2} q^{(n-b)/2},$$

por lo que podemos concluir que

$$\begin{aligned}
 P(M_n \geq r, S_n = b) &= N_n^r(0, b) p^{(n+b)/2} q^{(n-b)/2} \\
 &= N_n(0, 2r - b) p^{(n+b)/2} q^{(n-b)/2} \\
 &= \left(\frac{q}{p}\right)^{r-b} N_n(0, 2r - b) p^{(n+2r-b)/2} q^{(n-2r+b)/2} \\
 &= \left(\frac{q}{p}\right)^{r-b} P(S_n = 2r - b).
 \end{aligned}$$

■

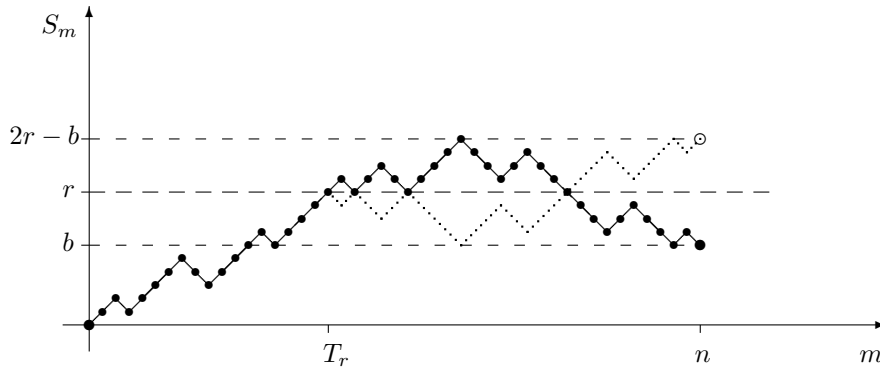


Figura 2.4

Observación 2.3 1. Una consecuencia del teorema anterior es la fórmula

$$\begin{aligned}
 P(M_n \geq r) &= P(S_n \geq r) + \sum_{b=-\infty}^{r-1} \left(\frac{q}{p}\right)^{r-b} P(S_n = 2r - b) \\
 &= P(S_n = r) + \sum_{l=r+1}^{\infty} P(S_n = l) \left(1 + (q/p)^{l-r}\right)
 \end{aligned}$$

2. Observemos que en particular si la caminata aleatoria es simétrica

$$P(M_n \geq r, S_n = b) = P(S_n = 2r - b).$$

¿Cuál es la probabilidad de que una caminata aleatoria alcance un máximo en un instante n dado? Más precisamente, ¿cuál es la probabilidad de que una caminata aleatoria que parte de 0 alcance un nivel b por la primera vez al tiempo n ? Denotemos por $f_b(n)$ a esta probabilidad. Tenemos el siguiente teorema.

Teorema 2.12 *La probabilidad de que una caminata aleatoria simple alcance el nivel b por la primera vez al tiempo n habiendo empezado de 0 está dada por*

$$f_b(n) = \frac{|b|}{n} P(S_n = b), \quad n \geq 1.$$

Demostración.

$$\begin{aligned}
f_b(n) &= P(M_{n-1} = S_{n-1} = b-1, S_n = b) \\
&= P(S_n = b | M_{n-1} = S_{n-1} = b-1) P(M_{n-1} = S_{n-1} = b-1) \\
&= p(P(M_{n-1} \geq b-1, S_{n-1} = b-1) - P(M_{n-1} \geq b, S_{n-1} = b-1)) \\
&= p \left(P(S_{n-1} = b-1) - \frac{q}{p} P(S_{n-1} = b+1) \right) \\
&= \frac{b}{n} P(S_n = b).
\end{aligned}$$

El mismo razonamiento es válido si $b < 0$. ■

Primer Regreso al Origen

Vamos a estudiar ahora la distribución de probabilidad del instante del primer retorno al estado inicial, que vamos a suponer que es 0. Queremos hallar

$$g(n) = P_0(T_0 = n), \quad n \geq 1.$$

Es sencillo ver que

$$P_0(T_0 = 1) = 0, \quad P_0(T_0 = 2) = 2pq, \quad P_0(T_0 = 3) = 0, \quad P_0(T_0 = 4) = 2p^2q^2$$

pero no es fácil generalizar este resultado para valores de $n \geq 6$. Vamos a resolver el problema hallando la f.g.p. ϕ_{T_0} de T_0 .

Proposición 2.8 *La función $g : \mathbb{N} \rightarrow [0, 1]$ definida por $g(n) = P_0(T_0 = n)$ satisface la ecuación de convolución*

$$h(n) = \sum_{k=0}^{n-2} g(n-k)h(k), \quad n \geq 2,$$

con condición inicial $g(1) = 0$, donde

$$h(n) = P_0(S_n = 0) = \begin{cases} \binom{n}{n/2} p^{n/2} q^{n/2} & \text{si } n \text{ es par,} \\ 0 & \text{si } n \text{ es impar.} \end{cases}$$

Demostración. Escribimos el evento $\{S_n = 0\}$ como

$$\{S_n = 0\} = \bigcup_{k=0}^{n-2} \{S_k = 0, S_{k+1} \neq 0, \dots, S_{n-1} \neq 0, S_n = 0\}$$

donde $k = 0, 1, \dots, n-2$ representa el instante de la última visita al 0 antes del instante n . Entonces

$$\begin{aligned}
h(n) &= P_0(S_n = 0) = \sum_{k=0}^{n-2} P_0(S_k = 0, S_{k+1} \neq 0, \dots, S_{n-1} \neq 0, S_n = 0) \\
&= \sum_{k=0}^{n-2} P_0(S_{k+1} \neq 0, \dots, S_{n-1} \neq 0, S_n = 0 | S_k = 0) P_0(S_k = 0) \\
&= \sum_{k=0}^{n-2} P_0(S_1 \neq 0, \dots, S_{n-k-1} \neq 0, S_{n-k} = 0 | S_0 = 0) P_0(S_k = 0) \\
&= \sum_{k=0}^{n-2} P_0(T_0 = n-k) P_0(S_k = 0) = \sum_{k=0}^{n-2} h(k) g(n-k)
\end{aligned}$$

Para resolver la ecuación de convolución para $g(n) = P_0(T_0 = n)$, $n \geq 1$ con $g(1) = 0$ vamos a usar la f.g.p. ■

$$\phi_{T_0}(s) = E(S^{T_0} \mathbf{1}_{\{T_0 < \infty\}})$$

Definimos la función $H : \mathbb{R} \rightarrow \mathbb{R}$ por

$$H(s) = \sum_{k=0}^{\infty} s^k P_0(S_k = 0) = \sum_{k=0}^{\infty} s^k h(k), \quad s \in [-1, 1].$$

Proposición 2.9 La función $H(s)$ está dada por

$$H(s) = (1 - 4pqs^2)^{-1/2}, \quad |s| < 1/2\sqrt{pq}.$$

y satisface la ecuación

$$\phi_{T_0}(s)H(s) = H(s) - 1, \quad s \in [-1, 1].$$

Demostración. Usaremos la fórmula para $h(n) = P_0(S_n = 0)$.

2.8. Procesos de Ramificación

Consideremos una partícula (neutrones, bacterias, virus informático, etc.) que puede generar nuevas partículas del mismo tipo. El grupo inicial de individuos pertenece a la generación 0 y suponemos que cada individuo produce una cantidad aleatoria ξ de descendientes con distribución de probabilidad

$$P(\xi = k) = p_k, \quad k = 0, 1, 2, \dots \quad (2.37)$$

donde $p_k \geq 0$, $\sum_k p_k = 1$. Suponemos que todos los individuos actúan de manera independiente, que todos viven el mismo período de tiempo y todos siguen la misma ley P dada por (2.37) para generar su descendencia (ver figura 2.5). El proceso $(X_n)_{n \geq 1}$ donde X_n representa el tamaño de la n -ésima generación, es una cadena de Markov y se conoce como un *proceso de ramificación*.

El espacio de estados de esta cadena es $\{0, 1, 2, \dots\}$ donde 0 es un estado absorbente. Por otro lado, si $X_n = k$, los k miembros de esta generación producen

$$\xi_1^n + \xi_2^n + \dots + \xi_k^n = X_{n+1} \quad (2.38)$$

descendientes, que forman la siguiente generación de modo que

$$P_{kj} = P(\xi_1^n + \xi_2^n + \dots + \xi_k^n = j | X_n = k). \quad (2.39)$$

Si una partícula produce $\xi = 0$ descendientes, lo interpretamos como que la partícula muere o desaparece. Puede ocurrir que luego de varias generaciones todos los descendientes de la partícula inicial hayan muerto o desaparecido. Decimos entonces que todos los descendientes de la partícula inicial se extinguieron. Un problema interesante es calcular la probabilidad de extinción U_∞ de un proceso de ramificación que comienza con una sola partícula. Una vez que resolvamos este problema, podemos hallar la probabilidad de que una cadena que comienza con k partículas se extinga, pues como las partículas actúan independientemente, esta probabilidad es $(U_\infty)^k$.

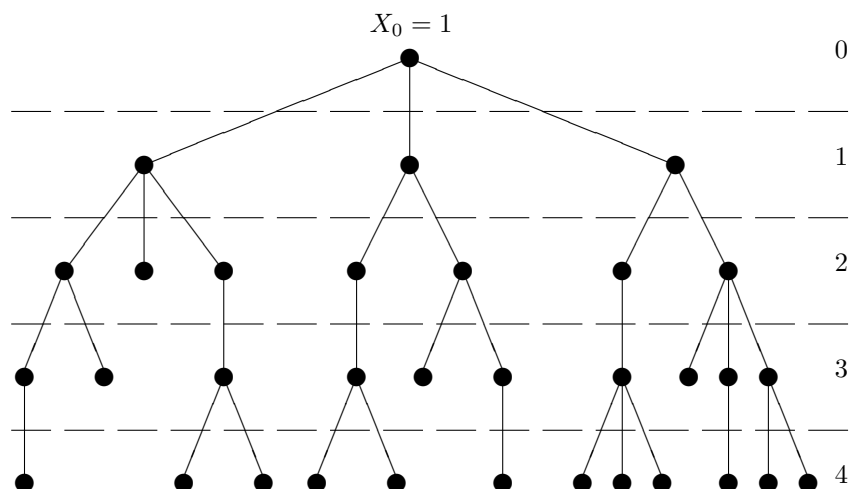


Figura 2.5

Media y Varianza de un Proceso de Ramificación.

La ecuación (2.38) caracteriza la evolución del proceso de ramificación y se puede escribir como una suma aleatoria:

$$X_{n+1} = \xi_1^n + \xi_2^n + \cdots + \xi_{X_n}^n$$

Supongamos que $E[\xi] = \mu$, $\text{Var}[\xi] = \sigma^2$ y sean $M(n)$ y $V(n)$ la media y varianza de la n -ésima generación X_n bajo la condición inicial $X_0 = 1$, $M(n) = E_1[X_n]$, $V(n) = \text{Var}_1[X_n]$. Usando las propiedades de sumas aleatorias tenemos

$$M(n+1) = \mu M(n), \quad (2.40)$$

$$V(n+1) = \sigma^2 M(n) + \mu^2 V(n). \quad (2.41)$$

La condición inicial $X_0 = 1$ hace que las relaciones recursivas (2.40) y (2.41) comiencen con $M(0) = 1$ y $V(0) = 0$. A partir de (2.40) obtenemos $M(1) = \mu \cdot 1 = \mu$, $M(2) = \mu M(1) = \mu^2$, y en general

$$M(n) = \mu^n \quad \text{para } n = 0, 1, \dots \quad (2.42)$$

Por lo tanto, el tamaño medio de la población crece geométricamente si $\mu > 1$, decrece geométricamente si $\mu < 1$ y es constante si $\mu = 1$.

Sustituyendo $M(n) = \mu^n$ en (2.41) obtenemos $V(n+1) = \sigma^2 \mu^n + \mu^2 V(n)$, y con $V(0) = 0$ se tiene

$$\begin{aligned} V(1) &= \sigma^2 \\ V(2) &= \sigma^2 \mu + \mu^2 V(1) = \sigma^2 \mu + \sigma^2 \mu^2 \\ V(3) &= \sigma^2 \mu^2 + \mu^2 V(2) = \sigma^2 \mu^2 + \sigma^2 \mu^3 + \sigma^2 \mu^4 \end{aligned}$$

y en general

$$\begin{aligned} V(n) &= \sigma^2 (\mu^{n-1} + \mu^n + \cdots + \mu^{2n-2}) \\ &= \sigma^2 \mu^{n-1} (1 + \mu + \cdots + \mu^{n-1}) \\ &= \sigma^2 \mu^{n-1} \times \begin{cases} n & \text{si } \mu = 1, \\ \frac{1-\mu^n}{1-\mu} & \text{si } \mu \neq 1. \end{cases} \end{aligned} \quad (2.43)$$

Probabilidades de Extinción.

La población se extingue cuando el tamaño de la población es 0. El instante (aleatorio) de extinción N es el primer índice n para el cual $X_n = 0$ y luego, obviamente, $X_k = 0$ para todo $k \geq N$. 0 es un estado absorbente y podemos calcular la probabilidad de extinción haciendo un análisis de la primera transición. Llamemos

$$U_n = P_1(N \leq n) = P_1(X_n = 0) \quad (2.44)$$

la probabilidad de extinción antes de n o en n . El único miembro inicial de la población produce $\xi_1^{(0)} = k$ descendientes. Por su parte, cada uno de estos descendientes generará una población de descendientes y cada una de estas líneas de descendencias debe desaparecer en $n - 1$ generaciones o antes.

Las k poblaciones generadas por el individuo inicial son independientes entre sí y tienen las mismas propiedades estadísticas que la población inicial. Por lo tanto, la probabilidad de que una cualquiera de ellas desaparezca en $n - 1$ generaciones es U_{n-1} por definición, y la probabilidad de que las k subpoblaciones mueran en $n - 1$ generaciones es $(U_{n-1})^k$, por independencia.

Por la ley de la probabilidad total

$$U_n = \sum_{k=0}^{\infty} p_k (U_{n-1})^k, \quad n = 1, 2, \dots \quad (2.45)$$

con $U_0 = 0$ y $U_1 = p_0$.

Funciones Generadoras de Probabilidad

Consideremos una v.a. ξ con valores enteros positivos y distribución de probabilidad

$$P(\xi = k) = p_k, \quad k = 0, 1, \dots$$

La función generadora de probabilidad (f.g.p.) $\phi(s)$ asociada a la v.a. ξ (o equivalentemente a su distribución (p_k)) se define por

$$\phi(s) = E[s^\xi] = \sum_{k=0}^{\infty} s^k p_k, \quad 0 \leq s \leq 1. \quad (2.46)$$

Tenemos los siguientes resultados fundamentales, algunos de los cuales estudiamos en el capítulo 1. Supondremos que $p_0 + p_1 < 1$, salvo en la propiedad 5, para evitar trivialidades:

1. *La relación entre funciones de probabilidad y funciones generadoras es 1-1.* Es posible obtener las probabilidades (p_k) a partir de ϕ usando la siguiente fórmula

$$p_k = \frac{1}{k!} \left. \frac{d^k \phi(s)}{ds^k} \right|_{s=0}. \quad (2.47)$$

2. *Si ξ_1, \dots, ξ_n son v.a.i. con funciones generadoras $\phi_1(s), \phi_2(s), \dots, \phi_n(s)$ respectivamente, la f. g. p. de su suma $X = \xi_1 + \xi_2 + \dots + \xi_n$ es el producto de las funciones generadoras respectivas*

$$\phi_X(s) = \phi_1(s)\phi_2(s) \cdots \phi_n(s). \quad (2.48)$$

3. *Los momentos de una variable que toma valores en los enteros no-negativos se pueden obtener derivando la función generadora y evaluando en 1.* Por ejemplo,

$$\left. \frac{d\phi(s)}{ds} \right|_{s=1} = p_1 + 2p_2 + 3p_3 + \dots = E[\xi]. \quad (2.49)$$

Para la segunda derivada tenemos

$$\begin{aligned} \left. \frac{d^2 \phi(s)}{ds^2} \right|_{s=1} &= \sum_{k=2}^{\infty} k(k-1)p_k \\ &= \mathbb{E}[\xi(\xi-1)] = \mathbb{E}[\xi^2] - \mathbb{E}[\xi] \end{aligned} \quad (2.50)$$

de modo que

$$\mathbb{E}[\xi^2] = \left. \frac{d^2 \phi(s)}{ds^2} \right|_{s=1} + \mathbb{E}[\xi] = \left. \frac{d^2 \phi(s)}{ds^2} \right|_{s=1} + \left. \frac{d\phi(s)}{ds} \right|_{s=1},$$

y en consecuencia

$$\text{Var}[\xi] = \mathbb{E}[\xi^2] - (\mathbb{E}[\xi])^2 = \left. \frac{d^2 \phi(s)}{ds^2} \right|_{s=1} + \left. \frac{d\phi(s)}{ds} \right|_{s=1} - \left(\left. \frac{d\phi(s)}{ds} \right|_{s=1} \right)^2.$$

4. ϕ es estrictamente convexa y creciente en $[0, 1]$. Esto es una consecuencia inmediata del hecho que ϕ es una serie de potencias con coeficientes positivos.
5. Si $\mathbb{E}(\xi) = 1$, $\text{Var}(\xi) = 0$, entonces $\phi(s) = s$.

Funciones Generadoras y Probabilidades de Extinción.

Regresamos ahora a la consideración de los procesos de ramificación. La función generadora de probabilidad para el número de descendientes ξ de cada individuo es

$$\phi(s) = \mathbb{E}[s^\xi] = \sum_{k=0}^{\infty} p_k s^k.$$

Podemos ahora escribir la relación (2.45) en términos de la función generadora:

$$U_n = \sum_{k=0}^{\infty} p_k (U_{n-1})^k = \phi(U_{n-1}) \quad (2.51)$$

es decir, si conocemos la función generadora de probabilidades $\phi(s)$, podemos calcular iterativamente las probabilidades de extinción U_n comenzando con $U_0 = 0$: $U_1 = \phi(U_0) = \phi(0)$, $U_2 = \phi(U_1)$, etc.

Ejemplo 2.19

En esta población un individuo no tiene descendientes con probabilidad $1/4$ o tiene dos descendientes con probabilidad $3/4$. La relación recursiva (2.45) es en este caso

$$U_n = \frac{1 + 3(U_{n-1})^2}{4}.$$

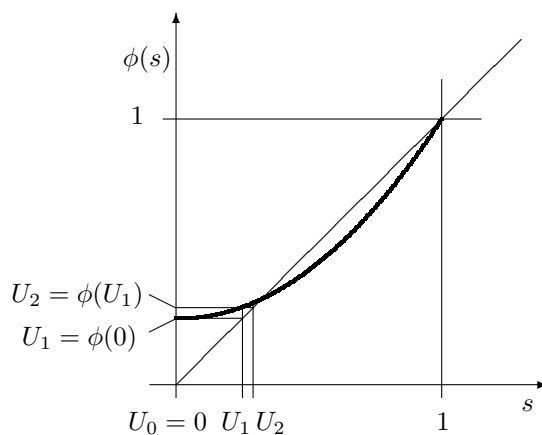


Figura 2.6

La función generadora es

$$\phi(s) = E[s^\xi] = 1 \cdot \frac{1}{4} + s^2 \cdot \frac{3}{4} = \frac{1 + 3s^2}{4}$$

y vemos que $U_n = \phi(U_{n-1})$. Representamos esta función en la Figura 2.6. Podemos ver que las probabilidades de extinción convergen de manera creciente a la menor solución de la ecuación $u = \phi(u)$.

Esto también ocurre en el caso general: Si U_∞ es la menor solución de la ecuación $u = \phi(u)$, entonces U_∞ es la probabilidad de que la población se extinga en algún momento finito. La alternativa es que la población exista indefinidamente, lo que ocurre con probabilidad $1 - U_\infty$.

En el ejemplo que estamos considerando, la ecuación $u = \phi(u)$ es

$$u = \frac{1}{4} + \frac{3}{4}u^2,$$

con soluciones 1 y $1/3$, y la menor solución es $1/3$. ▲

Puede ocurrir que $U_\infty = 1$, en cuyo caso es seguro que la población desaparece en algún momento.

Ejemplo 2.20

Si las probabilidades son ahora $p_0 = 3/4$ y $p_2 = 1/4$, la función generadora es

$$\phi(s) = \frac{3}{4} + \frac{1}{4}s^2.$$

La ecuación $u = \phi(u)$ es ahora

$$u = \frac{3}{4} + \frac{1}{4}u^2$$

con soluciones 1 y 3. Como la menor solución es 1, $U_\infty = 1$.

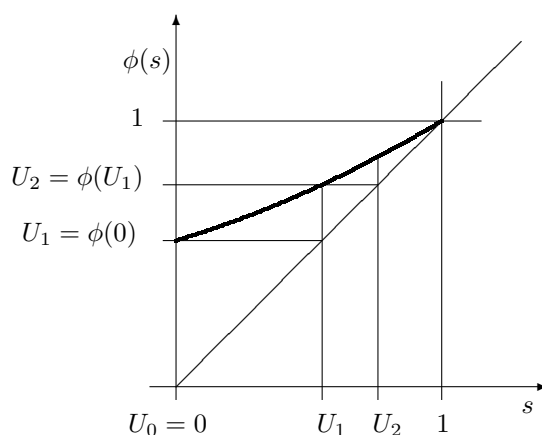


Figura 2.7

▲

Para determinar en cuál caso nos encontramos hay que ver si la curva de la función generadora $\phi(s)$ cruza la recta $y = x$ por debajo de 1, y esto se puede determinar por la pendiente de ϕ en 1:

$$\phi'(1) = \left. \frac{d\phi(s)}{ds} \right|_{s=1} = E(\xi) = \mu.$$

- Si $0 < \mu \leq 1$ entonces $\phi(t) > t$, para todo $t \in [0, 1)$. Para probarlo, definimos una función $g(t) = \phi(t) - t$, esta función satisface que $g(0) = \phi(0)$, $g(1) = 0$ y es estrictamente decreciente puesto que su derivada $g'(t) = \phi'(t) - 1$ es estrictamente negativa, y esto se debe al hecho que ϕ' es estrictamente creciente y $\phi'(1) = \mu \leq 1$. Entonces, $g(t) > 0$, para $0 \leq t < 1$. En particular, la ecuación $\phi(t) = t$, no tiene raíces en $(0, 1)$.
- Si $\mu > 1$, entonces la ecuación $\phi(t) = t$ tiene una única solución en $[0, 1)$. Esto implica que $\lim_{t \uparrow 1} \phi'(t) = \phi'(1) = \mu > 1$. Por continuidad existe un $t_0 < 1$, tal que $\phi'(t) > 1$ para todo $t_0 < t \leq 1$, por el teorema del valor intermedio vemos que

$$\frac{\phi(1) - \phi(t_0)}{1 - t_0} = \frac{1 - \phi(t_0)}{1 - t_0} = \phi'(t') > 1, \quad \text{para algún } t' \in (t_0, 1),$$

de donde sigue que $g(t_0) = \phi(t_0) - t_0 < 0$, y puesto que g es continua y $g(0) = P(\xi = 0) > 0$, podemos afirmar que existe un $0 < \eta < 1$ con $g(\eta) = 0$. Por la convexidad estricta de ϕ es claro que g no puede tener ninguna otra raíz en $(\eta, 1)$, ni en $(0, \eta)$.

Sea η la raíz más pequeña de la ecuación $\phi(t) = t$, en $[0, 1]$. Los hechos anteriores implican que esta solución existe y además: si $\mu \leq 1$, entonces $\eta = 1$; si $\mu > 1$, entonces $\eta < 1$.

Tenemos entonces

$$\begin{array}{lll} \phi'(1) < 1 & \text{no hay cruce} & U_\infty = 1, \\ \phi'(1) > 1 & \text{hay cruce} & U_\infty < 1. \end{array}$$

Pero hemos visto que $\phi'(1) = E[\xi]$ y por lo tanto,

$$\begin{array}{ll} E[\xi] < 1 & \Rightarrow U_\infty = 1, \\ E[\xi] > 1 & \Rightarrow U_\infty < 1. \end{array}$$

El caso límite corresponde a $E[\xi] = 1$, donde $E[X_n|X_0 = 1] = 1$ para todo n , de modo que el tamaño promedio de la población es constante pero la población desaparece con seguridad, a menos que la varianza sea 0, es decir, que con probabilidad 1 todo individuo tenga exactamente un descendiente, en cuyo caso la población no se extingue nunca.

Ejemplo 2.21

Supongamos que el tamaño de las familias se distribuye según una ley geométrica con parámetro q ,

$$P(\xi = k) = qp^k, \quad k \geq 0, \quad \text{para algún } p \in (0, 1).$$

Es fácil de calcular la función generadora ϕ ,

$$\phi(s) = \sum_{n=0}^{\infty} qs^n p^n = \frac{q}{1 - ps}, \quad |s| < p^{-1}.$$

La media vale $\mu = \frac{p}{q}$. Se puede verificar usando un argumento de inducción que la n -ésima composición de ϕ consigo misma puede ser escrita como

$$\phi_n(s) = \begin{cases} \frac{n - (n-1)s}{n+1 - ns} & \text{si } p = q = 1/2, \\ \frac{q[p^n - q^n - ps(p^{n-1} - q^{n-1})]}{p^{n+1} - q^{n+1} - ps(p^n - q^n)} & \text{si } p \neq q. \end{cases}$$

¿Cuál es la probabilidad de extinción en este caso? Usaremos la forma explícita de ϕ_n para responder a esta pregunta. Recordemos que

$$P(X_n = 0) = \phi_n(0) = \begin{cases} \frac{n}{n+1} & \text{si } p = q = 1/2, \\ \frac{q(p^n - q^n)}{p^{n+1} - q^{n+1}} & \text{si } p \neq q, \end{cases}$$

por lo que al hacer $n \rightarrow \infty$, vemos que

$$\lim_{n \rightarrow \infty} P(X_n = 0) = \begin{cases} 1 & \text{si } p \leq q \\ \frac{q}{p}, & \text{si } p > q. \end{cases}$$

Observemos que si para algún $n \geq 1$, $X_n = 0$ entonces $X_{n+k} = 0$, para todo $k \geq 0$. Es decir que la población se extingue en un tiempo anterior o igual a n . Tenemos que

$$\begin{aligned} P(\text{extinción en un tiempo finito}) &= \lim_{n \rightarrow \infty} P(\text{extinción antes del instante } n) \\ &= \lim_{n \rightarrow \infty} P(X_n = 0) \\ &= \begin{cases} 1 & \text{si } p \leq q \\ \frac{q}{p}, & \text{si } p > q. \end{cases} \end{aligned}$$

Conclusión: La extinción ocurre con probabilidad 1, solamente en el caso en que $p/q = \mu = E(X_1) \leq 1$; esta condición es bastante natural, puesto que $E(X_n) = E(X_1)^n \leq 1$, y es de esperarse que $X_n = 0$ tarde o temprano.

Veamos que el resultado de los ejercicios anteriores es consecuencia de un resultado más general.

