

Cadenas de Markov en tiempo continuo. Introducción a las Martingalas y al Movimiento Browniano.

Cadenas de Markov en tiempo continuo.

En las cadenas de Markov en tiempo discreto, utilizábamos como índice un *tiempo discreto* $n = 0, 1, 2, \dots$ y se deducían numerosas propiedades.

Las nociones de cadenas de Markov se puede extender a un tiempo continuo $t \geq 0$.

En tiempo continuo es complicado definir la distribución condicionada, dados todos los valores X_r para $r \leq s$, por lo que decimos en su lugar que X_t , $t \geq 0$, es una cadena de Markov si para cualquier $0 \leq s_0 < s_1 < \dots < s_n < s$ y posibles estados i_0, \dots, i_n, i, j se tiene que

$$P(X_{t+s} = j \mid X_s = i, X_{s_n} = i_n, \dots, X_{s_0} = i_0) = P(X_t = j \mid X_0 = i)$$

es decir, dado el estado actual, el resto del pasado es irrelevante para predecir el futuro. En la definición se observa que la probabilidad de ir desde i en el tiempo s hasta j en el tiempo $s + t$ solo depende de t , esto es, de las diferencias de tiempos.

Ejemplo. Sea $N(t)$, $t \geq 0$, un proceso de Poisson con tasa λ , y sea Y_n una cadena de Markov discreta con probabilidad de transición, digamos, $u(i, j)$.

Entonces el proceso definido como $X_t = Y_{N(t)}$ es una cadena de Markov en tiempo continuo, es decir, X_t realiza un salto de acuerdo a $u(i, j)$ en cada llegada de $N(t)$.

Esto se sigue de manera intuitiva de la falta de memoria de la distribución exponencial. Si $X_n = i$, entonces, independientemente de lo que ocurriese en el pasado, el tiempo hasta el siguiente salto se distribuye de manera exponencial con tasa λ , de modo que la cadena irá al estado j con probabilidad $u(i, j)$.

En las cadenas discretas, se tenía la matriz de transición y de modo paralelo, aquí, se tiene la probabilidad de transición para $t > 0$ definida como

$$p_t(i, j) = P(X_t = j \mid X_0 = i)$$

que en el ejemplo, como $N(t)$ se distribuye como una Poisson con media λt , tiene la forma

$$p_t(i, j) = \sum_{n=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^n}{n!} u^n(i, j)$$

donde $u^n(i, j)$ es la potencia n -ésima de la probabilidad de transición $u(i, j)$. Es decir, la probabilidad de estar en j en el tiempo t , es igual a que aparezcan n llegadas de $N(t)$ por la probabilidad de transición en $N(t) = n$ etapas: $u^n(i, j)$.

La probabilidad de transición satisface la ecuación de Chapman-Kolmogorov:

$$\sum_k p_s(i, k) p_t(k, j) = p_{s+t}(i, j).$$

Esto se justifica porque para ir de i a j en el tiempo $s + t$, debe estar en algún estado k en el tiempo s , y la propiedad markoviana implica que las dos partes del recorrido son independientes entre sí. De este modo, si se conoce la probabilidad de transición para $t < t_0$ para cualquier $t_0 > 0$, entonces se sabe para todo t .

Se puede demostrar que las probabilidades de transición p_t se pueden determinar a partir de las derivadas en 0 de las propias $p_t(i, j)$:

$$q(i, j) = \lim_{h \rightarrow 0} \frac{p_h(i, j)}{h}$$

para $j \neq i$. A esta cantidad, $q(i, j)$ se denomina *tasa de saltos* de i a j y tiene el sentido de una velocidad de salto de un estado a otro.

En el ejemplo,

$$q(i, j) = \lim_{h \rightarrow 0} \frac{p_h(i, j)}{h} \approx \lambda e^{-\lambda h} u(i, j) \rightarrow \lambda u(i, j)$$

cuando $h \rightarrow 0$.

Es decir, se abandona i con una tasa λ y se va a j con probabilidad $u(i, j)$.

En la mayoría de los casos se describe el sistema determinando previamente las tasas de transición $q(i, j)$ para $j \neq i$, que determina el ritmo al que se salta de i a j .

Ejemplo. En un proceso de Poisson, se tiene el número de llegadas $X(t)$ que hay hasta el tiempo t con una tasa λ . Entonces el número de llegadas se incrementa de n a $(n + 1)$ a un ritmo de λ , es decir,

$$q(n, n + 1) = \lambda$$

para todo $n \geq 0$.

Ejemplo. Cola $M/M/s$: Supongamos un banco con s ventanillas que atienden a clientes que forman una sola cola si todas las ventanillas están ocupadas. Supongamos que los clientes aparecen según un proceso de Poisson de tasa λ y que cada tiempo de servicio se distribuye según una exponencial de tasa μ . Como en el ejemplo anterior, $q(n, n + 1) = \lambda$, y las salidas del banco siguen este esquema:

$$q(n, n - 1) = \begin{cases} n\mu & 0 \leq n < s \\ s\mu & n \geq s \end{cases}$$

Se puede observar que cuando hay $n \leq s$ personas en el banco, todas son atendidas y las salidas de los clientes ocurren a una tasa de $n\mu$. Cuando $n > s$ todas las ventanillas están ocupadas y las salidas se producen con tasa $s\mu$.

Construcción de la cadena, conocidas las tasas

Denotamos

$$\lambda_i = \sum_{j \neq i} q(i, j)$$

la tasa a la que X_t abandona el estado i . Si $\lambda_i = 0$ significaría que nunca abandona dicho estado, luego asumimos que $\lambda_i > 0$ y se puede calcular la probabilidad de que la cadena vaya a j cuando deja i como,

$$r(i, j) = \frac{q(i, j)}{\lambda_i}$$

Es decir, X_t permanece en i durante un tiempo distribuido según una exponencial de tasa λ_i y luego va al estado j con probabilidad $r(i, j)$.

Ejemplo. *Proceso de Yule:* Este es el modelo más simple de crecimiento de una población (e.g. bacterias) donde no hay muertes y cada individuo se subdivide con tasa β , de modo que

$$q(i, i + 1) = \beta i$$

y el resto de valores $q(i, j) = 0$.

Si se empieza con un solo individuo entonces el salto a $n + 1$ individuos se hace en el tiempo $T_n = t_1 + \dots + t_n$, donde t_