Teorema 2.13 Si $X_0 = 1$ tenemos que

$$U_\infty = \lim_{n \rightarrow \infty} P_1(X_n = 0) = P(\text{extinción en un tiempo finito}) = \eta,$$

donde η es la menor solución a la ecuación, $\phi(t) = t$. Además, $\eta = 1$, si $\mu < 1$, (el caso subcrítico) y $\eta < 1$, si $\mu > 1$ (caso super-crítico), mientras que en el caso en que $\mu = 1$, (el caso crítico) $\eta = 1$ si el tamaño de las familias tiene varianza estrictamente positiva.

Demostración. Sea $U_n = P_1(X_n = 0)$. Sabemos que

$$U_n = \phi(U_{n-1}).$$

Es claro que $\{X_n = 0\} \subset \{X_{n+1} = 0\}$, para todo $n \geq 1$, entonces U_n es una sucesión creciente y acotada; en consecuencia el límite de U_n existe y por la continuidad de ϕ debe de satisfacer

$$\eta = \lim_{n \rightarrow \infty} U_n \leq 1, \quad \eta = \phi(\eta).$$

Veamos ahora que si ν es otra raíz positiva de la ecuación, entonces $\eta \leq \nu$. Dado que ϕ es una función estrictamente creciente, tenemos que

$$U_1 = \phi(0) \leq \phi(\nu) = \nu,$$

y se sigue que

$$U_2 = \phi(U_1) \leq \phi(\nu) = \nu,$$

y por inducción se ve que $U_n \leq \nu$, para todo $n \geq 1$, y por lo tanto que $\eta \leq \nu$. En consecuencia, η es la menor solución de la ecuación $t = \phi(t)$.

Ya vimos que si $\mu > 1$ entonces la ecuación $\phi(t) = t$, tiene una única solución η en $[0, 1)$, y de hecho la otra solución a la ecuación es $t = 1$. La menor solución es $\eta < 1$. Por otro lado, en el caso en que $\mu < 1$, vimos que $\phi(t) > t$ para todo $t \in [0, 1)$, y es claro que $\phi(1) = 1$, por lo tanto la solución positiva más pequeña a la ecuación $\phi(t) = t$ es $\eta = 1$. En el caso especial en que $\mu = 1$, el caso crítico, necesitamos distinguir entre el caso en que $\sigma^2 = 0$, donde $\phi(s) = s$ y por lo tanto $\eta = 0$, y el caso $\sigma^2 > 0$, donde $\phi(s) > s$, $s \in [0, 1)$ y por lo tanto $\eta = 1$. ■

2.9. Cadenas de Nacimiento y Muerte.

Sean $\mathcal{E} = \{0, 1, 2, \dots, N\}$ con $N \leq \infty$, y $\{X_n, n \geq 0\}$ una cadena de Markov con espacio de estados \mathcal{E} y matriz de transición

$$P_{i,j} = \begin{cases} q_i & \text{si } j = i - 1 \\ r_i & \text{si } j = i \\ p_i & \text{si } j = i + 1 \\ 0 & \text{en otro caso,} \end{cases}$$

con $0 \leq p_i, r_i, q_i \leq 1$ y $q_i + r_i + p_i = 1$ para todo $i \in \mathcal{E}$ y $q_0 = 0$ y $p_N = 0$ si $N < \infty$. Diremos que una cadena de Markov con espacio de estados y matriz de transición de esta forma pertenece a la clase de Cadenas de Nacimiento y Muerte. Dado que se tiene mucha libertad en los valores que pueden tomar los parámetros p_i, r_i y q_i , varias cadenas de Markov entran en esta familia, como por ejemplo la cadena de Ehrenfest, la ruina del jugador, la caminata aleatoria con barreras absorbentes o reflejantes, etc.

En esta sección daremos métodos para calcular las probabilidades de que el primer tiempo de salida de la cadena fuera de una región $[a, b]$ ocurra por la barrera superior o inferior.

Proposición 2.10 Sea H_j el primer tiempo de llegada al estado j

$$H_j = \min\{n \geq 0 : X_n = j\}, \quad i \in \mathcal{E}.$$

Supongamos que $p_j > 0$ y $q_j > 0$ para todo $j \in \{1, 2, \dots, N-1\}$ y que $p_0 > 0$ y $q_N > 0$ si $N < \infty$. Se tiene que la cadena es irreducible y para todo $a < k < b$, $a, b, k \in \mathcal{E}$,

$$P_k(H_a < H_b) = \frac{\sum_{j=k}^{b-1} \gamma_j}{\sum_{j=a}^{b-1} \gamma_j}, \quad P_k(H_b < H_a) = \frac{\sum_{j=a}^{k-1} \gamma_j}{\sum_{j=a}^{b-1} \gamma_j},$$

con $\gamma_0 = 1$ y

$$\gamma_j = \prod_{i=1}^j \left(\frac{q_i}{p_i} \right), \quad j > 0.$$

Demostración. Denotaremos por

$$h(j) = P_j(H_a < H_b), \quad \text{para todo } j \in [a, b].$$

Es claro que $h(a) = 1$ y $h(b) = 0$. Haciendo un estudio de las transiciones iniciales se demuestra que la función h satisface

$$h(j) = q_j h(j-1) + r_j h(j) + p_j h(j+1), \quad j \in (a, b).$$

En efecto,

$$\begin{aligned} h(j) &= P_j(H_a < H_b) \\ &= \sum_{i \in \mathcal{E}} P(H_a < H_b | X_1 = i, X_0 = j) P_{j,i} \\ &= \sum_{i \in \mathcal{E}} P(H_a < H_b | X_1 = i) P_{j,i} \\ &= q_j h(j-1) + r_j h(j) + p_j h(j+1). \end{aligned}$$

Dado que $r_j = 1 - p_j - q_j$ se tiene que h satisface el siguiente sistema de ecuaciones

$$h(j) = q_j h(j-1) + r_j h(j) + p_j h(j+1) \Leftrightarrow p_j (h(j+1) - h(j)) = q_j (h(j) - h(j-1)),$$

para todo $j \in (a, b)$. Multiplicando y dividiendo por γ_{j-1} se ve que el sistema anterior es equivalente a

$$h(j+1) - h(j) = \frac{\left(\frac{q_j}{p_j}\right) \gamma_{j-1}}{\gamma_{j-1}} (h(j) - h(j-1)), \quad \forall j \in (a, b),$$

que es lo mismo que

$$h(j+1) - h(j) = \frac{\gamma_j}{\gamma_{j-1}} (h(j) - h(j-1)), \quad \forall j \in (a, b).$$

Iterando este razonamiento se tiene que

$$h(j+1) - h(j) = \frac{\gamma_j}{\gamma_{j-1}} \frac{\gamma_{j-1}}{\gamma_{j-2}} \dots \frac{\gamma_{a+1}}{\gamma_a} (h(a+1) - h(a)), \quad \forall j \in (a, b),$$

esto implica que

$$h(j+1) - h(j) = \frac{\gamma_j}{\gamma_a} (h(a+1) - h(a)), \quad \forall j \in (a, b). \quad (2.52)$$

Sumando sobre $j = a+1, \dots, b-1$ se obtiene

$$h(b) - h(a+1) = \sum_{j=a+1}^{b-1} (h(j+1) - h(j)) = \frac{1}{\gamma_a} \left(\sum_{j=a+1}^{b-1} \gamma_j \right) (h(a+1) - h(a)),$$

usando que $h(b) = 0$ llegamos a

$$h(a+1) \left(\frac{\sum_{j=a}^{b-1} \gamma_j}{\gamma_a} \right) = h(a) \left(\frac{\sum_{j=a+1}^{b-1} \gamma_j}{\gamma_a} \right)$$

de donde se obtiene que

$$h(a+1) = \frac{\sum_{j=a+1}^{b-1} \gamma_j}{\sum_{j=a}^{b-1} \gamma_j} h(a) \quad \Leftrightarrow \quad h(a+1) - h(a) = \frac{-\gamma_a}{\sum_{j=a}^{b-1} \gamma_j}$$

Recordando la ecuación (2.52) vemos que

$$h(j+1) - h(j) = \frac{\gamma_j}{\gamma_a} (h(a+1) - h(a)) = \frac{\gamma_j}{\gamma_a} \left(\frac{-\gamma_a}{\sum_{j=a}^{b-1} \gamma_j} \right) = \frac{-\gamma_j}{\sum_{j=a}^{b-1} \gamma_j}, \quad a < j < b.$$

Finalmente, se tiene que

$$P_k(H_a < H_b) = h(k) = \sum_{j=k}^{b-1} (h(j) - h(j+1)) = \frac{\sum_{j=k}^{b-1} \gamma_j}{\sum_{j=a}^{b-1} \gamma_j}.$$

y la probabilidad de que $P_z(H_b < H_a)$ se calcula usando que $P_z(H_b < H_a) = 1 - P_z(H_a < H_b)$. ■

Corolario 2.4 *Bajo la hipótesis de la proposición anterior y si $a < k < b$*

$$P_k(H_a < H_b) = P_k(T_a < T_b),$$

con la notación usual.

Corolario 2.5 *Bajo las hipótesis de la proposición anterior y si $N = \infty$,*

$$P_1(T_0 < \infty) = 1 - \frac{1}{\sum_{j=0}^{\infty} \gamma_j}.$$

Demostración. Supongamos que $X_0 = 1$, después de una breve reflexión es fácil convencerse de que $P_1(T_n \geq n-1) = 1$ y por lo tanto $P_1(\lim_{n \rightarrow \infty} T_n = \infty) = 1$, ya que la medida de probabilidad P_1 es continua, es decir que si B_n es una sucesión creciente de eventos y $B = \cup_{n=0}^{\infty} B_n$ entonces

$$P_1(B) = \lim_{n \rightarrow \infty} P_1(B_n).$$

En este caso $B_n = \{\omega \in \Omega : T_n \geq n-1\}$. Ahora, sean $C_n = \{\omega \in \Omega : T_0 < T_n\}$, es claro que $C_n \subseteq C_{n+1}$, y que

$$\cup_{n=1}^{\infty} C_n = \{\omega \in \Omega : T_0 < \infty\},$$

y usando de nuevo la continuidad de la probabilidad P_1 llegamos a

$$\lim_{n \rightarrow \infty} P_1(T_0 < T_n) = P_1(T_0 < \infty),$$

y dado que

$$P_1(T_0 < T_{n+1}) = \frac{\sum_{j=1}^n \gamma_j}{\sum_{j=0}^n \gamma_j} = 1 - \frac{\gamma_0}{\sum_{j=0}^n \gamma_j} \xrightarrow{n \rightarrow \infty} 1 - \frac{1}{\sum_{j=0}^{\infty} \gamma_j},$$

ya que $\gamma_0 = 1$. ■

2.10. Simulación de Cadenas de Markov

En el capítulo 1 estudiamos la simulación de variables discretas usando el método de la transformada inversa y el método del rechazo. Vimos además algunos métodos más eficientes para ciertas distribuciones particulares como la binomial. Podemos usar estas técnicas que hemos estudiado para simular las trayectorias de una cadena de Markov con espacio de estados discreto y tiempo discreto. Supongamos que el espacio de estados es $\mathcal{E} = \{0, 1, 2, \dots, N\}$.

Para poder simular las trayectorias de una cadena a Markov a tiempo discreto necesitamos conocer la distribución del estado inicial π y la matriz de transición P . A partir de la primera podemos generar el valor inicial usando, por ejemplo, el método de la transformada inversa: Generamos una variable aleatoria con distribución uniforme en $(0, 1)$ y observamos en cual de los intervalos

$$[0, \pi_0), [\pi_0, \pi_0 + \pi_1), [\pi_0 + \pi_1, \pi_0 + \pi_1 + \pi_2), \dots, \left[\sum_{i=0}^{N-1} \pi_i, 1 \right),$$

cae la variable U que simulamos y le asignamos a X_0 el valor j correspondiente.

Una vez que obtenemos el valor de X_0 , digamos $X_0 = k$, para generar el siguiente valor X_1 tenemos que ver la fila k de la matriz de transición, que corresponde a las probabilidades de transición de la cadena partiendo del estado k : $P_{k,j}$, $j \in \mathcal{E}$. Usando el método descrito anteriormente con estas probabilidades en lugar de π_j obtenemos el valor correspondiente a X_1 . De aquí en adelante se usa este mismo procedimiento: Para generar X_{j+1} sabiendo que $X_j = i$, usamos la i -ésima fila de la matriz de transición para generar el valor de esta variable.

Algoritmo:

1. Generamos el valor inicial X_0 usando la distribución inicial π .
2. Dado el valor $X_{i-1} = k$ generamos X_i usando la i -ésima fila de la matriz de transición P .
3. Repetimos el paso 2 tantas veces como pasos de la cadena queramos generar.

Ejemplo 2.22

Queremos simular las trayectorias de una cadena de Markov con espacio de estados $\mathcal{E} = \{1, 2, 3\}$, distribución inicial uniforme y matriz de transición

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{6} & \frac{1}{2} \end{pmatrix}$$

Para generar el valor inicial tenemos que generar el valor de una variable con distribución uniforme en $\{1, 2, 3\}$. Por lo tanto generamos una variable U con distribución uniforme en $(0, 1)$ y si $U < 1/3$, $X_0 = 1$, si $1/3 \leq U < 2/3$, $X_0 = 2$ y finalmente si $U > 2/3$, $X_0 = 3$.

El siguiente paso depende del estado en el cual se encuentre a cadena:

- Si estamos en el estado 1, para generar el siguiente estado tenemos que usar la distribución de probabilidad dada por la primera fila de la matriz: la probabilidad de que la cadena permanezca en 1 es 0.5, de que pase a 2 es 0.25 y de que pase a 3 también es 0.25. Por lo tanto generamos una variable U con distribución uniforme en $(0, 1)$ y si $U < 0.5$, el siguiente valor de la cadena es 1; si $0.5 \leq U < 0.75$ el siguiente valor es 2 y si $U \geq 0.75$ el siguiente valor es 3.
- Si estamos en el estado 2, para generar el siguiente estado tenemos que usar la distribución de probabilidad dada por la segunda fila de la matriz: la probabilidad de que la cadena pase a 1 es 0.5, de que permanezca en 2 es 0 y de que pase a 3 también es 0.5. Por lo tanto generamos una variable U con distribución uniforme en $(0, 1)$ y si $U < 0.5$, el siguiente valor de la cadena es 1 y si $U \geq 0.5$ el siguiente valor es 3.
- Si estamos en el estado 3, para generar el siguiente estado tenemos que usar la distribución de probabilidad dada por la tercera fila de la matriz: la probabilidad de que la cadena pase a 1 es $1/3$, de que pase a 2 es $1/6$ y de que permanezca en 3 es 0.5. Por lo tanto generamos una variable U con distribución uniforme en $(0, 1)$ y si $U < 1/3$, el siguiente valor de la cadena es 1; si $1/3 \leq U < 0.5$ el siguiente valor es 2 y si $U \geq 0.5$ el siguiente valor es 3.

▲

Capítulo 3

Propiedades Asintóticas

3.1. Distribuciones Estacionarias

Definición 3.1 Sea $X_n, n \geq 1$, una cadena de Markov con espacio de estados \mathcal{E} y matriz de transición P . Sea $\pi(i), i \in \mathcal{E}$, una distribución de probabilidad, es decir,

$$\pi(i) \geq 0, \text{ para todo } i \in \mathcal{E}, \quad \sum_{i \in \mathcal{E}} \pi(i) = 1.$$

Si

$$\sum_i \pi(i)P_{i,j} = \pi(j), \text{ para todo } j \in \mathcal{E}, \quad (3.1)$$

decimos que π es una *distribución estacionaria* o una *medida invariante* para la cadena X_n .

Para una cadena con espacio de estados finito, observamos que si P es la matriz de transición y π es el vector que representa la distribución estacionaria, podemos escribir matricialmente la relación (3.1) como $\pi'P = \pi$, donde π es un vector columna y π' es su traspuesto.

Dada una cadena de Markov, no siempre es posible encontrar una distribución estacionaria, como veremos más adelante.

Sea π una distribución estacionaria, entonces, usando las ecuaciones de Chapman-Kolmogorov,

$$\begin{aligned} \sum_i \pi(i)P_{ij}^{(2)} &= \sum_i \pi(i) \sum_k P_{ik}P_{kj} \\ &= \sum_k \left(\sum_i \pi(i)P_{ik} \right) P_{kj} \\ &= \sum_k \pi(k)P_{kj} = \pi(j). \end{aligned}$$

De manera similar, por inducción obtenemos que para todo n ,

$$\sum_i \pi(i)P_{ij}^{(n)} = \pi(j), \text{ para todo } j \in \mathcal{E}. \quad (3.2)$$

Por lo tanto, si la distribución del estado inicial X_0 es π , (3.2) implica que para todo n ,

$$P(X_n = j) = \pi(j), \text{ para todo } j \in \mathcal{E},$$

y en consecuencia la distribución de X_n es independiente de n . Esto quiere decir que la distribución estacionaria representa una distribución de equilibrio del proceso: si el proceso se inicia con una distribución estacionaria entonces es estrictamente estacionario.

Supongamos ahora que la distribución de X_n es independiente de n , entonces la distribución inicial π_0 debe satisfacer

$$\pi_0(j) = P(X_0 = j) = P(X_1 = j) = \sum_i \pi_0(i)P_{ij}$$

y en consecuencia π_0 satisface la condición (3.1) de una distribución estacionaria. Resumiendo, la distribución de X_n es independiente de n si y sólo si la distribución inicial es una distribución estacionaria.

Ejemplo 3.1

Consideremos una cadena de Markov con espacio de estados $\mathcal{E} = \{0, 1, 2\}$ y matriz de transición

$$P = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/4 & 1/2 & 1/4 \\ 1/6 & 1/3 & 1/2 \end{pmatrix}$$

Veamos que esta cadena tiene una única distribución estacionaria π . La relación que debe satisfacer esta distribución es $\pi P = \pi$, es decir,

$$(\pi_0, \pi_1, \pi_2) \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/4 & 1/2 & 1/4 \\ 1/6 & 1/3 & 1/2 \end{pmatrix} = \begin{pmatrix} \pi_0 \\ \pi_1 \\ \pi_2 \end{pmatrix}$$

y obtenemos las tres ecuaciones siguientes,

$$\begin{aligned} \frac{\pi_0}{3} + \frac{\pi_1}{4} + \frac{\pi_2}{6} &= \pi_0, \\ \frac{\pi_0}{3} + \frac{\pi_1}{2} + \frac{\pi_2}{3} &= \pi_1, \\ \frac{\pi_0}{3} + \frac{\pi_1}{4} + \frac{\pi_2}{2} &= \pi_2, \end{aligned}$$

junto con la condición adicional de que el vector π represente una distribución de probabilidad, es decir

$$\pi_0 + \pi_1 + \pi_2 = 1.$$

Resolviendo este sistema obtenemos

$$\pi_0 = \frac{6}{25}, \quad \pi_1 = \frac{2}{5}, \quad \pi_2 = \frac{9}{25},$$

que son positivos y satisfacen las cuatro ecuaciones, de modo que representan la única distribución estacionaria para la cadena. \blacktriangle

Dada una cadena de Markov, supongamos ahora que existe una distribución ν tal que para todo $i \in \mathcal{E}$,

$$\lim_{n \rightarrow \infty} P_{ij}^{(n)} = \nu(j), \quad \text{para todo } j \in \mathcal{E}. \quad (3.3)$$

Entonces la distribución de X_n se aproxima a ν cuando $n \rightarrow \infty$, sin importar cual sea la distribución inicial de la cadena. En este caso decimos que ν es una *distribución asintótica* para la cadena.

Supongamos que (3.3) vale y sea π_0 la distribución inicial de la cadena, entonces

$$P(X_n = j) = \sum_i \pi_0(i) P_{ij}^{(n)}.$$

Suponiendo que podemos intercambiar límites con series, haciendo $n \rightarrow \infty$ y usando (3.3) obtenemos

$$\lim_{n \rightarrow \infty} P(X_n = j) = \sum_i \pi_0(i) \nu(j),$$

y como $\sum_i \pi_0(i) = 1$, concluimos que

$$\lim_{n \rightarrow \infty} P(X_n = j) = \nu(j), \quad \text{para todo } j \in \mathcal{E}. \quad (3.4)$$

Esta fórmula indica que, sin importar cual sea la distribución inicial, para valores grandes de n la distribución de X_n es aproximadamente igual a la distribución asintótica ν .

Si la cadena tiene una distribución estacionaria π entonces, necesariamente, $\pi = \nu$, porque podemos iniciar la cadena con la distribución π y entonces

$$P(X_n = j) = \sum_i \pi(i) P_{ij}^{(n)} = \pi(j), \quad \text{para todo } j \in \mathcal{E}.$$

Comparando con (3.4) vemos que ambas distribuciones tienen que coincidir.

Por lo tanto, de acuerdo a este argumento, si una cadena de Markov tiene una distribución estacionaria y una distribución asintótica, ambas deben coincidir.

Teorema 3.1 *Sea $X = \{X_n, n \geq 0\}$ una cadena de Markov con espacio de estados finito y matriz de transición P . Supongamos que para algún $i \in \mathcal{E}$ se cumple que*

$$\lim_{n \rightarrow \infty} P_{i,j}^{(n)} := \pi(j), \quad \text{para todo } j \in \mathcal{E}, \quad (3.5)$$

Entonces, el vector $(\pi(j), j \in \mathcal{E})$ es una distribución de probabilidad invariante.

Demostración. Es inmediato que $0 \leq \pi(j) \leq 1$, para todo $j \in \mathcal{E}$ pues esto se vale para las potencias de la matriz de transición P : $0 \leq P_{i,j}^{(n)} \leq 1$, para todo $n \geq 1$ y $i, j \in \mathcal{E}$. Veamos que, en efecto, π es un vector de probabilidad, puesto que \mathcal{E} es finito los siguientes intercambios de suma y límite los podemos hacer sin correr riesgo a equivocación

$$\sum_{j \in \mathcal{E}} \pi(j) = \sum_{j \in \mathcal{E}} \lim_{n \rightarrow \infty} P_{i,j}^{(n)} = \lim_{n \rightarrow \infty} \sum_{j \in \mathcal{E}} P_{i,j}^{(n)} = 1.$$

Para finalizar veamos que π es un vector de probabilidad invariante. Por las ecuaciones de Chapman-Kolmogorov tenemos que para todo $j \in \mathcal{E}$.

$$\begin{aligned} \pi(j) &= \lim_{n \rightarrow \infty} P_{i,j}^{(n+1)} = \lim_{n \rightarrow \infty} \sum_{k \in \mathcal{E}} P_{i,k}^{(n)} P_{k,j} \\ &= \sum_{k \in \mathcal{E}} \lim_{n \rightarrow \infty} P_{i,k}^{(n)} P_{k,j} = \sum_{k \in \mathcal{E}} \pi(k) P_{k,j} \end{aligned}$$

■

Observación 3.1 1. Como el espacio de estados en el teorema anterior es finito, se tiene que $\pi(k) > 0$ para algún $k \in \mathcal{E}$ pues π es un vector de probabilidad. En el caso en que \mathcal{E} es infinito esto no se puede garantizar. Por ejemplo, tomemos una cadena de Markov con espacio de estados \mathcal{E} infinito y tal que todos sus estados son transitorios, como la caminata aleatoria no simétrica. Puesto que todos los estados son transitorios se tiene que

$$\lim_{n \rightarrow \infty} P_{i,j}^{(n)} = 0 = \pi(j), \quad \forall j \in \mathcal{E}.$$

El vector π es sin duda invariante, $0P = 0$, pero no es un vector de probabilidad pues la suma de sus entradas es 0.

2. En el enunciado del teorema no se pide que la relación (3.5) se cumpla para todos los estados iniciales $i \in \mathcal{E}$, se pide que el límite exista para algún $i \in \mathcal{E}$. Si este límite existe para todo $i \in \mathcal{E}$ y no depende del valor inicial, entonces π es a la vez, una distribución asintótica y una distribución estacionaria, y por lo tanto es única. Sin embargo, como veremos, hay casos en los cuales la distribución límite depende del estado inicial.

3.2. Visitas a un Estado Recurrente

Veamos qué cosas pueden ocurrir para que una cadena no tenga distribución asintótica. Consideremos una cadena de Ehrenfest con tres bolas. En este caso la matriz de transición es

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

Si calculamos la matriz de transición en dos pasos, la disposición de los ceros en la matriz cambia,

$$P^2 = \begin{pmatrix} 1/3 & 0 & 2/3 & 0 \\ 0 & 7/9 & 0 & 2/9 \\ 2/9 & 0 & 7/9 & 0 \\ 0 & 2/3 & 0 & 1/3 \end{pmatrix}$$

Para ver que esta situación se mantiene para valores mayores de n , observamos que si inicialmente tenemos un número impar de bolas en la caja de la izquierda, no importa si añadimos o quitamos una, el resultado será un número par. De manera similar, si hay un número par de bolas inicialmente, en el próximo paso habrá un número impar. Esta situación de alternancia entre números pares e impares indica que es imposible regresar al estado inicial después de un número impar de pasos, es decir, si n es impar, $P_{ii}^n = 0$ para todo i .

Hay una manera de manejar esta situación en la cual no existe el límite de P_{ij}^n cuando $n \rightarrow \infty$. Sea a_n , $n \geq 0$, una sucesión de números. Si

$$\lim_{n \rightarrow \infty} a_n = L \tag{3.6}$$

para algún L finito, entonces

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=1}^n a_m = L. \tag{3.7}$$

Si (3.7) es cierto decimos que (a_n) converge a L en el sentido de Cesàro. Este tipo de convergencia es más general que la convergencia usual: es posible que (3.7) sea cierta sin que lo sea (3.6). Por ejemplo, si $a_n = 0$ para n par y $a_n = 1$ para n impar, entonces la sucesión no tiene límite cuando $n \rightarrow \infty$ pero

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=1}^n a_m = \frac{1}{2}.$$

Veremos a continuación que para cualquier par de estados i, j de cualquier cadena de Markov, el límite

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=1}^n P_{ij}^{(m)}$$

existe.

Recordemos que $N_n(j) = \sum_{m=1}^n \mathbf{1}_j(X_m)$ representa el número de visitas de la cadena al estado j durante $m = 1, \dots, n$. El valor esperado del número de visitas para una cadena que comienza en i está dado por

$$E_i[N_n(j)] = G_n(i, j) = \sum_{m=1}^n P_{ij}^{(m)}.$$

Sea j un estado transitorio, entonces hemos visto que $\lim_{n \rightarrow \infty} N_n(j) = N(j) < \infty$ con probabilidad 1, y $\lim_{n \rightarrow \infty} G_n(i, j) = G(i, j) < \infty$ para todo $i \in \mathcal{E}$. En consecuencia

$$\lim_{n \rightarrow \infty} \frac{N_n(j)}{n} = 0 \quad \text{con probabilidad 1,}$$

y también

$$\lim_{n \rightarrow \infty} \frac{G_n(i, j)}{n} = 0 \quad \text{para todo } i \in \mathcal{E}. \quad (3.8)$$

Observamos que $N_n(j)/n$ es la proporción de las primeras n unidades de tiempo que la cadena está en el estado j y que $G_n(i, j)/n$ es el valor esperado de esta proporción para una cadena que comienza en i .

Sea ahora j un estado recurrente y llamemos $m_j = E_j[T_j]$ al tiempo medio de regreso a j para una cadena que comienza en j , si este tiempo de regreso tiene esperanza finita, y ponemos $m_j = \infty$ si no.

Para probar el próximo teorema, necesitamos la Ley Fuerte de los Grandes Números y el Teorema de Convergencia Acotada:

Teorema 3.2 (Ley Fuerte de los Grandes Números) *Sea ξ_i , $i \geq 1$ una sucesión de v.a.i.i.d. Si estas variables tienen media finita μ , entonces*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \xi_i = \mu$$

con probabilidad 1. Si $\xi_i \geq 0$ y las variables no tienen esperanza finita, el resultado vale si ponemos $\mu = +\infty$.

Teorema 3.3 (Teorema de Convergencia Acotada) *Sea ξ_i , $i \geq 1$, una sucesión de v.a. Si existe una constante K tal que $|\xi_i| < K$ para todo $i \geq 1$, y si $\xi_i \rightarrow \xi$ cuando $i \rightarrow \infty$, entonces $E(\xi_i) \rightarrow E(\xi)$.*

Sea X_n , $n \geq 1$ una cadena de Markov con probabilidades de transición estacionarias que comienza en un estado recurrente j . Con probabilidad 1 la cadena regresa a j infinitas veces. Para $r \geq 1$ sea T_j^r el instante de la r -ésima visita a j :

$$T_j^r = \min\{n \geq 1 : N_n(j) = r\}.$$

Ponemos $W_j^1 = T_j^1 = T_j$ y para $r \geq 2$ sea $W_j^r = T_j^r - T_j^{r-1}$, el tiempo de espera entre la $(r-1)$ -ésima visita a j y la r -ésima visita. Claramente,

$$T_j^r = W_j^1 + \cdots + W_j^r.$$

Lema 3.1 *Sea X_n , $n \geq 1$ una cadena de Markov con probabilidades de transición estacionarias que comienza en un estado recurrente j . Las variables W_j^r , $r \geq 1$, son i.i.d.*

Demostración. Veamos en primer lugar que

$$P_j(W_j^{r+1} = z_{r+1} | W_j^1 = z_1, \dots, W_j^r = z_r) = P_j(W_j^1 = z_{r+1}). \quad (3.9)$$

Definimos $t_0 = 0$, $t_i = t_{i-1} + z_i = \sum_{q=1}^i z_q$ para $1 \leq i \leq r+1$ y sea

$$A_r = \{t_1, t_2, \dots, t_r\}$$

el conjunto de los instantes en los cuales el proceso realiza las primeras r visitas al estado j . Por lo tanto, para $1 \leq s \leq t_r$, $X_s = j$ si $s \in A_r$, $X_s \neq j$ si $s \notin A_r$. Usando esta notación vemos que el evento

$$\{W_j^1 = z_1, W_j^2 = z_2, \dots, W_j^r = z_r\}$$

se puede escribir como

$$\{X_s = j \text{ para } s \in A_r, X_s \neq j \text{ para } s \notin A_r, 1 \leq s \leq t_r\}.$$

Por lo tanto el lado izquierdo de (3.9) es ahora

$$P_j(X_{t_{r+1}} = j, X_s \neq j \text{ para } t_r + 1 \leq s < t_{r+1} | X_s = j \text{ para } s \in A_r, \\ X_s \neq j \text{ para } s \notin A_r, 1 \leq s \leq t_r),$$

por la propiedad de Markov esto es

$$P(X_{t_{r+1}} = j, X_s \neq j \text{ para } t_r + 1 \leq s < t_{r+1} | X_{t_r} = j)$$

y teniendo en cuenta la definición de los t_i y la homogeneidad de la cadena, esto es igual a

$$P_j(W_j^1 = z_{r+1})$$

de modo que hemos probado (3.9).

Para ver que todas tienen la misma distribución veamos la distribución de W_j^2 :

$$P_j(W_j^2 = z) = \sum_{s=1}^{\infty} P_j(W_j^2 = z | W_j^1 = s) P_j(W_j^1 = s) \\ = \sum_{s=1}^{\infty} P_j(W_j^1 = z) P_j(W_j^1 = s) = P_j(W_j^1 = z)$$

y el resultado se obtiene por inducción.

También por inducción, se prueba fácilmente que

$$P_j(W_j^1 = z_1, \dots, W_j^r = z_r) = P_j(W_j^1 = z_1) \cdots P_j(W_j^r = z_r),$$

y esto muestra que las variables W_j^k son independientes. ■

Teorema 3.4 Sea j un estado recurrente, entonces

$$\lim_{n \rightarrow \infty} \frac{N_n(j)}{n} = \frac{\mathbf{1}_{\{T_j < \infty\}}}{m_j} \quad \text{con probabilidad 1,} \quad (3.10)$$

y

$$\lim_{n \rightarrow \infty} \frac{G_n(i, j)}{n} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=1}^n P_{ij}^{(m)} = \frac{\rho_{ij}}{m_j} \quad \text{para todo } i \in \mathcal{E}. \quad (3.11)$$

Demostración. Consideremos una cadena de Markov que comienza en un estado recurrente j . Con probabilidad 1 regresa a j infinitas veces. Por el lema 3.1, las variables W_j^1, W_j^2, \dots son i.i.d. y tienen media común $E_j(W_j^1) = E_j(T_j) = m_j$. Por la LFGN tenemos

$$\lim_{k \rightarrow \infty} \frac{1}{k} (W_j^1 + W_j^2 + \dots + W_j^k) = m_j \quad \text{c. p. 1,}$$

es decir, que

$$\lim_{k \rightarrow \infty} \frac{T_j^k}{k} = m_j \quad \text{c. p. 1.} \quad (3.12)$$

Sea $N_n(j) = k$, entonces, al instante n la cadena ha hecho exactamente k visitas a j . Por lo tanto, la visita k ocurre en o antes del instante n , mientras que la visita $k + 1$ ocurre después de n . Esto lo expresamos en la siguiente desigualdad

$$T_j^{N_n(j)} \leq n < T_j^{N_n(j)+1},$$

y por lo tanto,

$$\frac{T_j^{N_n(j)}}{N_n(j)} \leq \frac{n}{N_n(j)} < \frac{T_j^{N_n(j)+1}}{N_n(j)},$$

siempre que $N_n(j) \geq 1$. Como $N_n(j) \rightarrow \infty$ con probabilidad 1 cuando $n \rightarrow \infty$, estas desigualdades y (3.12) implican que

$$\lim_{n \rightarrow \infty} \frac{n}{N_n(j)} = m_j \quad \text{c. p. 1.}$$

Sea, de nuevo, j un estado recurrente pero supongamos ahora que X_0 tiene distribución arbitraria, entonces es posible que la cadena nunca llegue a j . Si llega, el argumento anterior es válido y (3.10) es cierta. Por lo tanto

$$\lim_{n \rightarrow \infty} \frac{N_n(j)}{n} = \frac{\mathbf{1}_{\{T_j < \infty\}}}{m_j} \quad \text{c.p.1.}$$

Por definición $0 \leq N_n(j) \leq n$, y en consecuencia

$$0 \leq \frac{N_n(j)}{n} \leq 1.$$

Usando el Teorema de Convergencia Acotada tenemos

$$\lim_{n \rightarrow \infty} E_i \left[\frac{N_n(j)}{n} \right] = E_i \left[\frac{\mathbf{1}_{\{T_j < \infty\}}}{m_j} \right] = \frac{P_i(T_j < \infty)}{m_j} = \frac{\rho_{ij}}{m_j}$$

y por lo tanto (3.11) vale. ■

El teorema anterior tiene la siguiente interpretación. Supongamos que estamos considerando una cadena de Markov irreducible y finita, de modo que todos los estados son recurrentes y se comunican. Entonces, con probabilidad 1 todos los estados serán visitados y $\rho_{ij} = P_i(T_j < \infty) = 1$ para cualesquiera i, j . Por lo tanto, el tiempo promedio que la cadena pasa en el estado j cuando n es grande es, aproximadamente, $1/m_j = 1/E_j(T_j)$, es decir, el inverso del tiempo medio de retorno.

Ejemplo 3.2 (Cadena con dos estados)

Consideremos una cadena de Markov con dos estados posibles y matriz de transición

$$P = \begin{pmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{pmatrix}.$$

Supongamos que $0 < \alpha, \beta \leq 1$, entonces tenemos una fórmula explícita para las potencias de la matriz de transición (ver ejemplo 2.9)

$$P^n = \frac{1}{\alpha + \beta} \left[\begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix} + (1 - \alpha - \beta)^n \begin{pmatrix} \alpha & -\alpha \\ -\beta & \beta \end{pmatrix} \right].$$

Si $\alpha + \beta < 2$, haciendo $n \rightarrow \infty$ el factor $(1 - \alpha - \beta)^n \rightarrow 0$ y por lo tanto

$$\lim_{n \rightarrow \infty} P^n = \begin{pmatrix} \frac{\beta}{\alpha + \beta} & \frac{\alpha}{\alpha + \beta} \\ \frac{\beta}{\alpha + \beta} & \frac{\alpha}{\alpha + \beta} \end{pmatrix}.$$

En este ejemplo tenemos convergencia de las potencias de las probabilidades de transición P_{ij}^n cuando $n \rightarrow \infty$ y como consecuencia también hay convergencia en el sentido de Cesáro.

En el caso límite $\alpha = \beta = 1$, la cadena sigue siendo recurrente, pero ahora no hay convergencia de P_{ij}^n cuando $n \rightarrow \infty$ ya que

$$P^{2n} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad P^{2n+1} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Sin embargo, es fácil ver que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=1}^n P_{ij}^{(m)} = \frac{1}{2},$$

que es consistente con el resultado anterior (si $\alpha = \beta = 1$, $\alpha/(\alpha + \beta) = 1/2$ para $i = 1, 2$).

Una interpretación de este resultado es que, a largo plazo, la cadena estará en el estado 1 una fracción de tiempo $\beta/(\alpha + \beta)$ y en el otro estado la fracción complementaria $\alpha/(\alpha + \beta)$. \blacktriangle

3.3. Estados Recurrentes

Definición 3.2 Un estado recurrente j es *recurrente nulo* si $m_j = \infty$.

Por el teorema 3.4 vemos que si j es recurrente nulo,

$$\lim_{n \rightarrow \infty} \frac{G_n(i, j)}{n} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=1}^n P_{ij}^{(m)} = 0, \quad \forall i \in \mathcal{E}. \quad (3.13)$$

Es posible mostrar un resultado más fuerte: Si j es recurrente nulo entonces $\lim_{n \rightarrow \infty} P_{ij}^{(n)} = 0$ para $i \in \mathcal{E}$.

Definición 3.3 Un estado recurrente j es *recurrente positivo* si $m_j < \infty$.

Por el teorema 3.4 vemos que si j es recurrente positivo,

$$\lim_{n \rightarrow \infty} \frac{G_n(j, j)}{n} = \frac{1}{m_j} > 0.$$

Consideremos una cadena que comienza en un estado recurrente j . A partir del teorema 3.4 vemos que si j es recurrente nulo entonces, con probabilidad 1, la proporción del tiempo que la cadena está en el estado j durante las primeras n unidades de tiempo tiende a cero cuando $n \rightarrow \infty$, mientras que si j es recurrente positivo, con probabilidad 1 esta proporción tiende al límite positivo $1/m_j$ cuando $n \rightarrow \infty$.

Teorema 3.5 *Sea i un estado recurrente positivo y supongamos que desde i podemos acceder a j . Entonces j es recurrente positivo.*

Demostración. Ya vimos que en este caso desde j también se accede a i . Por lo tanto existen enteros positivos n_1 y n_2 tales que

$$P_{ji}^{(n_1)} > 0 \quad \text{y} \quad P_{ij}^{(n_2)} > 0.$$

Tenemos

$$P_{jj}^{(n_1+m+n_2)} \geq P_{ji}^{(n_1)} P_{ii}^{(m)} P_{ij}^{(n_2)},$$

sumando sobre $m = 1, 2, \dots, n$ y dividiendo por n concluimos que

$$\frac{1}{n} (G_{n_1+n+n_2}(j, j) - G_{n_1+n_2}(j, j)) \geq P_{ji}^{(n_1)} P_{ij}^{(n_2)} \frac{1}{n} G_n(i, i).$$

Haciendo $n \rightarrow \infty$, el lado izquierdo de la desigualdad converge a $1/m_j$ y el lado derecho converge a

$$\frac{P_{ji}^{(n_1)} P_{ij}^{(n_2)}}{m_i}.$$

Por lo tanto

$$\frac{1}{m_j} \geq \frac{P_{ji}^{(n_1)} P_{ij}^{(n_2)}}{m_i} > 0,$$

y en consecuencia $m_j < \infty$. Esto muestra que j es recurrente positivo. ■

A partir de este teorema y los resultados que vimos anteriormente, sabemos ahora que si $C \subset \mathcal{E}$ es un conjunto cerrado e irreducible, entonces o bien todos los estados de C son transitorios o todos son recurrentes nulos o todos son recurrentes positivos.

Si $C \subset \mathcal{E}$ es finito y cerrado, entonces tiene al menos un estado recurrente positivo: Como

$$\sum_{j \in C} P_{ij}^{(m)} = 1, \quad i \in C,$$

sumando sobre $m = 1, \dots, n$ y dividiendo por n obtenemos que

$$\sum_{j \in C} \frac{G_n(i, j)}{n} = 1, \quad i \in C.$$

Si C es finito y todos los estados de C fuesen transitorios o recurrentes nulos, entonces (3.8) o (3.14) valdrían y tendríamos

$$1 = \lim_{n \rightarrow \infty} \sum_{j \in C} \frac{G_n(i, j)}{n} = \sum_{j \in C} \lim_{n \rightarrow \infty} \frac{G_n(i, j)}{n} = 0,$$

lo cual es una contradicción.

Teorema 3.6 *Sea $C \subset \mathcal{E}$ finito, cerrado e irreducible. Entonces todos los estados de C son recurrentes positivos.*

Demostración. Como C es finito y cerrado, hay al menos un estado recurrente positivo. Como es irreducible, todos los estados se comunican y por el teorema 3.5 deben ser recurrentes positivos. ■

Corolario 3.1 *Una cadena de Markov irreducible con un número finito de estados es recurrente positiva.*

Corolario 3.2 *Una cadena de Markov con un número finito de estados no tiene estados recurrentes nulos.*

Demostración. Si j es un estado recurrente, está contenido en un conjunto cerrado e irreducible C de estados recurrentes. Como C es finito, por el teorema anterior todos los estados en C , incluyendo a j , son recurrentes positivos. Por lo tanto no hay estados recurrentes nulos. ■

3.4. Existencia y Unicidad de Distribuciones Estacionarias.

Vamos a necesitar la siguiente versión del Teorema de Convergencia Acotada para series.

Teorema 3.7 (Teorema de Convergencia Acotada) *Sea $a_i \geq 0$, $i \in \mathcal{E}$, una sucesión de números con $\sum a_i < \infty$ y sean $b_{i,n}$, $i \in \mathcal{E}$ y $n \geq 1$ tales que $|b_{i,n}| \leq 1$ para $i \in \mathcal{E}$ y $n \geq 1$, y*

$$\lim_{n \rightarrow \infty} b_{i,n} = b_i,$$

para todo $i \in \mathcal{E}$. Entonces

$$\lim_{n \rightarrow \infty} \sum_i a_i b_{i,n} = \sum_i a_i b_i.$$

Sea π una distribución estacionaria y sea $m \in \mathbb{N}$. Por la ecuación (3.2) tenemos

$$\sum_{k \in \mathcal{E}} \pi(k) P_{ki}^{(m)} = \pi(i), \quad i \in \mathcal{E}.$$

Sumando sobre $m = 1, \dots, n$ y dividiendo por n concluimos que

$$\sum_{k \in \mathcal{E}} \pi(k) \frac{G_n(k, i)}{n} = \pi(i), \quad i \in \mathcal{E}. \quad (3.14)$$

Teorema 3.8 *Sea π una distribución estacionaria. Si i es un estado transitorio o recurrente nulo, entonces $\pi(i) = 0$.*

Demostración. Si i es un estado transitorio o recurrente nulo,

$$\lim_{n \rightarrow \infty} \frac{G_n(k, i)}{n} = 0, \quad k \in \mathcal{E} \quad (3.15)$$

por (3.8) y (3.13). Por las ecuaciones (3.14) y (3.15) y el teorema 3.7,

$$\pi(i) = \lim_{n \rightarrow \infty} \sum_{k \in \mathcal{E}} \pi(k) \frac{G_n(k, i)}{n} = 0. \quad \blacksquare$$

Como consecuencia del teorema anterior vemos que una cadena sin estados recurrentes positivos no tiene una distribución estacionaria.

Teorema 3.9 *Una cadena de Markov irreducible y recurrente positiva tiene una única distribución estacionaria dada por*

$$\pi(i) = \frac{1}{m_i}, \quad i \in \mathcal{E}. \quad (3.16)$$

Demostración. Haremos la demostración para el caso en el cual el espacio de estados \mathcal{E} es finito. De las hipótesis del teorema y del teorema 3.4 vemos que

$$\lim_{n \rightarrow \infty} \frac{G_n(k, i)}{n} = \frac{1}{m_i}, \quad i, k \in \mathcal{E}. \quad (3.17)$$

Supongamos que π es una distribución estacionaria. A partir de la ecuaciones (3.14), (3.17) y el teorema 3.7 vemos que

$$\pi(i) = \lim_{n \rightarrow \infty} \sum_{k \in \mathcal{E}} \pi(k) \frac{G_n(k, i)}{n} = \frac{1}{m_i} \sum_{k \in \mathcal{E}} \pi(k) = \frac{1}{m_i}.$$

Por lo tanto, si el proceso tiene una distribución estacionaria, debe estar dada por (3.16).

Para completar la demostración del teorema tenemos que mostrar que la función $\pi(i)$, $i \in \mathcal{E}$ definida por (3.16) es una distribución estacionaria. Es claro que es no-negativa, así que sólo tenemos que verificar que

$$\sum_{i \in \mathcal{E}} \frac{1}{m_i} = 1 \quad (3.18)$$

y que

$$\sum_{i \in \mathcal{E}} \frac{1}{m_i} P_{ij} = \frac{1}{m_j}, \quad j \in \mathcal{E}. \quad (3.19)$$

Observemos inicialmente que

$$\sum_{i \in \mathcal{E}} P_{ki}^{(m)} = 1.$$

Sumando sobre $m = 1, \dots, n$ y dividiendo por n , concluimos que

$$\sum_{i \in \mathcal{E}} \frac{G_n(k, i)}{n} = 1, \quad k \in \mathcal{E}. \quad (3.20)$$

Si \mathcal{E} es finito, haciendo $n \rightarrow \infty$ en (3.20) y usando (3.17), obtenemos que

$$1 = \lim_{n \rightarrow \infty} \sum_{i \in \mathcal{E}} \frac{G_n(k, i)}{n} = \sum_{i \in \mathcal{E}} \frac{1}{m_i},$$

es decir, que (3.18) vale.

Por otro lado, por la ecuación de Chapman-Kolmogorov

$$\sum_{i \in \mathcal{E}} P_{ki}^{(m)} P_{ij} = P_{kj}^{(m+1)}.$$

Sumando de nuevo sobre $m = 1, \dots, n$ y dividiendo por n , obtenemos que

$$\sum_{i \in \mathcal{E}} \frac{G_n(k, i)}{n} P_{ij} = \frac{G_{n+1}(k, j)}{n} - \frac{P_{kj}}{n}. \quad (3.21)$$

Haciendo $n \rightarrow \infty$ en (3.21) concluimos que (3.19) vale. Esto completa la demostración cuando \mathcal{E} es finito. ■

A partir de los dos últimos teoremas obtenemos los siguientes corolarios.

Corolario 3.3 *Una cadena de Markov irreducible es recurrente positiva si y sólo si tiene una distribución estacionaria.*

Corolario 3.4 *Si una cadena de Markov con espacio de estados finito es irreducible, tiene una única distribución estacionaria.*

Corolario 3.5 *Consideremos una cadena de Markov irreducible, recurrente positiva con distribución estacionaria π . Entonces con probabilidad 1*

$$\lim_{n \rightarrow \infty} \frac{N_n(i)}{n} = \pi(i), \quad i \in \mathcal{E}. \quad (3.22)$$

Ejemplo 3.3 (Cadena con dos estados)

Consideremos una cadena de Markov con dos estados posibles y matriz de transición

$$P = \begin{pmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{pmatrix}$$

donde $0 < \alpha, \beta < 1$, $i = 1, 2$. Las ecuaciones para hallar la distribución estacionaria son

$$\begin{aligned} (1 - \alpha)\pi_1 + \beta\pi_2 &= \pi_1 \\ \alpha\pi_1 + (1 - \beta)\pi_2 &= \pi_2 \end{aligned}$$

que son la misma ecuación. Tenemos además la condición para que π sea una distribución de probabilidad: $\pi_1 + \pi_2 = 1$. La solución es

$$\pi_1 = \frac{\beta}{\alpha + \beta}, \quad \pi_2 = \frac{\alpha}{\alpha + \beta}.$$

que coincide con la distribución asintótica que hallamos en el ejemplo 3.2. ▲

3.5. Cadenas Reducibles

Definición 3.4 Sea π una distribución de probabilidad sobre \mathcal{E} y sea $C \subset \mathcal{E}$. Decimos que π está concentrada en C si $\pi(i) = 0$ siempre que $i \notin C$.

Los resultados que hemos demostrado anteriormente implican el siguiente teorema.

Teorema 3.10 *Sea C un conjunto cerrado e irreducible de estados recurrentes positivos. Entonces la cadena de Markov tiene una única distribución estacionaria π concentrada en C que está dada por*

$$\pi(i) = \begin{cases} \frac{1}{m_i}, & \text{si } i \in C, \\ 0, & \text{si no.} \end{cases} \quad (3.23)$$

Supongamos ahora que C_0 y C_1 son dos conjuntos distintos, cerrados e irreducibles de estados recurrentes positivos. Por el teorema anterior sabemos que la cadena tiene una distribución estacionaria π_0 concentrada en C_0 y otra distribución estacionaria π_1 concentrada en C_1 . Entonces, es posible demostrar que las distribuciones π_α , definidas para $0 \leq \alpha \leq 1$ por

$$\pi_\alpha(i) = (1 - \alpha)\pi_0(i) + \alpha\pi_1(i), \quad i \in \mathcal{E},$$

son distribuciones estacionarias distintas. Por lo tanto tenemos el siguiente resultado,

Corolario 3.6 Sea \mathcal{E}_P el conjunto de los estados recurrentes positivos de una cadena de Markov.

1. Si \mathcal{E}_P es vacío, la cadena no tiene ninguna distribución estacionaria.
2. Si \mathcal{E}_P es un conjunto irreducible no vacío, la cadena tiene una única distribución estacionaria.
3. Si \mathcal{E}_P no es vacío pero tampoco es irreducible, la cadena tiene un número infinito de distribuciones estacionarias distintas.

Ejemplo 3.4 (Cadena con dos estados)

Consideremos de nuevo la cadena de Markov con dos estados y matriz de transición

$$P = \begin{pmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{pmatrix}$$

y supongamos ahora que $\alpha = \beta = 0$, de modo que P es la matriz identidad. El espacio de estados tiene ahora dos conjuntos cerrados irreducibles: $\{0\}$ y $\{1\}$ y hay una distribución estacionaria concentrada en cada uno de ellos: Para $\{0\}$ es $(1, 0)$ mientras que para $\{1\}$ es $(0, 1)$. Cualquier combinación convexa de ellas es también una distribución estacionaria de la cadena y por lo tanto hay infinitas distribuciones estacionarias. ▲

3.6. Convergencia a la Distribución Estacionaria

Hasta ahora hemos visto que si X_n , $n \geq 0$ es una cadena de Markov irreducible y recurrente positiva con distribución estacionaria π , entonces

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=1}^n P_{ij}^{(m)} = \lim_{n \rightarrow \infty} \frac{G_n(i, j)}{n} = \pi(j), \quad i, j \in \mathcal{E}.$$

Estudiaremos en esta sección cuándo vale el resultado más fuerte

$$\lim_{n \rightarrow \infty} P_{ij}^{(n)} = \pi(j), \quad i, j \in \mathcal{E}$$

y qué ocurre cuando no vale.

Definición 3.5 Sea i un estado de la cadena con $P_{ii}^{(n)} > 0$ para algún $n \geq 1$, es decir, tal que $\rho_{ii} = P_i(T_i < \infty) > 0$. Llamemos $\mathcal{C}_i = \{n \geq 1 : P_{ii}^{(n)} > 0\}$. Definimos el *período* del estado i , d_i o $d(i)$ por

$$d_i = m.c.d. \{n \geq 1 : P_{ii}^{(n)} > 0\} = m.c.d. \mathcal{C}_i,$$

donde *m.c.d.* denota al máximo común divisor del conjunto.

Como consecuencia de la definición tenemos que

$$1 \leq d_i \leq \min \mathcal{C}_i.$$

y si $P_{ii} > 0$, entonces $d_i = 1$.

Lema 3.2 Si i y j son dos estados que se comunican, entonces $d_i = d_j$.

Demostración. Para ver esto sean n_1 y n_2 enteros positivos tales que

$$P_{ij}^{(n_1)} > 0 \quad \text{y} \quad P_{ji}^{(n_2)} > 0.$$

Entonces

$$P_{ii}^{(n_1+n_2)} \geq P_{ij}^{(n_1)} P_{ji}^{(n_2)} > 0,$$

y por lo tanto d_i divide a $n_1 + n_2$. Si $n \in \mathcal{C}_j$ tenemos que $P_{jj}^n > 0$ y en consecuencia

$$P_{ii}^{(n_1+n+n_2)} \geq P_{ij}^{(n_1)} P_{jj}^n P_{ji}^{(n_2)} > 0,$$

de modo que d_i es divisor de $n_1 + n + n_2$. Como d_i es divisor de $n_1 + n_2$ también debe ser divisor de n . Por lo tanto d_i es divisor de todos los números en el conjunto \mathcal{C}_j . Como d_j es el mayor de todos esos divisores, concluimos que $d_i \leq d_j$. De manera similar se muestra que $d_j \leq d_i$ y en consecuencia $d_1 = d_j$. ■

Hemos mostrado que los estados en una cadena de Markov irreducible tienen período común d .

Definición 3.6 Decimos que una cadena irreducible es *periódica* con período d si $d > 1$ y *aperiódica* si $d = 1$.

Una condición suficiente sencilla para que una cadena irreducible sea aperiódica es que $P_{ii} > 0$ para algún $i \in \mathcal{E}$.

Ejemplo 3.5

Consideremos una cadena con espacio de estados $\mathcal{E} = \{-2, -1, 0, 1, 2\}$ y matriz de transición

	-2	-1	0	1	2	3
-2	0	0	1	0	0	0
-1	1	0	0	0	0	0
0	0	0.5	0	0.5	0	0
1	0	0	0	0	1	0
2	0	0	0	0	0	1
3	0	0	1	0	0	0

Veamos el diagrama de transiciones para esta cadena

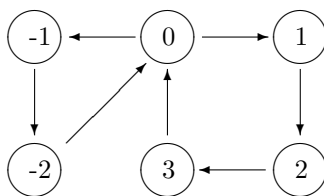


Figura 3.1

Considerando el estado 0, hay dos maneras de regresar a él: $0 \rightarrow -1 \rightarrow -2 \rightarrow 0$, que requiere tres pasos, y $0 \rightarrow 1 \rightarrow 2 \rightarrow 3 \rightarrow 0$, que requiere 4. Por lo tanto 3 y 4 están en \mathcal{C}_0 y el m.c.d. de este conjunto es 1. En consecuencia esta cadena es aperiódica. ▲

Ejemplo 3.6 (Paseo al azar simple con barreras reflectoras)

Consideremos un paseo al azar simple con barreras reflectoras en los extremos 0 y $N = 4$. La matriz de transición es

$$P = \begin{pmatrix} 0.5 & 0.5 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 \\ 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 0.5 & 0.5 \end{pmatrix}$$

Vemos que todos los estados de esta cadena se comunican y que $P_{00} > 0$, de modo que 0 tiene período 1 y, en consecuencia, todos los otros estados también. ▲

Teorema 3.11 *Sea X_n , $n \geq 0$, una cadena de Markov irreducible y recurrente positiva con distribución estacionaria π . Si la cadena es aperiódica,*

$$\lim_{n \rightarrow \infty} P_{ij}^{(n)} = \pi(j), \quad i, j \in \mathcal{E}. \quad (3.24)$$

Si la cadena es periódica con período d , entonces para cada par de estados i, j en \mathcal{E} existe un entero r , $0 \leq r < d$, tal que $P_{ij}^{(n)} = 0$ a menos que $n = md + r$ para algún $m \in \mathbb{N}$ y

$$\lim_{m \rightarrow \infty} P_{ij}^{(md+r)} = d\pi(j), \quad i, j \in \mathcal{E}. \quad (3.25)$$

Demostración. Ver Hoel, Port & Stone, pags. 75-79. ■

Ejemplo 3.7

En el problema 7 de la lista de problemas 7 se pedía calcular las potencias 2, 4, 8, 16, 17, 32 y 33 de la matriz

$$P = \begin{pmatrix} 0.4 & 0.6 & 0 \\ 0.2 & 0.5 & 0.3 \\ 0.1 & 0.7 & 0.2 \end{pmatrix}$$

La última de estas potencias es

$$P^{33} = \begin{pmatrix} 0.22353 & 0.56471 & 0.21177 \\ 0.22353 & 0.56471 & 0.21177 \\ 0.22353 & 0.56471 & 0.21177 \end{pmatrix}$$

Calculemos ahora la distribución estacionaria para esta matriz, resolviendo el sistema de ecuaciones $\pi P = \pi$:

$$-0.6\pi_1 + 0.2\pi_2 + 0.1\pi_3 = 0$$

$$0.6\pi_1 - 0.5\pi_2 + 0.7\pi_3 = 0$$

$$0.3\pi_2 - 0.8\pi_3 = 0$$

y la ecuación adicional $\pi_1 + \pi_2 + \pi_3 = 1$, obtenemos

$$\pi_1 = \frac{19}{85} = 0.22353; \quad \pi_2 = \frac{48}{85} = 0.56471; \quad \pi_3 = \frac{18}{85} = 0.21177.$$

Vemos que en este ejemplo no sólo hay convergencia a la distribución estacionaria, sino que esta convergencia ocurre rápidamente. ▲

3.7. Invertibilidad

Sea $\{X_n, n \geq 0\}$ una cadena de Markov homogénea con matriz de transición P y distribución estacionaria π tal que

$$\pi(i) > 0, \quad \forall i \in \mathcal{E}.$$

Definimos la matriz Q con índices en \mathcal{E} por

$$\pi(i)Q_{ij} = \pi(j)P_{ji}. \quad (3.26)$$

Al igual que P , esta matriz también es estocástica ya que

$$\sum_{j \in \mathcal{E}} Q_{ij} = \sum_{j \in \mathcal{E}} \frac{\pi(j)}{\pi(i)} P_{ji} = \frac{1}{\pi(i)} \sum_{j \in \mathcal{E}} \pi(j) P_{ji} = \frac{\pi(i)}{\pi(i)} = 1.$$

La interpretación de esta matriz es la siguiente. Supongamos que la distribución inicial de la cadena es π . En este caso el proceso es estacionario y para todo $n \geq 0$ y todo $i \in \mathcal{E}$

$$P(X_n = i) = \pi(i).$$

Usando la definición de la probabilidad condicional tenemos

$$\begin{aligned} P(X_n = j | X_{n+1} = i) &= \frac{P(X_{n+1} = i | X_n = j)P(X_n = j)}{P(X_{n+1} = i)} \\ &= \frac{\pi(j)P_{ji}}{\pi(i)} = Q_{ij} \end{aligned}$$

Por lo tanto Q es la matriz de transición de la cadena cuando invertimos el sentido del tiempo.

Teorema 3.12 *Sea $X_n, n \geq 0$ una cadena de Markov con matriz de transición P y espacio de estados \mathcal{E} y sea π una distribución de probabilidad sobre \mathcal{E} . Sea Q una matriz estocástica con índices en \mathcal{E} tal que para todo $i, j \in \mathcal{E}$*

$$\pi(i)Q_{ij} = \pi(j)P_{ji} \quad (3.27)$$

Entonces π es una distribución estacionaria para la cadena.

Demostración. Fijamos $i \in \mathcal{E}$ y sumamos (3.27) sobre $j \in \mathcal{E}$:

$$\sum_{j \in \mathcal{E}} \pi(i)Q_{ij} = \sum_{j \in \mathcal{E}} \pi(j)P_{ji}$$

Pero $\sum_j Q_{ij} = 1$ y por lo tanto, para todo $i \in \mathcal{E}$

$$\pi(i) = \sum_{j \in \mathcal{E}} \pi(j)P_{ji}$$

que es la condición para una distribución estacionaria. ■

Definición 3.7 *Sea $\{X_n, n \geq 0\}$ una cadena de Markov homogénea con distribución estacionaria π tal que $\pi(i) > 0$ para todo $i \in \mathcal{E}$. Supongamos que π es la distribución inicial de la cadena. Decimos que $\{X_n, n \geq 0\}$ es invertible si se satisfacen las ecuaciones de balance detallado:*

$$\pi(i)P_{ij} = \pi(j)P_{ji}, \quad \forall i, j \in \mathcal{E}. \quad (3.28)$$

En este caso $Q_{ij} = P_{ij}$ y por lo tanto la cadena original y la cadena invertida en el tiempo tienen la misma distribución ya que la distribución de una cadena de Markov homogénea está determinada por su distribución inicial y su matriz de transición.

Otra manera de expresar la condición de balance detallado (3.28) es

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

Corolario 3.7 (Prueba de Balance Detallado) *Sea P una matriz de transición sobre \mathcal{E} y sea π una distribución de probabilidad sobre \mathcal{E} . Si las ecuaciones de balance detallado (3.28) son válidas para todo $i, j \in \mathcal{E}$ entonces π es una distribución estacionaria para la cadena.*

Una cadena invertible tiene la misma distribución si invertimos la dirección del tiempo, es decir, $(X_0, X_1, \dots, X_{n-1}, X_n)$ tiene la misma distribución de probabilidad que $(X_n, X_{n-1}, \dots, X_1, X_0)$ para todo n . En particular tenemos el siguiente corolario

Corolario 3.8 *Una cadena invertible es estacionaria.*

Ejemplo 3.8 (Paseo al azar sobre grafos.)

Un grafo es una estructura compuesta por dos partes: Un conjunto de vértices o nodos \mathcal{V} , que supondremos finito, y una matriz de estructura $A = A(i, j)$ en la cual $A(i, j)$ vale 1 si i y j están conectados por un arco o arista (diremos que i y j son vecinos) y 0 si no. Por convención ponemos $A(i, i) = 0$ para todo $i \in \mathcal{V}$.

El grado de un vértice i es igual al número de vecinos que tiene:

$$d(i) = \sum_{j \in \mathcal{V}} A(i, j)$$

ya que cada vecino contribuye una unidad a la suma. Por lo tanto

$$P_{ij} = \frac{A(i, j)}{d(i)}$$

define una probabilidad de transición. Si $X_n = i$, la cadena salta a alguno de los vecinos de i con distribución uniforme. La cadena de Markov homogénea con estas probabilidades de transición se conoce como un paseo al azar sobre el grafo.

Por ejemplo, en el grafo de la figura 3.2

$$P_{0,i} = \frac{1}{4}, \quad i = 1, 2, 3, 4; \quad P_{10} = P_{12} = \frac{1}{2}; \quad P_{20} = P_{21} = P_{23} = \frac{1}{3}; \quad P_{30} = P_{32} = \frac{1}{2}; \quad P_{40} = 1$$

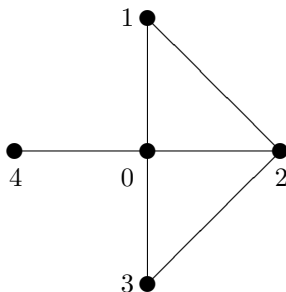


Figura 3.2

Es fácil hallar la distribución estacionaria en este caso. Vemos que si C es una constante positiva, $\pi(i) = Cd(i)$ satisface la condición de balance detallado:

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

Por lo tanto, si tomamos $C = 1/\sum_i d(i)$, tenemos una distribución estacionaria. Observamos que para cualquier grado, $\sum_i d(i) = 2|\mathcal{A}|$, donde $|\mathcal{A}|$ es el número de arcos, y por lo tanto $C = 1/2|\mathcal{A}|$. ▲

3.8. Teorema Ergódico

Una versión más general de la ley fuerte de grandes números para cadenas de Markov es el Teorema Ergódico, que presentamos a continuación.

Teorema 3.13 (Teorema Ergódico) *Sea $\{X_n, n \geq 0\}$ una cadena de Markov irreducible, recurrente positiva y con distribución estacionaria π . Sea $f : \mathcal{E} \rightarrow \mathbb{R}$, una función acotada, entonces,*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n f(X_j) = \sum_{i \in \mathcal{E}} \pi(i) f(i) = E_\pi(f), \quad c. p. 1. \quad (3.29)$$

Demostración. Supondremos, sin pérdida de generalidad, que $|f| \leq 1$. Tenemos que

$$\frac{1}{n} \sum_{m=1}^n f(X_m) = \sum_{i \in \mathcal{E}} \frac{N_n(i)}{n} f(i).$$

Por lo tanto

$$\left| \frac{1}{n} \sum_{m=1}^n f(X_m) - \sum_{i \in \mathcal{E}} \pi(i) f(i) \right| = \left| \sum_{i \in \mathcal{E}} \left(\frac{N_n(i)}{n} - \pi(i) \right) f(i) \right|$$

Para cualquier $S \subset \mathcal{E}$ tenemos

$$\begin{aligned} \left| \frac{1}{n} \sum_{m=1}^n f(X_m) - \sum_{i \in \mathcal{E}} \pi(i) f(i) \right| &= \left| \left(\sum_{i \in S} + \sum_{i \notin S} \right) \left(\frac{N_n(i)}{n} - \pi(i) \right) f(i) \right| \\ &\leq \sum_{i \in S} \left| \frac{N_n(i)}{n} - \pi(i) \right| + \sum_{i \notin S} \left| \frac{N_n(i)}{n} - \pi(i) \right| \end{aligned} \quad (3.30)$$

Ahora bien, el segundo término en (3.30) es

$$\begin{aligned} \sum_{i \notin S} \left| \frac{N_n(i)}{n} - \pi(i) \right| &\leq \sum_{i \notin S} \frac{N_n(i)}{n} + \sum_{i \notin S} \pi(i) = 1 - \sum_{i \in S} \frac{N_n(i)}{n} + \sum_{i \notin S} \pi(i) \\ &= \sum_{i \in S} \pi(i) - \sum_{i \in S} \frac{N_n(i)}{n} + 2 \sum_{i \notin S} \pi(i) \\ &\leq \sum_{i \in S} \left| \pi(i) - \frac{N_n(i)}{n} \right| + 2 \sum_{i \notin S} \pi(i) \end{aligned}$$

y por lo tanto

$$\left| \frac{1}{n} \sum_{m=1}^n f(X_m) - \sum_{i \in \mathcal{E}} \pi(i) f(i) \right| \leq 2 \sum_{i \in S} \left| \pi(i) - \frac{N_n(i)}{n} \right| + 2 \sum_{i \notin S} \pi(i)$$

Dado $\varepsilon > 0$ escogemos $S \subset \mathcal{E}$ finito de modo que $\sum_{i \notin S} \pi(i) < \varepsilon/4$ y luego escogemos $N(\omega)$ tal que, para todo $n \geq N(\omega)$,

$$\sum_{i \in S} \left| \pi(i) - \frac{N_n(i)}{n} \right| < \frac{\varepsilon}{4}.$$

Por lo tanto para $n \geq N(\omega)$ tenemos

$$\left| \frac{1}{n} \sum_{m=1}^n f(X_m) - \sum_{i \in \mathcal{E}} \pi(i) f(i) \right| \leq \varepsilon.$$

■

3.9. Ejemplos

Definición 3.8 Una matriz de transición P es *doblemente estocástica* si sus columnas suman 1, es decir, si

$$\sum_{i \in \mathcal{E}} P_{ij} = 1, \quad \text{para todo } j \in \mathcal{E}.$$

Para una matriz de transición doblemente estocástica, la distribución estacionaria es sencilla.

Teorema 3.14 Si la matriz de transición P de una cadena de Markov con N estados es doblemente estocástica, entonces la distribución uniforme $\pi(i) = 1/N$ para todo i , es una distribución estacionaria.

Demostración. Observamos que

$$\sum_{i \in \mathcal{E}} \pi(i) P_{ij} = \frac{1}{N} \sum_{i \in \mathcal{E}} P_{ij} = \frac{1}{N}$$

de modo que la distribución uniforme satisface la condición $\pi'P = \pi$ que define una distribución estacionaria.

Vemos además, que si la distribución estacionaria es uniforme, necesariamente la matriz P es doblemente estocástica. ■

Ejemplo 3.9 (Paseo al azar simple con barreras reflectoras)

Consideremos de nuevo el paseo al azar simple simétrico con barreras reflectoras (ver ejemplo 3.6). Es inmediato en el caso particular $N = 4$ considerado antes que la matriz es doblemente estocástica, y en consecuencia $\pi(i) = 1/5$ es una distribución estacionaria. En general, si consideramos un paseo de este tipo con espacio de estados $\mathcal{E} = \{0, 1, \dots, N\}$, la distribución estacionaria será $\pi(i) = 1/(N+1)$. ▲

Ejemplo 3.10 (Paseo al azar en la circunferencia)

Colocamos $N+1$ puntos, que llamamos $0, 1, \dots, N$ sobre la circunferencia. En cada paso la cadena se mueve a la derecha o a la izquierda un paso, con probabilidades respectivas p y $1-p$, incluyendo los extremos 0 y N , es decir, la cadena pasa de N a 0 con probabilidad p y de 0 a N con probabilidad $1-p$. Para el caso particular $N = 4$ la matriz de transición es

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

Vemos que todas las columnas suman 1, y lo mismo es cierto en el caso general, de modo que la distribución estacionaria es uniforme $\pi(i) = 1/(N+1)$. ▲

En el siguiente ejemplo usamos las ecuaciones de balance detallado para encontrar la distribución estacionaria.

Ejemplo 3.11

Tres bolas blancas y tres negras se colocan en dos cajas de modo que cada caja contenga tres bolas. En cada paso extraemos una bola de cada caja y las intercambiamos. X_n es el número de bolas blancas en la caja de la izquierda al paso n . Halle la matriz de transición y obtenga la distribución estacionaria. Demuestre que ésta corresponde a seleccionar 3 bolas al azar para colocarlas en la caja de la izquierda.

X_n es una cadena de Markov homogénea con $\mathcal{E} = \{0, 1, 2, 3\}$. Si $X_n = i$, hay $3 - i$ bolas negras en la caja de la izquierda mientras que la derecha tiene $3 - i$ blancas e i negras. Por lo tanto

$$\begin{aligned} P_{i,i+1} &= P(X_{n+1} = i + 1 | X_n = i) \\ &= P(\text{seleccionar blanca en la derecha y negra en la izquierda}) \\ &= \left(\frac{3-i}{3}\right)^2, \quad \text{siempre que } i < 3, \\ P_{i,i-1} &= \left(\frac{i}{3}\right)^2, \quad \text{siempre que } i > 0, \\ P_{i,i} &= 1 - \left(\frac{3-i}{3}\right)^2 - \left(\frac{i}{3}\right)^2. \end{aligned}$$

Tenemos

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1/9 & 4/9 & 4/9 & 0 \\ 0 & 4/9 & 4/9 & 1/9 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

Para hallar la distribución estacionaria usamos las ecuaciones de balance detallado:

$$\pi(i)P_{i,i-1} = \pi(i-1)P_{i-1,i}$$

es decir,

$$\pi(i)\left(\frac{i}{3}\right)^2 = \pi(i-1)\left(\frac{3-(i-1)}{3}\right)^2$$

de donde obtenemos

$$\pi(i) = \left(\frac{4-i}{i}\right)^2 \pi(i-1).$$

Usando esta relación obtenemos

$$\pi(1) = 9\pi(0); \quad \pi(2) = 9\pi(0); \quad \pi(3) = \pi(0)$$

y

$$\pi(0) + \pi(1) + \pi(2) + \pi(3) = 1$$

Finalmente

$$\pi = \frac{1}{20}(1, 9, 9, 1).$$

Veamos que la distribución que obtuvimos corresponde al número de bolas blancas que se obtienen al seleccionar tres bolas de un conjunto de seis bolas, de las cuales tres son blancas y tres negras. En este contexto, la probabilidad de seleccionar i bolas blancas es

$$\frac{\binom{3}{i}\binom{3}{3-i}}{\binom{6}{3}} = \frac{1}{20}\binom{3}{i}^2$$

que es la distribución π .

▲

3.9.1. Cadenas de Nacimiento y Muerte

En el capítulo anterior consideramos las cadenas de nacimiento y muerte. A continuación queremos obtener una condición que nos permita determinar, en el caso de una cadena irreducible con espacio de estados infinito, cuándo la cadena es transitoria y cuándo es recurrente. Consideraremos cadenas de nacimiento y muerte irreducibles sobre los enteros no-negativos, por lo tanto

$$p_i > 0 \quad \text{para } i \geq 0, \quad q_i > 0 \quad \text{para } i \geq 1.$$

Hemos visto que

$$P_1(T_0 < \infty) = 1 - \frac{1}{\sum_{j=0}^{\infty} \gamma_j}, \quad (3.31)$$

con $\gamma_0 = 1$ y $\gamma_j = \prod_{i=1}^j (q_i/p_i)$, $j \geq 1$. Supongamos ahora que la cadena es recurrente, entonces $P_1(T_0 < \infty) = 1$ y necesariamente

$$\sum_{j=0}^{\infty} \gamma_j = \infty. \quad (3.32)$$

Para ver que esta condición también es suficiente, observemos que $P_{0j} = 0$ para $j \geq 2$ y en consecuencia

$$P_0(T_0 < \infty) = P_{00} + P_{01}P_1(T_0 < \infty). \quad (3.33)$$

Supongamos que (3.32) vale, por (3.31)

$$P_1(T_0 < \infty) = 1,$$

y usando esto en (3.33) concluimos que

$$P_0(T_0 < \infty) = P_{00} + P_{01} = 1,$$

de modo que 0 es un estado recurrente. Como la cadena es irreducible, debe ser una cadena recurrente.

Resumiendo, hemos mostrado que una cadena de nacimiento y muerte irreducible sobre $\{0, 1, 2, \dots\}$ es recurrente sí y sólo sí

$$\sum_{j=0}^{\infty} \gamma_j = \sum_{j=0}^{\infty} \frac{q_1 \cdots q_j}{p_1 \cdots p_j} = \infty.$$

Ejemplo 3.12

Consideremos la cadena de nacimiento y muerte sobre los enteros no negativos con probabilidades de transición

$$p_i = \frac{i+2}{2(i+1)}, \quad y \quad q_i = \frac{i}{2(i+1)}, \quad i \geq 0.$$

En este caso

$$\frac{q_i}{p_i} = \frac{i}{i+2},$$

y obtenemos que

$$\gamma_i = \frac{q_1 \cdots q_i}{p_1 \cdots p_i} = \frac{1 \cdot 2 \cdots i}{3 \cdot 4 \cdots (i+2)} = \frac{2}{(i+1)(i+2)} = 2 \left(\frac{1}{i+1} - \frac{1}{i+2} \right).$$

Por lo tanto,

$$\sum_{i=1}^{\infty} \gamma_i = 2 \sum_{i=1}^{\infty} \left(\frac{1}{i+1} - \frac{1}{i+2} \right) = 2 \left(\frac{1}{2} - \frac{1}{3} + \frac{1}{3} - \frac{1}{4} + \frac{1}{4} - \frac{1}{5} + \dots \right) = 1,$$

y concluimos que la cadena es transitoria. ▲

Finalmente, veamos cual es la distribución estacionaria para una cadena irreducible con espacio de estados infinito. Las ecuaciones (3.1) que definen la distribución estacionaria, son en este caso

$$\begin{aligned} \pi(0)r_0 + \pi(1)q_1 &= \pi(0), \\ \pi(i-1)p_{i-1} + \pi(i)r_i + \pi(i+1)q_{i+1} &= \pi(i), \quad i \geq 1, \end{aligned}$$

y teniendo en cuenta la relación $p_i + r_i + q_i = 1$, las ecuaciones anteriores son

$$\begin{aligned} p_0\pi(0) &= q_1\pi(1), \\ q_{i+1}\pi(i+1) - p_i\pi(i) &= q_i\pi(i) - p_{i-1}\pi(i-1), \quad i \geq 1. \end{aligned}$$

A partir de estas ecuaciones obtenemos por inducción que

$$q_{i+1}\pi(i+1) = p_i\pi(i), \quad i \geq 0. \quad (3.34)$$

Esta ecuación es un caso particular de la ecuación de balance detallado

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

para el caso de las cadenas de nacimiento y muerte. La condición de balance detallado es más fuerte que (3.1), como es fácil de verificar, y no siempre es válida. La ecuación (3.34) también vale en el caso de un espacio de estados finito.

Ejemplo 3.13 (Cadena de Ehrenfest con tres estados)

En este caso la matriz de transición es

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

y vemos que para todo i , $r_i = 0$. Las ecuaciones (3.34) son en este caso

$$1 \cdot \pi(0) = \frac{1}{3}\pi(1); \quad \frac{2}{3}\pi(1) = \frac{2}{3}\pi(2); \quad \frac{1}{3}\pi(2) = 1 \cdot \pi(3).$$

Poniendo $\pi(0) = \lambda$ y resolviendo obtenemos $\pi(1) = \pi(2) = 3\lambda$, $\pi(3) = \lambda$. Como la suma debe ser 1, obtenemos que $\lambda = 1/8$ y la distribución estacionaria en este caso es

$$\pi(0) = \frac{1}{8}, \quad \pi(1) = \frac{3}{8}, \quad \pi(2) = \frac{3}{8}, \quad \pi(3) = \frac{1}{8}. \quad \blacktriangle$$

Veamos ahora cómo se obtiene la distribución estacionaria para una cadena general de nacimiento y muerte. A partir de la ecuación (3.34) obtenemos

$$\pi(i+1) = \frac{p_i}{q_{i+1}}\pi(i), \quad i \geq 0 \quad (3.35)$$

y en consecuencia

$$\pi(i) = \frac{p_0 \cdots p_{i-1}}{q_1 \cdots q_i} \pi(0), \quad i \geq 1. \quad (3.36)$$

Definamos $\nu_0 = 1$ y

$$\nu_i = \frac{p_0 \cdots p_{i-1}}{q_1 \cdots q_i}, \quad i \geq 1, \quad (3.37)$$

entonces podemos escribir (3.36) como

$$\pi(i) = \nu_i \pi(0), \quad i \geq 0. \quad (3.38)$$

Supongamos ahora que

$$\sum_i \nu_i = \sum_{i=1}^{\infty} \frac{p_0 \cdots p_{i-1}}{q_1 \cdots q_i} < \infty, \quad (3.39)$$

a partir de (3.38) concluimos que la cadena tiene una única distribución estacionaria dada por

$$\pi(i) = \frac{\nu_i}{\sum_{j=0}^{\infty} \nu_j}, \quad i \geq 1. \quad (3.40)$$

Si en cambio (3.39) no vale, es decir, si la serie diverge, la relación (3.38) dice que la solución de (3.1) es idénticamente igual a 0 (si $\pi(0) = 0$) o tiene suma infinita (si $\pi(0) > 0$) y en consecuencia no existe distribución estacionaria.

Vemos que una cadena de nacimiento y muerte tiene distribución estacionaria si y sólo si (3.39) vale, y que la distribución estacionaria, cuando existe, está dada por (3.37) y (3.40). Teniendo en cuenta que la cadena es irreducible, observamos que la distribución estacionaria existe si y sólo si la cadena es recurrente positiva.

Resumiendo, podemos dar condiciones necesarias y suficientes para cada una de las tres posibilidades en una cadena de nacimiento y muerte sobre los enteros no negativos.

- La cadena es transitoria sí y sólo sí

$$\sum_{j=0}^{\infty} \gamma_j < \infty.$$

- La cadena es recurrente positiva sí y sólo sí

$$\sum_{j=0}^{\infty} \gamma_j = \infty, \quad \sum_{j=0}^{\infty} \nu_j < \infty.$$

- La cadena es recurrente nula sí y sólo sí

$$\sum_{j=0}^{\infty} \gamma_j = \infty, \quad \sum_{j=0}^{\infty} \nu_j = \infty.$$

Ejemplo 3.14

Consideremos una cadena de nacimiento y muerte sobre los enteros no negativos con las siguientes probabilidades de transición

$$p_0 = 1, \quad p_i = p, \quad q_i = q = 1 - p, \quad i \geq 1,$$

con $0 < p < 1$. Para determinar en cuál de las tres clases se encuentra la cadena tenemos que estudiar el comportamiento de las series $\sum \gamma_i$ y $\sum \nu_i$. En este caso es fácil ver que

$$\gamma_i = \left(\frac{q}{p}\right)^i, \quad \nu_i = \frac{p^{i-1}}{q^i}$$

para $i \geq 1$. En consecuencia vemos que hay tres casos:

- $0 < p < 1/2$. $\sum \gamma_i = \infty$, por lo tanto la cadena es recurrente. Para ver si es nula o positiva calculamos $\sum \nu_i$, que es convergente en este caso y en consecuencia la cadena es recurrente positiva.
- $p = 1/2$. Un análisis similar muestra que ambas series divergen y la cadena es recurrente nula.
- $1/2 < p < 1$. En este caso $\sum \gamma_i < \infty$ y la cadena es transitoria.

▲

Ejemplo 3.15

Consideremos una empresa que tiene tres máquinas que se dañan de manera independiente, con probabilidad 0.1 cada día. Cuando hay al menos una máquina dañada, con probabilidad 0.5 el técnico puede reparar una de ellas para que esté funcionando el próximo día. Para simplificar, suponemos que es imposible que dos máquinas se dañen el mismo día. El número de máquinas en funcionamiento en un día dado puede ser modelado como una cadena de nacimiento y muerte con la siguiente matriz de transición:

$$P = \begin{pmatrix} 0.5 & 0.5 & 0 & 0 \\ 0.05 & 0.5 & 0.45 & 0 \\ 0 & 0.1 & 0.5 & 0.4 \\ 0 & 0 & 0.3 & 0.7 \end{pmatrix}$$

Para ver como se obtiene esta matriz, consideremos la segunda fila, que corresponde a un día que se inicia con una máquina en buen estado. Al día siguiente estaremos en el estado 0 si una máquina se daña y el técnico no puede arreglar la máquina en la que está trabajando, lo cual ocurre con probabilidad 0.1×0.5 . Por otro lado, pasamos al estado 2 sólo si el técnico logra reparar la máquina en la que está trabajando y la que está en uso no se daña. Esto ocurre con probabilidad $(0.5)(0.9)$. Un razonamiento similar muestra que $P_{21} = (0.2)(0.5)$ y $P_{23} = (0.5)(0.8)$.

Para obtener la distribución estacionaria usamos la fórmula (3.35) y poniendo $\pi(0) = \lambda$ entonces

$$\begin{aligned} \pi(1) &= \pi(0) \frac{p_0}{q_1} = \lambda \frac{0.5}{0.05} = 10\lambda, \\ \pi(2) &= \pi(1) \frac{p_1}{q_2} = 10\lambda \frac{0.45}{0.1} = 45\lambda, \\ \pi(3) &= \pi(2) \frac{p_2}{q_3} = 45\lambda \frac{0.4}{0.3} = 60\lambda. \end{aligned}$$

La suma de las π es 116λ , haciendo $\lambda = 1/116$ obtenemos

$$\pi(3) = \frac{60}{116}, \quad \pi(2) = \frac{45}{116}, \quad \pi(1) = \frac{10}{116}, \quad \pi(0) = \frac{1}{116}.$$

▲

Ejemplo 3.16 (Rachas)

Sea $\{X_n, n \geq 0\}$ una cadena de Markov sobre $\{0, 1, 2, \dots\}$ con probabilidades de transición

$$P_{i,i+1} = p_i, \quad P_{i,0} = 1 - p_i.$$

¿Bajo que condiciones la matriz de transición de X admite alguna medida invariante?

Para responder a la pregunta planteada arriba estudiaremos cuando el sistema $\pi'P = \pi$, es decir,

$$\pi_0 = \sum_{i \geq 0} (1 - p_i) \pi_i, \quad \pi_i = p_{i-1} \pi_{i-1}, \quad \forall i \geq 1,$$

tiene solución no trivial. Usando un argumento de recursión vemos que el sistema anterior equivale a

$$\pi_i = \left(\prod_{j=0}^{i-1} p_j \right) \pi_0, \quad \forall i \geq 1,$$

y

$$\pi_0 = (1 - p_0) \pi_0 + \sum_{i \geq 1} (1 - p_i) \left(\prod_{j=0}^{i-1} p_j \right) \pi_0.$$

Esta última ecuación es la que nos permitirá estudiar la existencia de vectores invariantes. Para evitar casos no interesantes supondremos de aquí en adelante que $0 < \pi_0 < \infty$. Entonces, se tiene la siguiente sucesión de igualdades:

$$\begin{aligned} \pi_0 &= (1 - p_0) \pi_0 + \pi_0 \sum_{i \geq 1} (1 - p_i) \left(\prod_{j=0}^{i-1} p_j \right) \\ &= (1 - p_0) \pi_0 + \pi_0 \lim_{n \rightarrow \infty} \sum_{i=1}^n (1 - p_i) \left(\prod_{j=0}^{i-1} p_j \right) \\ &= (1 - p_0) \pi_0 + \pi_0 \lim_{n \rightarrow \infty} \sum_{i=1}^n \left[\left(\prod_{j=0}^{i-1} p_j \right) - p_i \left(\prod_{j=0}^{i-1} p_j \right) \right] \\ &= (1 - p_0) \pi_0 + \pi_0 \left(p_0 - \lim_{n \rightarrow \infty} \prod_{j=0}^{n-1} p_j \right) \end{aligned}$$

Deducimos de esto que una condición necesaria para que π_0 sea $0 < \pi_0 < \infty$ es que

$$\prod_{i=0}^{\infty} p_i = 0.$$

Pero eso no es suficiente para garantizar la existencia del vector de probabilidad π , ya que necesitamos saber cuando

$$\sum_{i=0}^{\infty} \pi_i < \infty,$$

y esto ocurre si y solamente si

$$\sum_{n=1}^{\infty} \prod_{i=0}^n p_i < \infty.$$

Si esto es cierto, para que π sea una distribución de probabilidad es necesario que

$$\pi_0 + \pi_0 \sum_{n=1}^{\infty} \left(\prod_{i=0}^n p_i \right) = 1$$

y por lo tanto,

$$\pi_0 = \frac{1}{1 + \sum_{n=0}^{\infty} \prod_{j=0}^n p_j}, \quad \pi_j = \pi_0 \prod_{i=0}^{j-1} p_i, \quad j \geq 1.$$

Podemos concluir que la cadena es positiva recurrente si y solamente si

$$\sum_{n=0}^{\infty} \left(\prod_{i=0}^n p_i \right) < \infty.$$

Mientras que en el caso en que

$$\lim_{n \rightarrow \infty} \prod_{j=0}^n p_j = p \in (0, 1],$$

no existe vector invariante diferente del vector $\mathbf{0}$. ▲

Ejemplo 3.17

Supongamos que en el ejercicio anterior todas las $\{p_i, i \geq 0\}$, son iguales a un valor $p \in (0, 1)$. Calcular las n -ésimas potencias de la matriz de transición P . Estudiar el comportamiento asintótico de estas cuando $n \rightarrow \infty$.

Observemos que la cadena de Markov que nos interesa se puede usar para simular la fortuna de un jugador muy avaricioso, que apuesta toda su fortuna cada vez que juega en un juego que le permite ganar un peso con probabilidad p o perder todo el dinero apostado con probabilidad $1 - p$; y en compensación el casino le da crédito, en el sentido que le da la oportunidad de seguir jugando aunque su fortuna sea cero. Denotaremos por X_n la fortuna del jugador al tiempo n ; se tiene que en un paso la cadena se comporta como

$$X_{n+1} = \begin{cases} X_n + 1 & \text{con probabilidad } p \\ 0 & \text{con probabilidad } 1 - p = q, \end{cases}$$

mientras que en n -pasos, se tienen dos posibilidades, sea el jugador no pierde ni una sola vez, o bien pierde alguna vez. Esto se refleja en las probabilidades de transición como

$$P_{i,k}^{(n)} = p^n, \quad \text{si } k = i + n,$$

esto ocurre cuando el jugador no pierde, mientras que si el jugador pierde en alguno de los n juegos:

$$P_{i,k}^{(n)} = qp^k, \quad 0 \leq k \leq n - 1.$$

La primera afirmación es evidente, para ver la segunda veamos primero como calcular esa probabilidad en el caso en que $k = 0$, por la ecuación de Chapman-Kolmogorov,

$$P_{i,0}^{(n)} = \sum_{z \geq 0} P_{i,z}^{(n-1)} P_{z,0} = q \sum_{z \geq 0} P_{i,z}^{(n-1)} = q, \quad \forall n \geq 2,$$

ahora bien, el evento $\{X_n = k\}$ dado que $X_0 = i$, ocurre cuando hay una “racha” de k juegos ganados, antecidos de una sucesión de $n - k$ juegos que se terminan por un juego perdido, $(\underbrace{i, \dots, 0}_{n-k \text{ juegos}})$. Esto se

puede ver con la ecuación de Chapman-Kolmogorov, para $n \geq 2$ y $0 \leq k \leq n - 1$

$$P_{i,k}^{(n)} = \sum_{z \geq 0} P_{i,z}^{(n-k)} P_{z,k}^{(k)} = P_{i,0}^{(n-k)} P_{0,k}^{(k)} = qp^k,$$

ya que la única manera de ir al estado k en exactamente k pasos es: partir de 0 y no perder ningún juego, lo cual ocurre con probabilidad p^k y además vimos arriba que de cualquier estado se va a cero en j pasos con probabilidad q .

Ahora veamos lo que pasa con dichas probabilidades cuando n tiende a infinito. Gracias al calculo anterior es fácil ver que para cualquier estado $i \geq 0$ y $k \geq 0$

$$\lim_{n \rightarrow \infty} P_{i,k}^{(n)} = qp^k.$$

Observemos lo siguiente: 1. Este límite no depende del estado i del cual parte la cadena, y 2. Usando el resultado del ejercicio anterior podemos asegurar que la cadena es recurrente positiva y que el vector de probabilidad invariante π esta dado por $\pi_k = qp^k, k \geq 0$. Dicho de otro modo, las entradas de las potencias de la matriz de transición convergen al vector invariante. ▲

Ejemplo 3.18

Consideremos una cadena de Markov con espacio de estados $\mathcal{E} = \{0, 1, 2, 3, 4, 5\}$ y la siguiente matriz de transición

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

Veamos que esta cadena tiene infinitas distribuciones estacionarias. El diagrama de transiciones de la cadena es

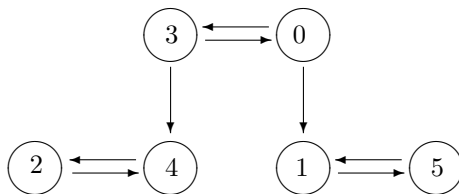


Figura 3.3

Vemos que hay tres clases de equivalencia, $C = \{0, 3\}$, que es transitoria, y $A = \{1, 5\}$ y $B = \{2, 4\}$ que son ambas recurrentes. Por lo tanto, existen distribuciones estacionarias que están concentradas tanto en A como B . Para hallarlas consideramos las matrices de transición concentradas en cada clase de equivalencia cerrada e irreducible:

$$P_A = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix} \quad P_B = \begin{pmatrix} 1/2 & 1/2 \\ 3/4 & 1/4 \end{pmatrix}$$

Es sencillo ver que la distribución estacionaria concentrada en A es uniforme: $\pi_A = (1/2, 1/2)$ mientras que la distribución concentrada en B es $\pi_B = (3/5, 2/5)$. Si $0 \leq \alpha \leq 1$, cualquier combinación de estas distribuciones de la forma

$$\alpha\pi_A + (1 - \alpha)\pi_B$$

también es una distribución estacionaria. ▲

3.10. Inferencia en Cadenas de Markov

Consideremos una cadena de Markov $X_n, n \geq 0$ con espacio de estados \mathcal{E} de tamaño m . Como es usual denotamos los elementos de \mathcal{E} por $1, 2, \dots, m$. La cadena tiene matriz de transición estacionaria P de entradas $P_{ij}, 1 \leq i, j \leq m$ y nuestro interés inicial en esta sección es describir procedimientos de estimación para las entradas de esta matriz a partir de la observación de una muestra de la cadena. Para simplificar vamos a suponer que la cadena es irreducible, de modo que todos sus estados son recurrentes positivos.

Supongamos que hemos observado los primeros n estados de la cadena x_1, x_2, \dots, x_n . La probabilidad de obtener esta realización de la cadena es

$$\begin{aligned} P((X_1, X_2, \dots, X_n) = (x_1, x_2, \dots, x_n)) &= P(X_1 = x_1) \prod_{j=2}^n P(X_j = x_j | X_{j-1} = x_{j-1}, \dots, X_1 = x_1) \\ &= P(X_1 = x_1) \prod_{k=2}^n P(X_k = x_k | X_{k-1} = x_{k-1}) \\ &= P(X_1 = x_1) \prod_{k=2}^n P_{x_{k-1}, x_k} \end{aligned}$$

donde hemos usado la propiedad de Markov. Esta función es la verosimilitud para la matriz de transición P asociada a esta muestra de la cadena:

$$L(P) = P(X_1 = x_1) \prod_{j=2}^n P_{x_{j-1}, x_j} \quad (3.41)$$

Vamos a reescribir esta función usando el conteo de las transiciones entre estados sucesivos. Llamemos n_{ij} al número de veces que hemos observado una transición del estado i al estado j . Podemos reescribir la verosimilitud como

$$L(P) = P(X_1 = x_1) \prod_{i=1}^m \prod_{j=1}^m P_{ij}^{n_{ij}} \quad (3.42)$$

y queremos maximizar esta expresión como función de P_{ij} . Tenemos que tomar en cuenta que la matriz P es una matriz estocástica y por lo tanto todas sus filas deben sumar 1:

$$\sum_{j=1}^m P_{ij} = 1 \quad \text{para } 1 \leq i \leq m. \quad (3.43)$$

Tomando logaritmos obtenemos la logverosimilitud

$$\ell(P) = \log L(P) = \log P(X_1 = x_1) + \sum_{i,j} n_{ij} \log P_{ij} \quad (3.44)$$

Las ecuaciones (3.43) representan m restricciones en el proceso de optimización. Sean $\lambda_i, 1 \leq i \leq m$ los multiplicadores de Lagrange asociados a estas restricciones, entonces la función objetivo es

$$\ell(P) - \sum_{i=1}^m \lambda_i \left(\sum_j P_{ij} - 1 \right) \quad (3.45)$$

Derivando respecto de P_{ij} e igualando a cero obtenemos

$$\frac{n_{ij}}{\widehat{P}_{ij}} - \lambda_i = 0$$

de donde

$$\hat{P}_{ij} = \frac{n_{ij}}{\lambda_i}.$$

Finalmente, usando las restricciones (3.43) obtenemos $\lambda_i = \sum_j n_{ij}$ y por lo tanto

$$\hat{P}_{ij} = \frac{n_{ij}}{\sum_j n_{ij}} = \frac{n_{ij}}{n_{i+}}, \quad \text{donde } n_{i+} = \sum_j n_{ij}. \quad (3.46)$$

Ejemplo 3.19

Consideremos las siguientes observaciones de una cadena de Markov con estados 0 y 1:

0 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 0 0 0 1 1 1 0 0

La tabla de contingencia para las transiciones es

3	4	7
4	19	23
7	23	30

y usando el estimador dado por (3.46) obtenemos los valores estimados para la matriz de transición

$$\hat{P} = \begin{pmatrix} 3/7 & 4/7 \\ 4/23 & 19/23 \end{pmatrix}.$$

▲

3.10.1. Comportamiento asintótico

Veamos ahora el comportamiento asintótico de este estimador. Podemos escribir n_{ij} como la suma de funciones indicadoras: Si $\mathbf{1}_j$ es la función indicadora del estado j tenemos

$$n_{ij} = \sum_{k=1}^{n-1} \mathbf{1}_i(X_k) \mathbf{1}_j(X_{k+1}) \quad (3.47)$$

y por lo tanto

$$\frac{n_{ij}}{n-1} = \frac{1}{n-1} \sum_{k=1}^{n-1} \mathbf{1}_i(X_k) \mathbf{1}_j(X_{k+1})$$

representa el promedio muestral del número de transiciones de i a j observados en la muestra. Usando el teorema ergódico que demostramos en la sección 3.8, obtenemos que

$$\frac{n_{ij}}{n-1} \rightarrow \mathbb{E}_\pi(\mathbf{1}_i(X_t) \mathbf{1}_j(X_{t+1})) = P(X_k = i, X_{k+1} = j) = \pi(i) P_{ij} \quad (3.48)$$

donde π es la distribución estacionaria de la cadena. Veamos ahora el comportamiento asintótico de n_{i+} :

$$n_{i+} = \sum_{j=1}^m n_{ij} = \sum_{k=1}^{n-1} \mathbf{1}_i(X_k) \quad (3.49)$$

y de nuevo dividiendo por $n - 1$ obtenemos

$$\frac{n_{i+}}{n-1} = \frac{1}{n-1} \sum_{k=1}^{n-1} \mathbf{1}_i(X_k) \rightarrow \mathbb{E}_\pi(\mathbf{1}_i(X_k)) = \pi(i). \quad (3.50)$$

A partir de (3.48) y (3.50) obtenemos que

$$\hat{P}_{ij} = \frac{n_{ij}}{n_{i+}} \rightarrow \frac{\pi(i)P_{ij}}{\pi(i)} = P_{ij}$$

de modo que el estimador es consistente. Observamos que esto ocurre para cualquier distribución inicial de la cadena, por las propiedades que hemos estudiado en este capítulo.

Es posible demostrar (ver Billingsley P.: *Statistical Inference for Markov Processes*. The University of Chicago Press, Chicago (1961)) que la distribución conjunta de $\sqrt{n}(\hat{P}_{ij} - P_{ij})$ para $i, j \in \mathcal{E}$ es asintóticamente normal con media 0 y varianzas y covarianzas dadas por

$$\text{Var}(\sqrt{n}(\hat{P}_{ij} - P_{ij})) \rightarrow \pi(i)P_{ij}(1 - P_{ij}) \quad (3.51)$$

$$\text{Cov}(\sqrt{n}(\hat{P}_{ij} - P_{ij}), \sqrt{n}(\hat{P}_{ik} - P_{ik})) \rightarrow -\pi(i)P_{ij}P_{ik}, \quad j \neq k \quad (3.52)$$

$$\text{Cov}(\sqrt{n}(\hat{P}_{ij} - P_{ij}), \sqrt{n}(\hat{P}_{hk} - P_{hk})) \rightarrow 0, \quad i \neq h \quad (3.53)$$

En nuestra consideración no hemos tomado en cuenta el primer término en (3.41). Si el tamaño de la muestra n es grande, este término va a contribuir poco a la logverosimilitud y es posible no tomarlo en cuenta. Alternativamente, si en lugar de tener una muestra de tamaño n tenemos m muestras independientes y finitas de la misma cadena de Markov con $m \rightarrow \infty$ entonces podemos estimarlo.

3.10.2. Pruebas de independencia

Queremos ahora considerar la posibilidad de que los datos que observamos sean en realidad una muestra i.i.d. de una cierta distribución Q con valores en \mathcal{E} , en lugar de provenir de una cadena de Markov con matriz de transición P . Si los datos son i.i.d., la matriz de transición de este proceso debería tener todas sus filas iguales y por lo tanto $Q_{ij} = q_j$ para todo i , $1 \leq j \leq m$.

Para hacer un contraste de hipótesis debemos hallar los estimadores de máxima verosimilitud en ambos casos. Ya sabemos que para el modelo de una cadena de Markov homogénea en el tiempo $\hat{P}_{ij} = n_{ij}/n_{i+}$. Bajo la hipótesis de independencia tenemos una distribución multinomial con $n_{+j} = \sum_i n_{ij}$ observaciones del valor j . La verosimilitud es

$$\ell(Q) = \sum_{j=1}^{m-1} n_{+j} q_j + n_{+m} \left(1 - \sum_{j=1}^{m-1} q_j\right) \quad (3.54)$$

que se maximiza en $\hat{q}_j = n_{+j}/n$. Por lo tanto, el estadístico para la prueba de cociente de verosimilitudes para la hipótesis nula de independencia es

$$2(\ell(\hat{P}) - \ell(\hat{Q})) = 2 \sum_{ij} n_{ij} \log \frac{n_{ij}/n_{i+}}{n_{+j}/n} \quad (3.55)$$

que tiene asintóticamente una distribución χ^2 con $m(m-1) - (m-1) = (m-1)^2$ grados de libertad.

3.10.3. Orden de la cadena

Definición 3.9 Un proceso estocástico $X_n, n \geq 0$ es una cadena de Markov de orden k si satisface la siguiente condición: Para $n > k$,

$$\begin{aligned} P(X_n = x_n | X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \dots, X_0 = x_0) \\ = P(X_n = x_n | X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \dots, X_{n-k} = x_{n-k}) \end{aligned} \quad (3.56)$$

En otras palabras, el futuro depende de los k estados previos.

Las cadenas que hemos estudiado hasta ahora han sido de primer orden. Si X_n es una cadena de Markov de orden k , es posible demostrar que se puede construir una cadena Y_n a partir de X_n que satisface la propiedad de Markov que definimos en la sección 2.2. Para ello hay que considerar como estados de la nueva cadena, a los vectores de k estados consecutivos ordenados del proceso X , es decir, $Y_n = (X_n, X_{n-1}, \dots, X_{n-k+1})$.

Podemos determinar si una cadena de orden 1 es un modelo razonable haciendo una prueba de hipótesis contra la alternativa de que es de orden 2. Si suponemos que la cadena es de orden 2, los estimadores de máxima verosimilitud de $P_{ij,k} = P(X_{n+2} = k | X_{n+1} = j, X_n = i)$ están dados por

$$\hat{P}_{ij,k} = \frac{n_{ijk}}{n_{ij+}}$$

donde

$$n_{ijk} = \sum_{n=1}^{n-2} \mathbf{1}_i(X_n) \mathbf{1}_j(X_{n+1}) \mathbf{1}_k(X_{n+2}), \quad n_{ij+} = \sum_k n_{ijk}$$

El estadístico de cociente de verosimilitudes está dado por

$$2 \left(\sum_{ijk} n_{ijk} \log \hat{P}_{ijk} - \sum_{ij} n_{ij} \log \hat{P}_{ij} \right) \sim \chi_{m(m-1)}^2$$

ya que la diferencia en el número de parámetros entre los dos modelos es

$$m^2(m-1) - m(m-1) = m(m-1)^2.$$

Capítulo 4

Procesos de Poisson

4.1. Distribución Exponencial

Definición 4.1 Una variable aleatoria T tiene *distribución exponencial* con parámetro $\lambda > 0$, $T \sim \text{Exp}(\lambda)$, si su función de distribución está dada por

$$F_T(t) = P(T \leq t) = 1 - e^{-\lambda t}, \quad \text{para } t \geq 0.$$

Equivalentemente, T tiene densidad $f_T(t)$ dada por

$$f_T(t) = \begin{cases} \lambda e^{-\lambda t} & \text{para } t \geq 0, \\ 0 & \text{para } t < 0. \end{cases}$$

Esta distribución tiene el siguiente valor esperado:

$$\begin{aligned} E[T] &= \int_{-\infty}^{\infty} t f_T(t) dt = \int_0^{\infty} t \lambda e^{-\lambda t} dt \\ &= -t e^{-\lambda t} \Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda t} dt = \frac{1}{\lambda}. \end{aligned}$$

De manera similar podemos calcular $E[T^2]$ integrando por partes,

$$\begin{aligned} E[T^2] &= \int_{-\infty}^{\infty} t^2 f_T(t) dt = \int_0^{\infty} t^2 \lambda e^{-\lambda t} dt \\ &= -t^2 e^{-\lambda t} \Big|_0^{\infty} + \int_0^{\infty} 2t e^{-\lambda t} dt = \frac{2}{\lambda^2} \end{aligned}$$

y por lo tanto, la varianza de T es

$$\text{Var}(T) = E[T^2] - (E[T])^2 = \frac{1}{\lambda^2}.$$

4.1.1. Falta de Memoria

Una de las propiedades fundamentales de la distribución exponencial es la siguiente: Para $s, t \in (0, \infty)$

$$P(T > t + s | T > t) = P(T > s).$$

Para demostrar esta propiedad usamos la definición de probabilidad condicional

$$P(T > t + s | T > t) = \frac{P(T > t + s)}{P(T > t)} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} = e^{-\lambda s} = P(T > s).$$

4.1.2. Mínimo de Variables Exponenciales

Sean $S \sim \text{Exp}(\lambda)$ y $T \sim \text{Exp}(\mu)$ variables independientes. Tenemos en primer lugar

$$\begin{aligned} P(\min(S, T) > t) &= P(S > t, T > t) \\ &= P(S > t)P(T > t) = e^{-(\lambda+\mu)t}, \end{aligned}$$

es decir, $\min(S, T)$ tiene distribución exponencial de parámetro $\lambda + \mu$. El mismo cálculo muestra que para una colección de variables independientes T_1, \dots, T_n con $T_i \sim \text{Exp}(\lambda_i)$, $1 \leq i \leq n$,

$$\begin{aligned} P(\min(T_1, \dots, T_n) > t) &= P(T_1 > t, \dots, T_n > t) \\ &= \prod_{i=1}^n P(T_i > t) = \prod_{i=1}^n e^{-\lambda_i t} = e^{-(\lambda_1 + \dots + \lambda_n)t} \end{aligned} \quad (4.1)$$

En consecuencia, el mínimo de varias variables independientes con distribuciones exponenciales tiene distribución exponencial con parámetro igual a la suma de los parámetros.

Veamos ahora con qué probabilidad una variable exponencial es menor que otra. Sean $S \sim \text{Exp}(\lambda)$ y $T \sim \text{Exp}(\mu)$ independientes, tenemos

$$\begin{aligned} P(T > S) &= \int_0^\infty P(T > s | S = s) f_S(s) ds \\ &= \int_0^\infty \lambda e^{-\lambda s} e^{-\mu s} ds = \frac{\lambda}{\lambda + \mu} \int_0^\infty (\lambda + \mu) e^{-(\lambda + \mu)s} ds \\ &= \frac{\lambda}{\lambda + \mu}. \end{aligned}$$

Para varias variables, el resultado es el siguiente

$$\begin{aligned} P(T_i = \min(T_1, \dots, T_n)) &= P(T_i < T_1, \dots, T_i < T_{i-1}, T_i < T_{i+1}, \dots, T_i < T_n) \\ &= \frac{\lambda_i}{\lambda_1 + \dots + \lambda_n}. \end{aligned}$$

Para demostrar esta propiedad llamemos $S = T_i$ y sea U el mínimo de T_j , $j \neq i$. Por (4.1) sabemos que U es exponencial con parámetro $\mu = (\lambda_1 + \dots + \lambda_n) - \lambda_i$. Usando el resultado para dos variables

$$P(T_i = \min(T_1, \dots, T_n)) = P(S < U) = \frac{\lambda_i}{\lambda_i + \mu} = \frac{\lambda_i}{\lambda_1 + \dots + \lambda_n}.$$

Sea I el índice (aleatorio) de la menor de las variables exponenciales, hemos demostrado que

$$P(I = i) = \frac{\lambda_i}{\lambda_1 + \dots + \lambda_n}.$$

Lema 4.1 I y $V = \min(T_1, \dots, T_n)$ son independientes.

Demostración. Calculamos la siguiente probabilidad conjunta

$$\begin{aligned} P(I = i, V > t) &= P(T_i > t, T_j > T_i, \forall j \neq i) = \int_t^\infty P(T_j > s, \forall j \neq i) f_{T_i}(s) ds \\ &= \int_t^\infty \lambda_i e^{-\lambda_i s} \prod_{j \neq i} e^{-\lambda_j s} ds = \lambda_i \int_t^\infty e^{-s(\sum_j \lambda_j)} ds \\ &= \frac{\lambda_i}{\lambda_1 + \dots + \lambda_n} e^{-t(\sum_j \lambda_j)} = P(I = i)P(V > t). \end{aligned}$$

Veamos a continuación cómo se distribuye una suma de exponenciales. ■

Teorema 4.1 Sean T_1, T_2, \dots v.a.i.i.d. con distribución exponencial de parámetro λ . La suma $\tau_n = T_1 + \dots + T_n$ tiene distribución $\Gamma(n, \lambda)$, es decir, la densidad está dada por

$$f_{\tau_n}(t) = \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!} \quad \text{para } t \geq 0$$

y 0 en otro caso.

Demostración. Haremos la prueba por inducción. Para $n = 1$, $\tau_1 = T_1$ tiene distribución exponencial de parámetro λ , que concuerda con la densidad de la fórmula anterior.

Supongamos ahora que la fórmula es cierta para n . Tenemos $\tau_{n+1} = \tau_n + T_{n+1}$ y por independencia

$$\begin{aligned} P(\tau_{n+1} \leq t) &= \int_0^t P(\tau_n + T_{n+1} \leq t | \tau_n = s) f_{\tau_n}(s) ds \\ &= \int_0^t P(T_{n+1} \leq t - s) f_{\tau_n}(s) ds \end{aligned}$$

Usamos ahora la distribución exponencial para el primer factor y la fórmula inductiva para el segundo obtenemos

$$\begin{aligned} \int_0^t (1 - e^{-\lambda(t-s)}) \lambda e^{-\lambda s} \frac{(\lambda s)^{n-1}}{(n-1)!} ds &= \frac{\lambda^n}{(n-1)!} \int_0^t e^{-\lambda s} s^{n-1} ds - \frac{\lambda^n}{(n-1)!} \int_0^t e^{-\lambda t} s^{n-1} ds \\ &= \frac{\lambda^n}{(n-1)!} \left[\frac{1}{n} s^n e^{-\lambda s} \Big|_0^t + \int_0^t \lambda \frac{s^n}{n} e^{-\lambda s} ds - \frac{t^n}{n} e^{-\lambda t} \right] \\ &= \int_0^t \lambda e^{-\lambda s} \frac{(\lambda s)^n}{n!} ds. \end{aligned}$$

Como consecuencia del teorema anterior, teniendo en cuenta que la distribución $\Gamma(n, \lambda)$ se obtiene como suma de v.a.i.i.d. con distribución exponencial de parámetro λ , vemos que

$$E[\tau_n] = \frac{n}{\lambda}, \quad \text{Var}(\tau_n) = \frac{n}{\lambda^2}.$$

También es posible demostrar que la función de distribución de τ_n se puede escribir de la siguiente manera:

$$P(\tau_n \leq x) = 1 - \sum_{i=0}^{n-1} \frac{(\lambda x)^i}{i!} e^{-\lambda x} = \sum_{i=n}^{\infty} \frac{(\lambda x)^i}{i!} e^{-\lambda x}$$

Observación 4.1 Tenemos los siguientes casos especiales de la distribución Gamma: $\Gamma(1, \lambda)$ es la distribución exponencial de parámetro λ mientras que $\Gamma(k, 2)$ es la distribución Ji-cuadrado con $2k$ grados de libertad, χ_{2k}^2 . Además, si $X \sim \Gamma(n, \lambda)$ entonces $cX \sim \Gamma(n, \lambda/c)$.

4.2. La Distribución de Poisson

Definición 4.2 Una variable aleatoria X tiene *distribución de Poisson* de parámetro $\lambda > 0$ si toma valores en el conjunto $\{0, 1, 2, \dots\}$, con probabilidad dada por

$$P(X = k) = p_k = e^{-\lambda} \frac{\lambda^k}{k!}.$$

Calculemos la función generadora de probabilidad de una variable de este tipo:

$$\phi_X(s) = E[s^X] = \sum_{k=0}^{\infty} s^k e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(s\lambda)^k}{k!} = e^{\lambda(s-1)}.$$

A partir de esta expresión podemos obtener los momentos de la distribución:

$$E[X] = \left. \frac{d\phi}{ds} \right|_{s=1} = \left. \lambda e^{\lambda(s-1)} \right|_{s=1} = \lambda,$$

$$E[X(X-1)] = \left. \frac{d^2\phi}{ds^2} \right|_{s=1} = \left. \lambda^2 e^{\lambda(s-1)} \right|_{s=1} = \lambda^2,$$

$$E[X^2] = E[X(X-1)] + E[X] = \lambda^2 + \lambda,$$

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

Si $X \sim \text{Pois}(\lambda)$ e $Y \sim \text{Pois}(\mu)$ son independientes entonces la suma tiene f.g.p.

$$\phi_{X+Y}(s) = \phi_X(s)\phi_Y(s) = e^{\lambda(s-1)} e^{\mu(s-1)} = e^{(\lambda+\mu)(s-1)}$$

y vemos que $X + Y$ tiene distribución de Poisson con parámetro $\lambda + \mu$.

Lema 4.2 Sea $N \sim \text{Pois}(\lambda)$ y condicional a N , M tiene distribución binomial con parámetros N y p . Entonces la distribución (incondicional) de M es Poisson con parámetro λp .

Demostración. Podemos considerar M como la suma de una cantidad aleatoria N de variables de Bernoulli con probabilidad de éxito p :

$$M = X_1 + \dots + X_N$$

donde X_i , $i \geq 1$ tiene distribución de Bernoulli con probabilidad de éxito p . La f.g.p. de una variable de Bernoulli es

$$\phi_X(s) = E[s^X] = q + sp$$

y ya hemos visto que la f.g.p. de N es $\phi_N(t) = e^{\lambda(t-1)}$. Por lo tanto, la f.g.p. de M es la composición de estas dos:

$$\phi_M(s) = \phi_N(\phi_X(s)) = e^{\lambda(q+sp-1)} = e^{\lambda(sp-p)} = e^{\lambda p(s-1)}$$

que es la f.g.p. de una v.a. de Poisson con parámetro λp . ■

4.3. El Proceso de Poisson

Definición 4.3 Sean T_1, T_2, \dots v.a.i.i.d. con distribución exponencial de parámetro λ , $\tau_0 = 0$ y $\tau_n = T_1 + \dots + T_n$ para $n \geq 1$. Definimos el *proceso de Poisson de parámetro o intensidad λ* por

$$N(s) = \text{máx}\{n : \tau_n \leq s\}, \quad s \geq 0.$$

Las variables T_n representan los intervalos de tiempo entre eventos sucesivos (llegadas de clientes a una cola, de llamadas a una central telefónica, de pacientes a la emergencia de un hospital, etc.) y en consecuencia $\tau_n = T_1 + \dots + T_n$ es el instante en el que ocurre el n -ésimo evento y $N(s)$ es el número de eventos que han ocurrido hasta el instante s (ver figura 4.1). Llamaremos *tiempos de llegada* del proceso a las variables τ_n , $n \geq 1$.

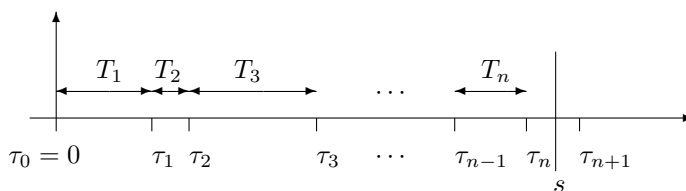


Figura 4.1

Para ver por qué $N(s)$, $s \geq 0$, recibe este nombre, calculemos su distribución: $N(s) = n$ si y sólo si $\tau_n \leq s < \tau_{n+1}$, es decir, el n -ésimo evento ocurre antes del instante s o en s , pero el $(n+1)$ -ésimo ocurre después de s . Usando la ley de la probabilidad total, condicionando respecto al instante en el cual ocurre τ_n , obtenemos

$$\begin{aligned} P(N(s) = n) &= P(\tau_{n+1} > s > \tau_n) = \int_0^s P(\tau_{n+1} > s | \tau_n = t) f_{\tau_n}(t) dt \\ &= \int_0^s P(T_{n+1} > s - t) f_{\tau_n}(t) dt. \end{aligned}$$

Usando ahora el resultado del teorema 4.1 obtenemos

$$\begin{aligned} &= \int_0^s \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!} e^{-\lambda(s-t)} dt \\ &= \frac{\lambda^n}{(n-1)!} e^{-\lambda s} \int_0^s t^{n-1} dt = \frac{(\lambda s)^n}{n!} e^{-\lambda s}. \end{aligned}$$

Por lo tanto hemos demostrado el siguiente resultado

Lema 4.3 $N(s)$ tiene distribución de Poisson de parámetro λs .

Veamos algunas propiedades del proceso que acabamos de definir.

Lema 4.4 $N(t+s) - N(s)$, $t \geq 0$ es un proceso de Poisson de parámetro λ y es independiente de $N(r)$, $0 \leq r \leq s$.

Demostración. Supongamos que $N(s) = n$ y que el n -ésimo evento ocurrió en el instante τ_n . Sabemos que el intervalo de tiempo para el siguiente evento debe satisfacer $T_{n+1} > s - \tau_n$, pero por la propiedad de falta de memoria de la distribución exponencial

$$P(T_{n+1} > s - \tau_n + t | T_{n+1} > s - \tau_n) = P(T_{n+1} > t) = e^{-\lambda t}.$$

Esto muestra que la distribución del tiempo de espera hasta el primer evento después de s es exponencial de parámetro λ y es independiente de T_i , $1 \leq i \leq n$. Por otro lado T_{n+1}, T_{n+2}, \dots son independientes de T_i , $1 \leq i \leq n$ y por lo tanto también de τ_i , $1 \leq i \leq n$. Esto muestra que los intervalos entre eventos que ocurren después de s son v.a.i.i.d. con distribución exponencial de parámetro λ , y por lo tanto $N(t+s) - N(s)$ es un proceso de Poisson. ■

Como consecuencia de este resultado tenemos

Lema 4.5 $N(t)$ tiene incrementos independientes: Si $t_0 < t_1 < \dots < t_n$, entonces

$$N(t_1) - N(t_0), N(t_2) - N(t_1), \dots, N(t_n) - N(t_{n-1})$$

son independientes.

Demostración. El lema 4.5 implica que $N(t_n) - N(t_{n+1})$ es independiente de $N(r)$, $r \leq t_{n-1}$ y en consecuencia también de $N(t_{n-1}) - N(t_{n-2}), \dots, N(t_1) - N(t_0)$. El resultado sigue por inducción. ■

Combinando los dos lemas anteriores tenemos la mitad del siguiente resultado, que es una caracterización fundamental del proceso de Poisson.

Teorema 4.2 Si $\{N(s), s \geq 0\}$ es un proceso de Poisson de parámetro $\lambda > 0$, entonces

1. $N(0) = 0$.
2. $N(t+s) - N(s) \sim \text{Pois}(\lambda t)$.
3. $N(t)$ tiene incrementos independientes.

Recíprocamente, si 1, 2 y 3 valen, entonces $\{N(s), s \geq 0\}$ es un proceso de Poisson.

Demostración. Los lemas 4.3 y 4.4 demuestran la primera afirmación. Para ver el recíproco, sea τ_n el instante en el cual ocurre el n -ésimo evento. El primer evento ocurre después de t si y sólo si no ocurre ningún evento en $[0, t]$. Usando la fórmula para la distribución de Poisson

$$P(\tau_1 > t) = P(N(t) = 0) = e^{-\lambda t}$$

lo cual muestra que $\tau_1 = T_1 \sim \text{Exp}(\lambda)$. Para $T_2 = \tau_2 - \tau_1$ observamos que

$$\begin{aligned} P(T_2 > t | T_1 = s) &= P(\text{no ocurre ningún evento en } (s, s+t] | T_1 = s) \\ &= P(N(t+s) - N(s) = 0 | N(r) = 0 \text{ para } r < s, N(s) = 1) \\ &= P(N(t+s) - N(s) = 0) = e^{-\lambda t} \end{aligned}$$

por la propiedad de incrementos independientes, de modo que $T_2 \sim \text{Exp}(\lambda)$ y es independiente de T_1 . Repitiendo este argumento vemos que T_1, T_2, \dots son i.i.d. con distribución exponencial de parámetro λ . ■

Ejemplo 4.1

Un cable submarino tiene defectos de acuerdo a un proceso de Poisson de parámetro $\lambda = 0.1$ por km. (a) ¿Cuál es la probabilidad de que no haya defectos en los primeros dos kilómetros de cable? (b) Si no hay defectos en los primeros dos kilómetros, ¿cuál es la probabilidad de que tampoco los haya en el tercer kilómetro?

(a) $N(2)$ tiene distribución de Poisson de parámetro $(0.1)(2) = 0.2$. Por lo tanto

$$P(N(2) = 0) = e^{-0.2} = 0.8187.$$

(b) $N(3) - N(2)$ y $N(2) - N(0) = N(2)$ son independientes, de modo que

$$P(N(3) - N(2) = 0 | N(2) = 0) = P(N(3) - N(2) = 0) = e^{-0.1} = 0.9048$$

▲

Ejemplo 4.2

Los clientes llegan a una tienda de acuerdo con un proceso de Poisson de tasa $\lambda = 4$ por hora. Si la tienda abre a las 9 a.m. ¿Cuál es la probabilidad de que exactamente un cliente haya entrado antes de las 9:30 a.m. y que un total de cinco hayan entrado antes de las 11:30 a.m.?

Medimos el tiempo t en horas a partir de las 9 a.m. Queremos hallar $P(N(1/2) = 1, N(5/2) = 5)$, y para esto usaremos la independencia de los incrementos:

$$\begin{aligned} P(N(1/2) = 1, N(5/2) = 5) &= P(N(1/2) = 1, N(5/2) - N(1/2) = 4) \\ &= \left(\frac{e^{-4(1/2)} 4(1/2)}{1!} \right) \left(\frac{e^{-4(2)} [4(2)]^4}{4!} \right) \\ &= (2e^{-2}) \left(\frac{512}{3} e^{-8} \right) = 0.0155. \end{aligned}$$

▲

La importancia de la distribución de Poisson, y del proceso de Poisson en particular, se debe, al menos en parte, al siguiente resultado, que se conoce como la ley de los eventos raros.

Consideremos una cantidad grande n de ensayos de Bernoulli independientes con probabilidad de éxito p constante. Sea S_n el número de éxitos en los n ensayos. Sabemos que S_n tiene distribución Binomial de parámetros n y p :

$$P(S_n = k) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}.$$

Supongamos ahora que el número de ensayos n tiende a infinito y la probabilidad de éxito p tiende a 0, de modo que $np = \lambda$. Veamos que ocurre con la distribución de S_n en este caso. Reemplacemos p por λ/n en la ecuación anterior

$$\begin{aligned} P(S_n = k) &= \frac{n(n-1) \cdots (n-k+1)}{k!} \left(\frac{\lambda}{n} \right)^k \left(1 - \frac{\lambda}{n} \right)^{n-k} \\ &= \frac{\lambda^k}{k!} \frac{n(n-1) \cdots (n-k+1)}{n^k} \left(1 - \frac{\lambda}{n} \right)^n \left(1 - \frac{\lambda}{n} \right)^{-k}. \end{aligned} \quad (4.2)$$

Veamos ahora el comportamiento de estos cuatro factores cuando $n \rightarrow \infty$. El primer factor no depende de n . En el segundo hay k factores en el numerador y k en el denominador y podemos escribirlo como

$$\frac{n}{n} \frac{n-1}{n} \cdots \frac{n-k+1}{n}.$$

En virtud de que k está fijo es fácil ver que todos estos factores tienden a 1, y en consecuencia su producto también. El tercer factor converge a $e^{-\lambda}$. Finalmente, el último converge a 1 ya que $\lambda/n \rightarrow 0$ y la potencia k de este factor está fija. Reuniendo estos resultados vemos que la probabilidad (4.2) converge a

$$\frac{\lambda^k}{k!} e^{-\lambda}$$

que es la distribución de Poisson de parámetro λ . El mismo resultado es cierto si en lugar de tener $np = \lambda$ tenemos que $p \rightarrow 0$ cuando $n \rightarrow \infty$ de modo que $np \rightarrow \lambda$.

En realidad la ley de eventos raros se cumple con mayor generalidad aún. Es posible suponer que los ensayos de Bernoulli no tienen una probabilidad de éxito común, como lo muestra el siguiente teorema. Primero enunciamos y demostramos un resultado auxiliar.

Lema 4.6 Sean S y T dos variables aleatorias y A un subconjunto medible de \mathbb{R} . Entonces

$$|P(S \in A) - P(T \in A)| \leq P(S \neq T).$$

Demostración.

$$\begin{aligned} P(S \in A) &= P(S \in A, S = T) + P(S \in A, S \neq T) = P(T \in A, S = T) + P(S \in A, S \neq T) \\ &= P(T \in A, S = T) + P(T \in A, S \neq T) - P(T \in A, S \neq T) + P(S \in A, S \neq T) \\ &= P(T \in A) - P(T \in A, S \neq T) + P(S \in A, S \neq T) \end{aligned}$$

Por lo tanto

$$P(S \in A) - P(T \in A) = P(S \in A, S \neq T) - P(T \in A, S \neq T) \leq P(S \in A, S \neq T) \leq P(S \neq T),$$

y de manera similar

$$P(T \in A) - P(S \in A) \leq P(S \neq T),$$

de modo que

$$|P(T \in A) - P(S \in A)| \leq P(S \neq T),$$

■

Teorema 4.3 (Le Cam) Sean X_m , $1 \leq m \leq n$, variables aleatorias independientes con

$$P(X_m = 1) = p_m, \quad P(X_m = 0) = 1 - p_m.$$

Sean

$$S_n = X_1 + \cdots + X_n, \quad \lambda_n = E[S_n] = p_1 + \cdots + p_n.$$

Entonces, para cualquier conjunto A ,

$$\left| P(S_n \in A) - \sum_{k \in A} e^{-\lambda_n} \frac{\lambda_n^k}{k!} \right| \leq \sum_{m=1}^n p_m^2$$

Demostración. Las variables X_m son independientes y tienen distribución de Bernoulli con parámetro p_m . Definimos variables independientes $Y_m \sim \mathcal{Pois}(p_m)$, y como la suma de variables Poisson independientes es Poisson, tenemos que $Z_n = Y_1 + \cdots + Y_n$ tiene distribución $\mathcal{Pois}(\lambda_n)$ y

$$P(Z_n \in A) = \sum_{k \in A} e^{-\lambda_n} \frac{\lambda_n^k}{k!}.$$

Por lo tanto queremos comparar $P(S_n \in A)$ y $P(Z_n \in A)$ para cualquier conjunto A de enteros positivos. Por el lema 4.6

$$|P(S_n \in A) - P(Z_n \in A)| \leq P(S_n \neq Z_n) = P\left(\sum_{m=1}^n X_m \neq \sum_{m=1}^n Y_m\right),$$

pero si S_n y Z_n difieren, al menos uno de los pares X_m y Y_m deben diferir también. En consecuencia

$$|P(S_n \in A) - P(Z_n \in A)| \leq \sum_{m=1}^n P(X_m \neq Y_m)$$

y para completar la demostración hay que ver que $P(X_m \neq Y_m) \leq p^2$. Para simplificar la notación sean $X \sim \text{Ber}(p)$ y $Y \sim \text{Pois}(p)$ y veamos que $P(X \neq Y) \leq p^2$, o equivalentemente, que

$$1 - p^2 \leq P(X = Y) = P(X = Y = 0) + P(X = Y = 1).$$

Este resultado no depende de la distribución conjunta entre X y Y pues no hemos supuesto ninguna propiedad de independencia entre ellas. Lo importante es que las distribuciones marginales de X y Y sigan siendo las mismas. Escogemos la distribución conjunta de (X, Y) de la siguiente manera: Sea U una variable con distribución uniforme en $(0, 1]$ y sean

$$X = \begin{cases} 0 & \text{si } 0 < U \leq 1 - p \\ 1 & \text{si } 1 - p < U \leq 1. \end{cases}$$

$Y = 0$ si $0 < U < e^{-p}$ y para $k = 1, 2, \dots$

$$Y = k \quad \text{si} \quad \sum_{i=0}^{k-1} e^{-p} \frac{p^i}{i!} < U \leq \sum_{i=0}^k e^{-p} \frac{p^i}{i!}.$$

Es sencillo verificar que X y Y tienen las distribuciones marginales adecuadas (esto no es más que el método de la transformada inversa de generación de variables aleatorias, aplicado las distribuciones de Bernoulli y de Poisson). Como $1 - p \leq e^{-p}$ tenemos que $X = Y = 0$ si y sólo si $U \leq 1 - p$, de modo que

$$P(X = Y = 0) = 1 - p.$$

De manera similar $X = Y = 1$ si y sólo si $e^{-p} < U \leq (1 + p)e^{-p}$ y por lo tanto

$$P(X = Y = 1) = pe^{-p}.$$

Sumando estas dos expresiones tenemos

$$P(X = Y) = 1 - p + pe^{-p} = 1 - p^2 + \frac{p^3}{2} + \dots \geq 1 - p^2.$$

■

Corolario 4.1 (Le Cam) Para cada n , sean $X_{n,m}$, $1 \leq m \leq n$, $n \geq 1$ variables aleatorias independientes con

$$P(X_{n,m} = 1) = p_{n,m}, \quad P(X_{n,m} = 0) = 1 - p_{n,m}.$$

Sean

$$S_n = X_{n,1} + \dots + X_{n,n}, \quad \lambda_n = E[S_n] = p_{n,1} + \dots + p_{n,n},$$

y $Z_n \sim \text{Pois}(\lambda_n)$. Entonces, para cualquier conjunto A ,

$$|P(S_n \in A) - P(Z_n \in A)| \leq \sum_{m=1}^n p_{n,m}^2$$

El teorema anterior nos da una cota para la diferencia entre la distribución de S_n y la distribución de Poisson de parámetro $\lambda_n = E[S_n]$, que podemos usar para obtener una versión general del teorema de aproximación de Poisson.

Corolario 4.2 *Supongamos que en la situación del corolario anterior $\lambda_n \rightarrow \lambda < \infty$ y $\max_k p_{n,k} \rightarrow 0$, cuando $n \rightarrow \infty$, entonces*

$$\max_A |P(S_n \in A) - P(Z_n \in A)| \rightarrow 0.$$

Demostración. Como $p_{n,m}^2 \leq p_{n,m}(\max_k p_{n,k})$, sumando sobre m obtenemos

$$\sum_{m=1}^n p_{n,m}^2 \leq \max_k p_{n,k} \sum_m p_{n,m}.$$

El primer factor de la derecha va a 0 por hipótesis. El segundo es λ_n que converge a $\lambda < \infty$ y en consecuencia el producto de los dos converge a 0. ■

4.4. Postulados para el Proceso de Poisson

Consideremos una sucesión de eventos que ocurren en $[0, \infty)$ como por ejemplo las emisiones de partículas por una sustancia radioactiva, la llegada de llamadas a una central telefónica, los accidentes que ocurren en cierto cruce carretero, la ubicación de fallas o defectos a lo largo de una fibra o las llegadas sucesivas de clientes a un establecimiento comercial. Sea $N((a, b])$ el número de eventos que ocurren en el intervalo $(a, b]$, es decir, si $\tau_1 < \tau_2 < \tau_3 \cdots$ representan los instantes (o ubicaciones) de los sucesivos eventos, entonces $N((a, b])$ es el número de estos instantes τ_i que satisfacen $a < \tau_i \leq b$.

Proponemos los siguientes postulados:

1. El número de eventos que ocurren en intervalos disjuntos son variables aleatorias independientes: Para cualquier entero $m \geq 2$ y cualesquiera instantes $t_0 = 0 < t_1 < t_2 < \cdots < t_m$, las variables aleatorias

$$N((t_0, t_1]), N((t_1, t_2]), \dots, N((t_{m-1}, t_m])$$

son independientes.

2. Para cualquier instante t y cualquier $h > 0$, la distribución de probabilidad de $N((t, t+h])$ depende sólo de la longitud del intervalo h y no del instante inicial t .
3. Hay una constante positiva λ para la cual la probabilidad de que ocurra al menos un evento en un intervalo de longitud h es

$$P(N((t, t+h]) \geq 1) = \lambda h + o(h), \quad \text{cuando } h \downarrow 0$$

(la notación $o(h)$ indica una función general indeterminada que representa el resto y satisface $o(h)/h \rightarrow 0$ cuando $h \downarrow 0$, es decir, que es de orden menor que h cuando $h \downarrow 0$). El parámetro λ se conoce como la intensidad del proceso.

4. La probabilidad de que haya dos o más eventos en un intervalo de longitud h es $o(h)$:

$$P(N((t, t+h]) \geq 2) = o(h), \quad \text{cuando } h \downarrow 0.$$

El número de sucesos que ocurren en intervalos disjuntos son independientes por 1, y 2 afirma que la distribución de $N((s, t])$ es la misma que la de $N((0, t-s])$. Por lo tanto, para describir la ley de probabilidad del sistema basta determinar la distribución de probabilidad de $N((0, t])$ para cualquier

valor de t . Llamemos $N((0, t]) = N(t)$. Mostraremos que los postulados anteriores implican que $N(t)$ tiene una distribución de Poisson:

$$P(N(t) = k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}, \quad \text{para } k = 0, 1, \dots \quad (4.3)$$

Para demostrar (4.3) dividimos el intervalo $(0, t]$ en n subintervalos de igual longitud $h = t/n$ y definimos las siguientes variables de Bernoulli: $\xi_{n,i} = 1$ si hay al menos un evento en el intervalo $((i-1)t/n, it/n]$ y $\xi_{n,i} = 0$ si no, para $1 \leq i \leq n$. $S_n = \xi_{n,1} + \dots + \xi_{n,n}$ representa el número de subintervalos que contienen al menos un evento y

$$p_{n,i} = P(\xi_{n,i} = 1) = \frac{\lambda t}{n} + o\left(\frac{t}{n}\right)$$

según el postulado 3. Sea

$$E(S_n) = \mu_n = \sum_{i=1}^n p_{n,i} = \lambda t + no\left(\frac{t}{n}\right).$$

Usando el teorema 4.3 vemos que

$$\begin{aligned} \left| P(S_n = k) - \frac{\mu_n^k e^{-\mu_n}}{k!} \right| &\leq n \left[\frac{\lambda t}{n} + o\left(\frac{t}{n}\right) \right]^2 \\ &= \frac{(\lambda t)^2}{n} + 2\lambda t o\left(\frac{t}{n}\right) + no^2\left(\frac{t}{n}\right), \end{aligned}$$

Como $o(h) = o(t/n)$ es un término de orden menor que $h = t/n$ para n grande, se tiene que

$$no(t/n) = t \frac{o(t/n)}{t/n} = t \frac{o(h)}{h}$$

tiende a 0 cuando n crece. Pasando al límite cuando $n \rightarrow \infty$ obtenemos que

$$\lim_{n \rightarrow \infty} P(S_n = k) = \frac{\mu^k e^{-\mu}}{k!}, \quad \text{con } \mu = \lambda t.$$

Para completar la demostración sólo falta ver que

$$\lim_{n \rightarrow \infty} P(S_n = k) = P(N((0, t]) = k)$$

pero S_n y $N((0, t])$ son diferentes si al menos uno de los subintervalos contiene dos o más eventos, y el postulado 4 impide esto porque

$$\begin{aligned} |P(N(t) = k) - P(S_n = k)| &\leq P(N(t) \neq S_n) \\ &\leq \sum_{i=1}^n P\left(N\left(\left(\frac{i-1}{n}t, \frac{it}{n}\right] \geq 2\right)\right) \\ &\leq no\left(\frac{t}{n}\right) \\ &\rightarrow 0 \quad \text{cuando } n \rightarrow \infty. \end{aligned}$$

En consecuencia, haciendo $n \rightarrow \infty$,

$$P(N((0, t]) = k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}, \quad \text{para } k \geq 0.$$

Esto completa la demostración de (4.3). ■

El proceso $N((a, b])$ se conoce como el Proceso Puntual de Poisson y sus valores, como hemos visto, se pueden calcular a partir de los del proceso $N(t)$:

$$N((s, t]) = N(t) - N(s)$$

Recíprocamente, $N(t) = N((0, t])$, de modo que ambos procesos son equivalentes, las diferencias son de enfoque, pero en algunos casos resulta útil considerar al proceso de una u otra manera.

A continuación presentamos sin demostración otra caracterización de los procesos de Poisson que resultará útil más adelante.

Teorema 4.4 $N(t), t \geq 0$ es un proceso de Poisson con intensidad λ si y sólo si

- a) Para casi todo ω , los saltos de $N(t, \omega)$ son unitarios.
- b) Para todo $s, t \geq 0$ se tiene que $E(N(t+s) - N(t) | N(u), u \leq t) = \lambda s$.

4.5. Distribuciones Asociadas a un Proceso de Poisson

Hemos visto que los intervalos de tiempo entre eventos sucesivos, $T_n, n \geq 0$ son v.a.i.i.d. con distribución exponencial de parámetro λ . Los instantes τ_n en los cuales ocurren los eventos, son sumas de las variables anteriores, y en consecuencia tienen distribución $\Gamma(n, \lambda)$.

Teorema 4.5 Sea $N(t), t \geq 0$ un proceso de Poisson de parámetro $\lambda > 0$. Para $0 < u < t$ y $0 \leq k \leq n$,

$$P(N(u) = k | N(t) = n) = \frac{n!}{k!(n-k)!} \left(\frac{u}{t}\right)^k \left(1 - \frac{u}{t}\right)^{n-k}, \quad (4.4)$$

Es decir, condicional a que para el instante t han ocurrido n eventos, la distribución del número de eventos que han ocurrido para el instante $u < t$ es binomial con parámetros n y (u/t) .

Demostración.

$$\begin{aligned} P(N(u) = k | N(t) = n) &= \frac{P(N(u) = k, N(t) = n)}{P(N(t) = n)} \\ &= \frac{P(N(u) = k, N(t) - N(u) = n - k)}{P(N(t) = n)} \\ &= \frac{[e^{-\lambda u} (\lambda u)^k / k!] [e^{-\lambda(t-u)} (\lambda(t-u))^{n-k} / (n-k)!]}{e^{-\lambda t} (\lambda t)^n / n!} \\ &= \frac{n!}{k!(n-k)!} \frac{u^k (t-u)^{n-k}}{t^n}. \end{aligned}$$

■

Ejemplo 4.3

Recordemos que la variable τ_n tiene distribución $\Gamma(n, \lambda)$ y por la observación 1 sabemos que $\lambda\tau_n/2 \sim \Gamma(n, 2) = \chi_{2n}^2$.

Si observamos un proceso de Poisson hasta que se registre un número prefijado m de eventos, el tiempo necesario τ_m puede usarse para construir intervalos de confianza para la intensidad λ del proceso, usando el hecho de que $\lambda\tau_m/2$ tiene distribución χ^2 con $2m$ grados de libertad. Sean $z_{\alpha/2}$ y $z_{1-\alpha/2}$ valores tales que si $Z \sim \chi_{2m}^2$, entonces $P(Z < z_{\alpha/2}) = P(Z > z_{1-\alpha/2}) = \alpha/2$. Tenemos

$$1 - \alpha = P(z_{\alpha/2} \leq \lambda\tau_m/2 \leq z_{1-\alpha/2}) = P\left(\frac{2z_{\alpha/2}}{\tau_m} \leq \lambda \leq \frac{2z_{1-\alpha/2}}{\tau_m}\right).$$

En consecuencia, $(2z_{\alpha/2}/\tau_m, 2z_{1-\alpha/2}/\tau_m)$ es un intervalo de confianza para λ a nivel $1 - \alpha$. ▲

Ejemplo 4.4

Sean N y M dos procesos de Poisson independientes con parámetros respectivos λ y μ . Sean n y m enteros, τ_n el tiempo de espera hasta el n -ésimo evento en el proceso N y γ_m el tiempo de espera hasta el m -ésimo evento en el proceso M . Las variables $\lambda\tau_n/2$ y $\mu\gamma_m/2$ son independientes y tienen distribuciones χ^2 con $2n$ y $2m$ grados de libertad, respectivamente. Por lo tanto, bajo la hipótesis de que $\lambda = \mu$, la variable $m\tau_n/n\gamma_m$ tiene distribución F con $2n$ y $2m$ grados de libertad, y podemos desarrollar una prueba de hipótesis para $\lambda = \mu$. ▲

4.6. Procesos de Poisson Compuestos

Asociamos ahora una variable aleatoria Y_i a cada evento de un proceso de Poisson. Suponemos que las variables Y_i , $i \geq 1$, son i.i.d y también son independientes del proceso. Por ejemplo, el proceso puede representar los carros que llegan a un centro comercial y las variables asociadas, el número de pasajeros que hay en cada uno de ellos; o el proceso puede representar los mensajes que llegan a un computador central para ser transmitidos via internet y las variables Y_i pueden representar el tamaño de los mensajes.

Es natural considerar la suma de las variables Y_i como una variable de interés:

$$S(t) = Y_1 + \cdots + Y_{N(t)}$$

donde ponemos $S(t) = 0$ si $N(t) = 0$. Ya hemos visto que para suma aleatorias, la media es el producto de las medias de N e Y , mientras que la varianza está dada por

$$\text{Var}(S(t)) = E[N(t)] \text{Var}(Y_i) + \text{Var}(N(t))(E[Y_i])^2.$$

En nuestro caso, $N(t) \sim \text{Pois}(\lambda t)$ y por lo tanto, $E[N(t)] = \text{Var}(N(t)) = \lambda t$. En consecuencia tenemos

$$E(S(t)) = \lambda t E(Y_i),$$

$$\text{Var}(S(t)) = \lambda t (\text{Var}(Y_i) + (E[Y_i])^2) = \lambda t E[Y_i^2].$$

Ejemplo 4.5

El número de clientes de una tienda durante el día tiene distribución de Poisson de media 30 y cada cliente gasta un promedio de \$150 con desviación típica de \$50. Por los cálculos anteriores sabemos que el ingreso medio por día es $30 \cdot \$150 = \4.500 . La varianza del ingreso total es

$$30 \cdot [(\$50)^2 + (\$150)^2] = 750.000$$

Sacando la raíz cuadrada obtenemos una desviación típica de \$ 866,02. ▲

La función de distribución para el proceso de Poisson compuesto $S(t)$ puede representarse explícitamente si condicionamos por los valores de $N(t)$. Recordemos que la distribución de una suma de variables independientes es la convolución de las distribuciones: Si Y tiene f.d. G ,

$$G^{(n)}(y) = P(Y_1 + \cdots + Y_n \leq y) = \int_{-\infty}^{\infty} G^{(n-1)}(y-z) dG(z)$$

con

$$G^{(0)}(y) = \begin{cases} 1 & \text{para } y \geq 0, \\ 0 & \text{para } y < 0. \end{cases}$$

Ahora

$$\begin{aligned}
 P(S(t) \leq z) &= P\left(\sum_{k=1}^{N(t)} Y_k \leq z\right) \\
 &= \sum_{n=0}^{\infty} P\left(\sum_{k=1}^{N(t)} Y_k \leq z \mid N(t) = n\right) \frac{(\lambda t)^n e^{-\lambda t}}{n!} \\
 &= \sum_{n=0}^{\infty} \frac{(\lambda t)^n e^{-\lambda t}}{n!} G^{(n)}(z). \tag{4.5}
 \end{aligned}$$

Ejemplo 4.6

Sea $N(t)$ el número de impactos que recibe un sistema mecánico hasta el instante t y sea Y_k el daño o desgaste que produce el k -ésimo impacto. Suponemos que los daños son positivos: $P(Y_k \geq 0) = 1$, y que se acumulan aditivamente, de modo que $S(t) = \sum_{k=1}^{N(t)} Y_k$ representa el daño total hasta el instante t . Supongamos que el sistema continúa funcionando mientras el daño total sea menor que un valor crítico a y en caso contrario falla. Sea T el tiempo transcurrido hasta que el sistema falla, entonces

$$\{T > t\} \quad \text{si y sólo si} \quad \{S(t) < a\}.$$

Teniendo en cuenta esta relación y (4.5) tenemos

$$P(T > t) = \sum_{n=0}^{\infty} \frac{(\lambda t)^n e^{-\lambda t}}{n!} G^{(n)}(a).$$

Para obtener el tiempo promedio hasta que el sistema falle podemos integrar esta probabilidad:

$$\begin{aligned}
 E[T] &= \int_0^{\infty} P(T > t) dt = \sum_{n=0}^{\infty} \left(\int_0^{\infty} \frac{(\lambda t)^n e^{-\lambda t}}{n!} dt \right) G^{(n)}(a) \\
 &= \lambda^{-1} \sum_{n=0}^{\infty} G^{(n)}(a),
 \end{aligned}$$

donde hemos intercambiado series e integrales porque todos los términos son positivos. Esta expresión se simplifica en el caso particular en el cual los daños tienen distribución exponencial de parámetro μ . En este caso la suma $\tau_n = Y_1 + \dots + Y_n$ tiene distribución $\Gamma(n, \mu)$; sea $M(t)$ el proceso de Poisson asociado a estas variables i.i.d. exponenciales, entonces

$$\begin{aligned}
 G^{(n)}(z) &= P(\tau_n \leq z) = P(M(z) \geq n) \\
 &= 1 - \sum_{k=0}^{n-1} \frac{(\mu z)^k e^{-\mu z}}{k!} = \sum_{k=n}^{\infty} \frac{(\mu z)^k e^{-\mu z}}{k!}
 \end{aligned}$$

y

$$\begin{aligned}
 \sum_{n=0}^{\infty} G^{(n)}(a) &= \sum_{n=0}^{\infty} \sum_{k=n}^{\infty} \frac{(\mu a)^k e^{-\mu a}}{k!} = \sum_{k=0}^{\infty} \sum_{n=0}^k \frac{(\mu a)^k e^{-\mu a}}{k!} \\
 &= \sum_{k=0}^{\infty} (k+1) \frac{(\mu a)^k e^{-\mu a}}{k!} = 1 + \mu a.
 \end{aligned}$$

Por lo tanto, cuando Y_i , $i \geq 1$, tienen distribución exponencial de parámetro μ ,

$$E[T] = \frac{1 + \mu a}{\lambda}.$$

▲

4.7. Descomposición de un Proceso de Poisson

En la sección anterior asociamos a cada evento de un proceso de Poisson una variable aleatoria Y_i , ahora vamos a usar estas variables para descomponer el proceso. Sea $N_j(t)$ el número de eventos del proceso que han ocurrido antes de t con $Y_i = j$. Si, por ejemplo, Y_i representa el número de personas en un carro que llega a un centro comercial, $N_j(t)$ representa el número de carros que han llegado antes del instante t con exactamente j personas dentro.

Veamos inicialmente el caso más sencillo, en el cual las variables Y_k son de Bernoulli:

$$P(Y_k = 1) = p, \quad P(Y_k = 0) = 1 - p,$$

para $0 < p < 1$ fijo y $k \geq 1$. Definimos ahora dos procesos, según el valor de las variables Y_k sea 0 ó 1:

$$N_1(t) = \sum_{k=1}^{N(t)} Y_k, \quad y \quad N_0(t) = N(t) - N_1(t).$$

Los valores de $N_1(t)$ sobre intervalos disjuntos son variables aleatorias independientes, $N_1(0) = 0$ y finalmente, el lema 4.2 nos dice que $N_1(t)$ tiene distribución de Poisson con media λpt . Un argumento similar muestra que $N_0(t)$ es un proceso de Poisson con parámetro $\lambda(1-p)$. Lo que resulta más sorprendente es que N_0 y N_1 son procesos independientes. Para ver esto calculemos

$$\begin{aligned} P(N_0(t) = j, N_1(t) = k) &= P(N(t) = j + k, N_1(t) = k) \\ &= P(N_1(t) = k | N(t) = j + k) P(N(t) = j + k) \\ &= \frac{(j+k)!}{j!k!} p^k (1-p)^j \frac{(\lambda t)^{j+k} e^{-\lambda t}}{(j+k)!} \\ &= \left[\frac{e^{-\lambda pt} (\lambda pt)^k}{k!} \right] \left[\frac{e^{-\lambda(1-p)t} (\lambda(1-p)t)^j}{j!} \right] \\ &= P(N_1(t) = k) P(N_0(t) = j) \end{aligned}$$

para $j, k = 0, 1, 2, \dots$

Ejemplo 4.7

Los clientes entran a una tienda de acuerdo a un proceso de Poisson con intensidad de 10 por hora. De manera independiente, cada cliente compra algo con probabilidad $p = 0.3$ o sale de la tienda sin comprar nada con probabilidad $q = 1 - p = 0.7$. ¿Cuál es la probabilidad de que durante la primera hora 9 personas entren a la tienda y que tres de estas personas compren algo y las otras 6 no?

Sea $N_1 = N_1(1)$ el número de clientes que hacen una compra durante la primera hora y $N_0 = N_0(1)$ el número de clientes que entran pero no compran nada. Entonces N_0 y N_1 son v.a.i. de Poisson con parámetros respectivos $(0.7)(10) = 7$ y $(0.3)(10) = 3$. Por lo tanto

$$P(N_0 = 6) = \frac{7^6 e^{-7}}{6!} = 0.149, \quad P(N_1 = 3) = \frac{3^3 e^{-3}}{3!} = 0.224.$$

y

$$P(N_0 = 6, N_1 = 3) = P(N_0 = 6)P(N_1 = 3) = (0.149)(0.224) = 0.0334. \quad \blacktriangle$$

En el caso general las variables Y_k toman valores sobre un conjunto numerable, por ejemplo sobre $\{0, 1, 2, \dots\}$, y el resultado correspondiente es el siguiente teorema, que no demostraremos.

Teorema 4.6 $N_j(t)$ son procesos de Poisson independientes con intensidad $\lambda P(Y_i = j)$.

4.8. Superposición de Procesos de Poisson

La situación inversa a la descomposición de un proceso de Poisson es la superposición de procesos. Ya que un proceso de Poisson puede descomponerse en procesos de Poisson independientes, es razonable esperar que el proceso inverso, la superposición de procesos de Poisson independientes, produzca un proceso de Poisson cuya intensidad sea la suma de las intensidades.

Teorema 4.7 Sean $N_1(t), \dots, N_k(t)$ procesos de Poisson independientes con parámetros $\lambda_1, \dots, \lambda_k$, entonces $N_1(t) + \dots + N_k(t)$ es un proceso de Poisson con parámetro $\lambda_1 + \dots + \lambda_k$.

Demostración. Haremos la demostración para el caso $k = 2$, el caso general se obtiene luego por inducción. Es inmediato que la suma tiene incrementos independientes y que $N_1(0) + N_2(0) = 0$. Para verificar que los incrementos tienen distribución de Poisson con parámetro igual a la suma de los parámetros observamos que si $Y = N_1(t + s) - N_1(s) \sim \text{Pois}(\lambda_1 t)$ y $Z = N_2(t + s) - N_2(s) \sim \text{Pois}(\lambda_2 t)$, entonces

$$\begin{aligned} N(t + s) - N(s) &= [N_1(t + s) - N_1(s)] + [N_2(t + s) - N_2(s)] \\ &= Y + Z \sim \text{Pois}((\lambda_1 + \lambda_2)t). \end{aligned}$$

■

Ejemplo 4.8

Consideremos dos procesos de Poisson, uno con parámetro λ , que representa las llegadas a la meta del equipo rojo, y otro, independiente del anterior y con parámetro μ , que representa las llegadas del equipo verde. ¿Cuál es la probabilidad de que haya 6 llegadas rojas antes que 4 verdes?

Observamos que el evento en cuestión equivale a tener al menos 6 rojos en los primeros 9. Si esto ocurre, tenemos a lo sumo tres verdes antes de la llegada del sexto rojo. Por otro lado, si hay 5 o menos rojos en los primeros 9, entonces tendremos al menos 4 verdes.

Podemos ahora ver el problema en el marco de un proceso de Poisson general que incluye rojos y verdes, y tiene parámetro $\lambda + \mu$. Para cada llegada escogemos al azar el color lanzando una moneda con probabilidad $p = \lambda/(\lambda + \mu)$ para rojo. La probabilidad que nos interesa es

$$\sum_{k=6}^9 \binom{9}{k} p^k (1-p)^{9-k}.$$

En el caso particular en el cual ambos procesos iniciales tienen la misma intensidad $\lambda = \mu$, $p = 1/2$ y la expresión anterior es

$$\frac{1}{512} \sum_{k=6}^9 \binom{9}{k} = \frac{140}{512} = 0.273.$$

▲

4.9. Procesos No Homogéneos

En el corolario 1.3 vimos qué ocurre si la probabilidad de cada evento individual no es homogénea. Si, en cambio, el parámetro del proceso, que representa la intensidad por unidad de tiempo con la cual ocurren los eventos, no es constante a lo largo del tiempo, tenemos un *proceso no-homogéneo*.

Definición 4.4 Decimos que $(N(t), t \geq 0)$ es un proceso de Poisson no homogéneo con tasa $\lambda(s)$, $s \geq 0$ si

1. $N(0) = 0$,

2. $N(t)$ tiene incrementos independientes,
3. $N(s+t) - N(s)$ tiene distribución de Poisson con media $\int_s^{s+t} \lambda(r) dr$.

En este caso los intervalos de tiempo entre eventos sucesivos, T_n , $n \geq 1$, ya no son independientes ni tienen distribución exponencial. Esta es la razón por la cual no usamos nuestra definición inicial para esta generalización. Veamos que esto es efectivamente cierto. Pongamos $\mu(t) = \int_0^t \lambda(s) ds$, entonces $N(t) \sim \text{Pois}(\mu(t))$ y

$$P(T_1 > t) = P(N(t) = 0) = e^{-\mu(t)}.$$

Derivando obtenemos la densidad

$$f_{T_1}(t) = -\frac{d}{dt}P(T_1 > t) = \lambda(t)e^{-\int_0^t \lambda(s) ds} = \lambda(t)e^{-\mu(t)}$$

para $t \geq 0$. La relación anterior se puede generalizar de la siguiente manera

$$f_{T_1, \dots, T_n}(t_1, \dots, t_n) = \lambda(t_1)\lambda(t_1 + t_2) \cdots \lambda(t_1 + \dots + t_n)e^{-\mu(t_1 + \dots + t_n)},$$

lo cual muestra que, en general, las variables T_i no son independientes ni tienen distribución exponencial.

Ejemplo 4.9

Los clientes llegan a una tienda de acuerdo a un proceso de Poisson no-homogéneo con intensidad

$$\lambda(t) = \begin{cases} 2t & \text{para } 0 \leq t < 1, \\ 2 & \text{para } 1 \leq t < 2, \\ 4-t & \text{para } 2 \leq t \leq 4, \end{cases}$$

donde t se mide en horas a partir de la apertura. ¿Cuál es la probabilidad de que dos clientes lleguen durante las primeras dos horas y dos más durante las dos horas siguientes?

Como las llegadas durante intervalos disjuntos son independientes, podemos responder las dos preguntas por separado. La media para las primeras dos horas es $\mu = \int_0^1 2t dt + \int_1^2 2 dt = 3$ y por lo tanto

$$P(N(2) = 2) = \frac{e^{-3}(3)^2}{2!} = 0.2240.$$

Para las siguientes dos horas, $\mu = \int_2^4 (4-t) dt = 2$ y

$$P(N(4) - N(2) = 2) = \frac{e^{-2}(2)^2}{2!} = 0.2707.$$

La probabilidad que nos piden es

$$P(N(2) = 2, N(4) - N(2) = 2) = P(N(2) = 2)P(N(4) - N(2) = 2) = 0.0606$$

▲

4.9.1. Postulados para un proceso de Poisson no-homogéneo

Al igual que para el caso del proceso homogéneo, es posible demostrar que los siguientes postulados implican que el proceso de conteo $N(t)$ es un proceso de Poisson no-homogéneo con función de intensidad $\lambda(t)$, $t \geq 0$:

- (a) $N(0) = 0$.

- (b) $\{N(t), t \geq 0\}$ tiene incrementos independientes.
 (c) $P(N(t+h) - N(t) = 1) = \lambda(t)h + o(h)$.
 (d) $P(N(t+h) - N(t) \geq 2) = o(h)$.

Muestrear en el tiempo un proceso de Poisson ordinario a una tasa que depende del tiempo produce un proceso de Poisson no-homogéneo. Esto es similar a lo que vimos para la descomposición de un proceso de Poisson sólo que ahora la probabilidad de observar un evento del proceso original no es una constante p como ocurría antes, sino que depende del tiempo: $p(t)$.

Sea $\{N(t), t \geq 0\}$ un proceso de Poisson con intensidad constante λ y supongamos que un evento que ocurre en el instante t se observa con probabilidad $p(t)$, independientemente de lo que haya ocurrido antes. Llamemos $M(t)$ al proceso de los eventos que hemos logrado contar hasta el instante t , entonces $\{M(t), t \geq 0\}$ es un proceso de Poisson no-homogéneo con función de intensidad $\lambda(t) = \lambda p(t)$. Podemos verificar esta afirmación comprobando que se satisfacen los axiomas anteriores.

- (a) $M(0) = 0$.
 (b) El número de eventos que contamos en el intervalo $(t, t+h]$ depende únicamente de los eventos del proceso de Poisson N que ocurren en $(t, t+h]$, que es independiente de lo que haya ocurrido antes de t . En consecuencia el número de eventos observados en $(t, t+h]$ es independiente del proceso de eventos observados hasta el tiempo t , y por lo tanto M tiene incrementos independientes.
 (c) Condicionando sobre $N((t, t+h])$:

$$\begin{aligned} P(M((t, t+h]) = 1) &= P(M((t, t+h]) = 1 | N((t, t+h]) = 1)P(N((t, t+h]) = 1) \\ &\quad + P(M((t, t+h]) = 1 | N((t, t+h]) \geq 2)P(N((t, t+h]) \geq 2) \\ &= P(M((t, t+h]) = 1 | N((t, t+h]) = 1)\lambda h + o(h) \\ &= p(t)\lambda h + o(h) \end{aligned}$$

- (d) $P(M((t, t+h]) \geq 2) \leq P(N((t, t+h]) \geq 2) = o(h)$.

Hay un recíproco (parcial) para este resultado: todo proceso no-homogéneo de Poisson con intensidad acotada se puede obtener a partir de un proceso homogéneo muestreado en el tiempo. Para ver esto necesitamos la siguiente proposición que enunciamos sin demostración

Proposición 4.1 Sean $N(t)$, $t \geq 0$ y $M(t)$, $t \geq 0$ procesos de Poisson independientes no-homogéneos, con funciones de intensidad respectivas $\alpha(t)$ y $\beta(t)$ y sea $S(t) = N(t) + M(t)$. Entonces

- (a) $\{S(t), t \geq 0\}$ es un proceso de Poisson no-homogéneo con función de intensidad $\lambda(t) = \alpha(t) + \beta(t)$.
 (b) Dado que un evento del proceso S ocurre en el instante t entonces, independientemente de lo que haya ocurrido antes de t , el evento en t viene del proceso N con probabilidad $\alpha(t)/(\alpha(t) + \beta(t))$.

Demostración. Ver S.M. Ross, *Introduction to Probability Models* 10th. Ed. p. 340.

Supongamos ahora que $\{N(t), t \geq 0\}$ es un proceso de Poisson no-homogéneo con función de intensidad acotada $\lambda(t)$ tal que $\lambda(t) \leq \lambda$ para todo t . Sea $\{M(t), t \geq 0\}$ otro proceso de Poisson no-homogéneo con intensidad $\mu(t) = \lambda - \lambda(t)$ e independiente de $N(t)$. Por la proposición anterior tenemos que $\{N(t), t \geq 0\}$ se puede considerar como el proceso que se obtiene a partir del proceso homogéneo $\{N(t) + M(t), t \geq 0\}$, donde un evento que ocurre en el tiempo t es observado con probabilidad $p(t) = \lambda(t)/\lambda$.

La función $\mu(t)$ definida por

$$\mu(t) = \int_0^t \lambda(s) ds$$

es continua y no decreciente y representa el valor esperado del número de eventos que ocurren en el intervalo $[0, t]$, $E(N(t)) = \mu(t)$. Definimos su inversa generalizada $\nu(t)$ por

$$\nu(t) = \inf\{s : \mu(s) > t\}, \quad t \geq 0.$$

Usando estas funciones tenemos el siguiente resultado

Teorema 4.8 *Sea N un proceso de Poisson no homogéneo y sea $M(t) = N(\nu(t))$, $t \geq 0$. Entonces M es un proceso de Poisson homogéneo con intensidad 1.*

Demostración. Fijamos s, t y ponemos $t' = \nu(t)$, $t' + s' = \nu(t + s)$, $s' = \nu(t + s) - \nu(t)$. Entonces

$$\begin{aligned} E(M(t + s) - M(t) | M(u), u \leq t) &= E(N(t' + s') - N(t') | N(u), u \leq t) \\ &= E(N(t' + s') - N(t')) = \mu(t' + s') - \mu(t') \\ &= t + s - t = s. \end{aligned}$$

Por el teorema 4.4 obtenemos el resultado. ■

Denotemos por τ_n el instante en el cual ocurre el n -ésimo evento de un proceso no-homogéneo $\{N(t), t \geq 0\}$. Entonces

$$\begin{aligned} P(t < \tau_n < t + h) &= P(N(t) = n - 1, \text{ y al menos un evento ocurre en } (t, t + h)) \\ &= P(N(t) = n - 1, \text{ y un evento ocurre en } (t, t + h)) + o(h) \\ &= P(N(t) = n - 1)P(\text{un evento ocurre en } (t, t + h)) + o(h) \\ &= e^{-\mu(t)} \frac{\mu(t)^{n-1}}{(n-1)!} [\lambda(t)h + o(h)] + o(h) \\ &= \lambda(t)e^{-\mu(t)} \frac{(\mu(t))^{n-1}}{(n-1)!} h + o(h). \end{aligned}$$

Dividiendo por h y haciendo $h \rightarrow 0$ obtenemos que la densidad de esta variable es

$$f_{\tau_n}(t) = \lambda(t)e^{-\mu(t)} \frac{(\mu(t))^{n-1}}{(n-1)!}.$$

4.9.2. Procesos de Cox

Un *proceso de Cox* es un proceso de Poisson no-homogéneo en el cual la intensidad ($\lambda(t)$, $t \geq 0$) es a su vez un proceso aleatorio. En general, los incrementos sobre intervalos disjuntos para un proceso de Cox no son independientes.

Sea $(N(t), t \geq 0)$ un proceso de Poisson con intensidad constante $\lambda = 1$. El proceso de Cox más simple requiere seleccionar el valor de una v.a. Θ y luego observar el proceso $M(t) = N(\Theta t)$. Dado el valor de Θ , M es, condicionalmente, un proceso de Poisson con intensidad constante $\lambda = \Theta$. Θ es aleatoria y, típicamente, no es observable. Si Θ tiene distribución continua con densidad $f(\theta)$ entonces, por la ley de probabilidad total obtenemos la distribución marginal

$$P(M(t) = k) = \int_0^\infty \frac{(\theta t)^k e^{-\theta t}}{k!} f(\theta) d\theta.$$

4.10. La Distribución Uniforme

Consideremos un segmento de longitud t y escojamos sobre él n puntos al azar, de manera independiente y con distribución uniforme, es decir, consideramos una muestra aleatoria simple de tamaño n de la distribución uniforme sobre $[0, t]$. Llamemos U_1, \dots, U_n a estas variables. La densidad de probabilidad de c/u de ellas es

$$f_U(u) = \frac{1}{t}, \quad \text{para } 0 \leq u \leq t.$$

Consideremos ahora esta misma muestra pero ordenada y llamemos $U_{(i)}$, $1 \leq i \leq n$, a sus valores, de modo que

$$U_{(1)} \leq U_{(2)} \leq \dots \leq U_{(n)}.$$

Como veremos a continuación, la densidad conjunta de $U_{(1)}, U_{(2)}, \dots, U_{(n)}$ es

$$f_{U_{(1)}, \dots, U_{(n)}}(u_1, \dots, u_n) = \frac{n!}{t^n} \quad \text{para } 0 < u_1 < \dots < u_n \leq t \quad (4.6)$$

Dada cualquier colección X_1, \dots, X_n de v.a.i.i.d. las variables $X_{(1)}, \dots, X_{(n)}$, que corresponden a los valores de las variables originales pero ordenadas

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)},$$

se conocen como los *estadísticos de orden*. El próximo teorema nos muestra cómo se obtiene su distribución en el caso general.

Teorema 4.9 Sean $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ los estadísticos de orden para una muestra aleatoria simple de variables aleatorias continuas con densidad $f(x)$. La densidad conjunta de los estadísticos de orden es

$$g_n(x_{(1)}, x_{(2)}, \dots, x_{(n)}) = \begin{cases} n! \prod_{i=1}^n f(x_{(i)}), & x_{(1)} < \dots < x_{(n)}, \\ 0, & \text{en otro caso.} \end{cases} \quad (4.7)$$

Demostración. Haremos la prueba para el caso general n y la ilustraremos detalladamente cuando $n = 2$. Definimos los conjuntos $A = \{(x_1, \dots, x_n) : x_i \in \mathbb{R}, x_i \neq x_j \text{ para } i \neq j\}$ y $B = \{(x_{(1)}, \dots, x_{(n)}) : -\infty < x_{(1)} < \dots < x_{(n)} < \infty\}$. La transformación que define los estadísticos de orden es una función de A a B pero no es 1-1, ya que cualquiera de las $n!$ permutaciones de los valores observados produce los mismos estadísticos de orden. Por ejemplo, cuando $n = 2$, $(x_1, x_2) = (1.6, 3.4)$ y $(x_1, x_2) = (3.4, 1.6)$ ambos producen $(x_{(1)}, x_{(2)}) = (1.6, 3.4)$

Si dividimos A en $n!$ subconjuntos de modo que c/u corresponda a un orden particular de la muestra observada, vemos que ahora la transformación que define los estadísticos de orden define una biyección de cada uno de estos conjuntos al conjunto B . Como ilustración, cuando $n = 2$, dividimos A en $A_1 = \{(x_1, x_2) : -\infty < x_1 < x_2 < \infty\}$ y $A_2 = \{(x_1, x_2) : -\infty < x_2 < x_1 < \infty\}$. En el primero de estos conjuntos la transformación es $x_1 = x_{(1)}$ y $x_2 = x_{(2)}$. Por lo tanto el valor absoluto del Jacobiano de la transformación es

$$|J_1| = \left| \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} \right| = 1$$

mientras que en A_2 la transformación es $x_1 = x_{(2)}$ y $x_2 = x_{(1)}$ y el valor absoluto del Jacobiano de la transformación es

$$|J_2| = \left| \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} \right| = 1.$$

En el caso general se ve similarmente que el Jacobiano de cada una de las biyecciones de una de las $n!$ regiones en que dividimos a A sobre B , tiene valor absoluto igual a 1. La densidad conjunta de los

estadísticos de orden es, en consecuencia, la suma de las contribuciones de cada conjunto de la partición. En el caso particular $n = 2$ tenemos para $-\infty < x_{(1)} < x_{(2)} < \infty$,

$$\begin{aligned} g(x_{(1)}, x_{(2)}) &= f(x_{(1)})f(x_{(2)})|J_1| + f(x_{(2)})f(x_{(1)})|J_2| \\ &= 2f(x_{(1)})f(x_{(2)}). \end{aligned}$$

Para n cualquiera, la densidad conjunta (4.7) se obtiene tomando en cuenta que la contribución de cada una de las $n!$ particiones es $\prod_{i=1}^n f(x_{(i)})$. ■

En el caso particular de la distribución uniforme obtenemos la ecuación (4.6) para la densidad de los estadísticos de orden.

Teorema 4.10 Sean τ_1, τ_2, \dots los instantes en los cuales ocurren los sucesivos eventos de un proceso de Poisson de parámetro λ . Dado que $N(t) = n$, las variables τ_1, τ_2, \dots tienen la misma distribución conjunta que los estadísticos de orden de n v.a.i. con distribución uniforme en $[0, t]$.

Demostración. Consideremos un proceso de Poisson $\{N(t), t \geq 0\}$ y supongamos que en el intervalo $[0, t]$ han ocurrido n eventos. Sea $[t_i, t_i + h_i]$, $1 \leq i \leq n$, una sucesión de intervalos disjuntos en $[0, t]$. Dado que han ocurrido n eventos hasta t , la probabilidad de que ocurra exactamente un evento en cada uno de los intervalos que listamos, y ningún evento fuera de ellos es

$$\begin{aligned} P(t_1 \leq \tau_1 \leq t_1 + h_1, \dots, t_n \leq \tau_n \leq t_n + h_n | N(t) = n) \\ &= \frac{\lambda h_1 e^{-\lambda h_1} \dots \lambda h_n e^{-\lambda h_n} e^{-\lambda(t-h_1-\dots-h_n)}}{e^{-\lambda t} (\lambda t)^n / n!} \\ &= \frac{n!}{t^n} h_1 \dots h_n. \end{aligned} \tag{4.8}$$

Pero, por definición de la función de densidad, el lado izquierdo de (4.8) es, para valores pequeños de h_i , $1 \leq i \leq n$, aproximadamente igual a

$$f_{\tau_1, \dots, \tau_n | N(t)=n}(t_1, \dots, t_n) h_1 \dots h_n.$$

Esto es suficiente para demostrar que la densidad condicional de los tiempos τ_i dado que han ocurrido n eventos en el intervalo $[0, t]$ es igual a $n!/t^n$. ■

Ejemplo 4.10

El teorema anterior nos da una manera de probar si un conjunto de observaciones es de Poisson. Supongamos que hemos observado el proceso por un período de tiempo t durante el cual han ocurrido n eventos. Sea τ_1, \dots, τ_n los instantes en los cuales han ocurrido los eventos y sea W_1, \dots, W_n una permutación de los instantes escogida al azar. Si los eventos ocurrieron de acuerdo a un proceso de Poisson, las variables W_i son independientes y tienen distribución uniforme sobre el intervalo $[0, t]$. Por lo tanto podemos hacer un test sobre estas variables para ver si cumplen esta hipótesis, para lo cual podemos hacer una prueba de Kolmogorov-Smirnov o de Cramér-von Mises. También es posible usar el TCL ya que para valores moderados o grandes de n , la suma $S_n = \sum_1^n U_i$ es aproximadamente normal con media $E(S_n) = n E(U_1) = nt/2$ y varianza $\text{Var}(S_n) = n \text{Var}(U_1) = nt^2/12$.

Por ejemplo, si en $t = 10$ minutos de observación, $n = 12$ eventos ocurren, entonces la suma S_{12} de los instantes en los cuales ocurren los eventos es aproximadamente normal con media 60 y desviación estándar 10. En consecuencia, si S_{12} satisface las desigualdades

$$60 - (1.96)10 \leq S_{12} \leq 60 + (1.96)10,$$

aceptaríamos la hipótesis de que los eventos provienen de un proceso de Poisson con un nivel de significación de 95%. ▲

Ejemplo 4.11

Consideremos una masa de material radioactivo que emite partículas alfa de acuerdo a un proceso de Poisson de intensidad λ . Cada partícula existe por un período aleatorio de tiempo y luego desaparece. Supongamos que los tiempos de vida sucesivos Y_1, Y_2, \dots de las diferentes partículas son v.a.i. con distribución común $G(y) = P(Y_k \leq y)$. Sea $M(t)$ el número de partículas que existen en el instante t . Queremos hallar la distribución de probabilidad de $M(t)$ bajo la condición de que $M(0) = 0$.

Sea $N(t)$ el número de partículas creadas hasta el tiempo t . Observamos que $M(t) \leq N(t)$. Dado que $N(t) = n$ sean $\tau_1, \dots, \tau_n \leq t$ los instantes en los cuales se crean las partículas. La partícula k existe en el instante t si y sólo si $\tau_k + Y_k \geq t$. Por lo tanto

$$P(M(t) = m | N(t) = n) = P\left(\sum_{k=1}^n \mathbf{1}_{\{\tau_k + Y_k \geq t\}} = m | N(t) = n\right).$$

Usando el teorema 4.10 y la simetría entre las partículas tenemos

$$P\left(\sum_{k=1}^n \mathbf{1}_{\{\tau_k + Y_k \geq t\}} = m | N(t) = n\right) = P\left(\sum_{k=1}^n \mathbf{1}_{\{U_k + Y_k \geq t\}} = m\right) \quad (4.9)$$

donde U_1, U_2, \dots, U_n son v.a.i. con distribución uniforme en $[0, t]$. El lado derecho de (4.9) es una distribución binomial con parámetros n y

$$\begin{aligned} p &= P(U_k + Y_k \geq t) = \frac{1}{t} \int_0^t P(Y_k \geq t - u) du \\ &= \frac{1}{t} \int_0^t [1 - G(t - u)] du = \frac{1}{t} \int_0^t [1 - G(z)] dz. \end{aligned} \quad (4.10)$$

Escribiendo explícitamente la distribución binomial tenemos

$$P(M(t) = m | N(t) = n) = \binom{n}{m} p^m (1 - p)^{n-m}$$

con p dado por la ecuación (4.10). Finalmente

$$\begin{aligned} P(M(t) = m) &= \sum_{n=m}^{\infty} P(M(t) = m | N(t) = n) P(N(t) = n) \\ &= \sum_{n=m}^{\infty} \frac{n!}{m!(n-m)!} p^m (1-p)^{n-m} \frac{(\lambda t)^n e^{-\lambda t}}{n!} \\ &= e^{-\lambda t} \frac{(\lambda p t)^m}{m!} \sum_{n=m}^{\infty} \frac{(1-p)^{n-m} (\lambda t)^{n-m}}{(n-m)!}. \end{aligned} \quad (4.11)$$

La suma es una serie exponencial que se reduce a

$$\sum_{n=m}^{\infty} \frac{(1-p)^{n-m} (\lambda t)^{n-m}}{(n-m)!} = \sum_{j=0}^{\infty} \frac{[\lambda t(1-p)]^j}{j!} = e^{\lambda t(1-p)}$$

y usando esto (4.11) se reduce a

$$P(M(t) = m) = \frac{e^{-\lambda p t} (\lambda p t)^m}{m!} \quad \text{para } m \geq 0,$$

es decir, el número de partículas que existen en el instante t tiene distribución de Poisson de media

$$\lambda pt = \lambda \int_0^t (1 - G(y)) dy. \quad (4.12)$$

Veamos que ocurre cuando $t \rightarrow \infty$. Sea $\mu = E[Y_k] = \int_0^\infty (1 - G(y)) dy$ la vida media de una partícula alfa. Vemos a partir de (4.12) que cuando $t \rightarrow \infty$, la distribución de $M(t)$ converge a una distribución de Poisson con parámetro $\lambda\mu$. Por lo tanto, asintóticamente, la distribución de probabilidad para el número de partículas que existen depende únicamente de la vida media μ . ▲

Ejemplo 4.12

Un procedimiento común en estadística es observar un número fijo n de v.a.i.i.d. X_1, \dots, X_n y usar su media muestral

$$\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$$

como estimador de la media de la población $E[X_1]$. Consideremos en cambio la siguiente situación: Una compañía nos pide estimar el tiempo medio de vida en servicio de cierto componente de una máquina. La máquina ha estado funcionando por dos años y se observó que el componente original duró 7 meses, el siguiente duró 5 meses y el tercero 9. No se observaron fallas en los tres meses restantes del período de observación. La pregunta es si es correcto estimar la vida media en servicio por el promedio observado $(7 + 9 + 5)/3 = 7$ meses.

Este ejemplo presenta una situación en la cual el tamaño de la muestra no está fijo de antemano sino que se determina a través de una 'cuota' prefijada $t > 0$: Observamos una sucesión de variables i.i.d. X_1, X_2, \dots y continuamos el muestreo mientras la suma de observaciones sea menor que la cuota t . Llamemos $N(t)$ al tamaño de la muestra,

$$N(t) = \text{máx}\{n \geq 0 : X_1 + \dots + X_n < t\}.$$

La media muestral es

$$\bar{X}_{N(t)} = \frac{X_1 + \dots + X_{N(t)}}{N(t)}.$$

Puede suceder que $X_1 \geq t$, en este caso $N(t) = 0$ y no podemos definir la media muestral. Por lo tanto tenemos que suponer que $N(t) \geq 1$. Una pregunta importante en estadística matemática es si este estimador es insesgado. Es decir, ¿cómo se relaciona el valor esperado de este estimador con el valor esperado de $E[X_1]$?

En general, determinar el valor esperado de la media muestral en esta situación es muy difícil. Es posible hacerlo, sin embargo, en el caso en el cual los sumandos tienen distribución exponencial de parámetro común λ , de modo que $N(t)$ es un proceso de Poisson. La clave es usar el teorema 4.10 para evaluar la esperanza condicional

$$E[\tau_{N(t)} | N(t) = n] = E[\text{máx}(U_1, \dots, U_n)] = t \left(\frac{n}{n+1} \right),$$

donde U_1, \dots, U_n son independientes y tienen distribución uniforme sobre el intervalo $[0, t]$. Observamos además que

$$P(N(t) = n | N(t) > 0) = \frac{(\lambda t)^n e^{-\lambda t}}{n!(1 - e^{-\lambda t})}.$$

Entonces,

$$\begin{aligned}
 E\left[\frac{\tau_{N(t)}}{N(t)} \mid N(t) > 0\right] &= \sum_{n=1}^{\infty} E\left[\frac{\tau_{N(t)}}{n} \mid N(t) = n\right] P(N(t) = n \mid N(t) > 0) \\
 &= \sum_{n=1}^{\infty} t \binom{n}{n+1} \left(\frac{1}{n}\right) \left(\frac{(\lambda t)^n e^{-\lambda t}}{n!(1 - e^{-\lambda t})}\right) \\
 &= \frac{1}{\lambda} \left(\frac{1}{e^{\lambda t} - 1}\right) \sum_{n=1}^{\infty} \frac{(\lambda t)^{n+1}}{(n+1)!} \\
 &= \frac{1}{\lambda} \left(\frac{1}{e^{\lambda t} - 1}\right) (e^{\lambda t} - 1 - \lambda t) \\
 &= \frac{1}{\lambda} \left(1 - \frac{\lambda t}{e^{\lambda t} - 1}\right).
 \end{aligned}$$

Podemos ver el efecto de este tipo de muestreo si expresamos el resultado anterior en términos del cociente del sesgo entre el verdadero valor de la esperanza $E(X_1) = 1/\lambda$. Tenemos

$$\frac{E[X_1] - E[\bar{X}_{N(t)}]}{E[X_1]} = \frac{\lambda t}{e^{\lambda t} - 1} = \frac{E[N(t)]}{e^{E[N(t)]} - 1}.$$

El lado izquierdo representa la fracción del sesgo y el lado derecho expresa esta fracción como función del tamaño esperado de la muestra para este tipo de muestreo. La siguiente tabla presenta algunos valores:

$E(N(t))$	Fracción
1	0.58
2	0.31
3	0.16
4	0.07
5	0.03
6	0.015
10	0.0005

En el ejemplo inicial, observamos $N(t) = 3$ fallas en un período de un año y en la tabla anterior vemos que la fracción del sesgo es del orden de 16%. Como mencionamos $\bar{X}_{N(t)} = 7$, una estimación más adecuada podría ser $7/0.84 = 8.33$, que intenta corregir, en promedio, el sesgo debido al método de muestreo. ▲

Los resultados anteriores se pueden generalizar al caso de procesos no-homogéneos con intensidad $\lambda(r)$. Sea $\mu(t) = \int_0^t \lambda(r) dr$ y $g(r) = \lambda(r)/\mu(t)$ para $0 < r < t$.

Teorema 4.11 Sean U_1, U_2, \dots, U_n v.a.i. con densidad g . Dado que $N(t) = n$, los tiempos de llegadas τ_1, \dots, τ_n tienen la misma distribución que los estadísticos de orden correspondientes a las variables U_1, \dots, U_n .

La demostración es similar a la del caso homogéneo y queda como ejercicio.

4.11. Procesos Espaciales de Poisson

Sea S un conjunto en un espacio n -dimensional y sea \mathcal{A} una familia de subconjuntos de S . Un proceso puntual en S es un proceso estocástico $N(A)$ indexado por los conjuntos A en \mathcal{A} , que tiene como valores posibles los elementos del conjunto $\{0, 1, 2, \dots\}$. La idea es que los puntos se encuentran dispersos en S

de manera aleatoria y $N(A)$ cuenta los puntos en el conjunto A . Como $N(A)$ es una función que cuenta, hay varias condiciones obvias que debe satisfacer. Por ejemplo, si A y B son disjuntos, están en \mathcal{A} y su unión $A \cup B$ también está en \mathcal{A} entonces necesariamente $N(A \cup B) = N(A) + N(B)$.

El caso unidimensional, en el cual S es la semirecta positiva y \mathcal{A} es la colección de los intervalos de la forma $A = (s, t]$ para $0 \leq s < t$, lo consideramos al estudiar el proceso puntual de Poisson. La generalización al plano o al espacio tridimensional tiene interés cuando consideramos la distribución espacial de estrellas o galaxias en Astronomía, de plantas o animales en Ecología, de bacterias sobre una placa de laboratorio en Biología o de defectos sobre una superficie en Ingeniería.

Definición 4.5 Sea S un subconjunto de \mathbb{R}, \mathbb{R}^2 o \mathbb{R}^3 . Sea \mathcal{A} una familia de subconjuntos de S y para cualquier $A \in \mathcal{A}$ sea $|A|$ el tamaño (longitud, área o volumen) de A . Entonces $\{N(A) : A \in \mathcal{A}\}$ es un proceso puntual homogéneo de Poisson de intensidad $\lambda > 0$ si

1. Para todo $A \in \mathcal{A}$, la variable $N(A)$ tiene distribución de Poisson con parámetro $\lambda|A|$.
2. Para toda colección finita A_1, \dots, A_n de conjuntos disjuntos de \mathcal{A} , las variables $N(A_1), \dots, N(A_n)$ son independientes.

Ejemplo 4.13

Una zona de Londres se dividió en $N = 576 = 24 \times 24$ áreas de $1/4$ de kilómetro cuadrado. Esta área recibió 535 impactos de bomba durante la II Guerra Mundial, un promedio de $535/576 = 0.9288$ por cuadrado. La siguiente tabla presenta N_k , el número de cuadrados que recibieron el impacto de exactamente k bombas y lo compara con el valor esperado si los impactos tuviesen una distribución de Poisson con esta media

k	0	1	2	3	4	≥ 5
N_k	229	211	93	35	7	1
Poisson	226.74	211.39	98.54	30.62	7.14	1.57

El ajuste es muy bueno y puede ser verificado haciendo una prueba χ^2 . ▲

Muchas de las propiedades que hemos estudiado para el caso unidimensional tienen una extensión natural para el caso de dimensiones mayores. Veamos, como ejemplo, la propiedad de uniformidad de la distribución de la ubicación de los puntos en una región dado que conocemos el número de puntos. Consideremos inicialmente una región A de tamaño positivo $|A| > 0$ y supongamos que sabemos que A contiene exactamente un punto: $N(A) = 1$. Entonces, la distribución de este punto es uniforme en el siguiente sentido:

$$P(N(B) = 1 | N(A) = 1) = \frac{|B|}{|A|}$$

para cualquier $B \subset A$. Para ver esto escribimos $A = B \cup C$ donde $C = A \setminus B$ y en consecuencia $N(B)$ y $N(C)$ son v.a.i. de Poisson con medias respectivas $\lambda|B|$ y $\lambda|C|$. Entonces

$$\begin{aligned} P(N(B) = 1 | N(A) = 1) &= \frac{P(N(B) = 1, N(C) = 0)}{P(N(A) = 1)} \\ &= \frac{\lambda|B|e^{-\lambda|B|}e^{-\lambda|C|}}{\lambda|A|e^{-\lambda|A|}} \\ &= \frac{|B|}{|A|}. \end{aligned}$$

Para generalizar este resultado consideremos una región A de tamaño positivo $|A| > 0$ que contiene $N(A) = n \geq 1$ puntos. Entonces estos puntos son independientes y están distribuidos uniformemente en

A en el sentido de que para cualquier partición disjunta A_1, \dots, A_m de A , donde $A = A_1 \cup \dots \cup A_m$ y para cualesquiera enteros positivos k_1, \dots, k_m con $k_1 + \dots + k_m = n$, tenemos

$$P(N(A_1) = k_1, \dots, N(A_m) = k_m | N(A) = n) = \frac{n!}{k_1! \cdots k_m!} \left(\frac{|A_1|}{|A|} \right)^{k_1} \cdots \left(\frac{|A_m|}{|A|} \right)^{k_m}$$

es decir, dado que $N(A) = n$, la distribución de $N(A_1), \dots, N(A_m)$ es multinomial.

Otras propiedades de los procesos de Poisson homogéneos en el plano son las siguientes:

1. Si N es un proceso de Poisson homogéneo en el plano, $\mathbf{x} \in \mathbb{R}^2$, para cualquier n y conjuntos B_1, \dots, B_n en \mathcal{A} , los vectores $(N(B_1 + \mathbf{x}), \dots, N(B_n + \mathbf{x}))$ y $(N(B_1), \dots, N(B_n))$ tienen la misma distribución.
2. N es invariante bajo rotaciones: Si φ es una rotación respecto al origen de coordenadas, $N(B)$ y $N(\varphi(B))$ tienen la misma distribución, para cualquier $B \in \mathcal{A}$.
3. Si a cada punto del proceso N le asociamos una v.a. Y de Bernoulli con probabilidad de éxito p , y estas variables son independientes entre sí y del proceso N , el proceso M formado por los puntos en los cuales las variables Y valen 1 es un proceso de Poisson homogéneo de intensidad $p\lambda$.
4. Si N y M son dos procesos de Poisson independientes y homogéneos con intensidades λ y μ , respectivamente, el proceso $S = N + M$ es un proceso de Poisson con intensidad $\lambda + \mu$.

Ejemplo 4.14

Consideremos un proceso de Poisson compuesto sobre la recta y supongamos que las variables asociadas U_1, U_2, \dots tienen distribución uniforme $\mathcal{U}(0, 1)$. Esto nos da una sucesión de puntos sobre la banda $\{(t, u) : 0 \leq t < \infty, 0 < u < 1\}$. En este caso, como las U_i tienen distribución continua, el número de puntos sobre una recta fija $\{(t, u) : u = x\}$ es 0 con probabilidad uno.

Si en lugar de rectas consideramos bandas

$$\{(t, u) : 0 \leq t < \infty, a_{m-1} < u \leq a_m\}$$

y si $0 \leq a_0 < a_1 < \dots < a_n \leq 1$, entonces los puntos en las bandas $1 \leq m \leq n$ son independientes.

Usando esta propiedad y la propiedad de incrementos independientes de los procesos de Poisson unidimensionales obtenemos el siguiente resultado: Sean $R_m = \{(t, u) : a_m < t \leq b_m, c_m < u \leq d_m\}$ rectángulos con $a_m \geq 0$ y $0 \leq c_m < d_m \leq 1$ y sea $N(R_m)$ el número de puntos (T_i, U_i) que están en el rectángulo R_m . Si los rectángulos R_m son disjuntos entonces las variables $N(R_m)$ son independientes y tienen distribución de Poisson con media

$$\lambda_m = \lambda(b_m - a_m)(d_m - c_m) = \lambda|R_m|.$$

▲

4.11.1. Procesos no homogéneos en el plano

Decimos que N es un proceso no homogéneo de Poisson con intensidad $\lambda(x, y)$ si cuando R_m , $1 \leq m \leq n$ son conjuntos disjuntos, las variables $N(R_m)$ son independientes con distribución de Poisson de media

$$\mu(R_m) = \int_{(x,y) \in R_m} \lambda(x, y) dy dx.$$

Si la integral vale ∞ , el número de puntos es ∞ .

Ejemplo 4.15

Sean τ_1, τ_2, \dots los instantes en los que ocurren eventos de un proceso de Poisson con intensidad λ . Supongamos que si un evento ocurre en el instante s , lo registramos con probabilidad $p(s)$. El proceso de eventos registrados es un proceso de Poisson no-homogéneo de intensidad $\lambda p(s)$.

Para ver esto, asociamos a cada τ_i una v.a.i. con distribución uniforme sobre $(0, 1)$, y aceptamos el punto si $U_i < p(\tau_i)$. El número de puntos τ_i aceptados en un intervalo (a, b) es igual al número de puntos (τ_i, U_i) que caen en la región

$$\{(t, u) : a < t < b, 0 < u < p(t)\}.$$

En consecuencia, este número es Poisson con media igual al producto de λ por el área del conjunto, es decir $\lambda \int_a^b p(s) ds$. Es claro que el número de eventos en intervalos disjuntos son independientes, de modo que tenemos un proceso de Poisson no-homogéneo. Este resultado nos da una manera de construir un proceso de Poisson no-homogéneo de intensidad dada. ▲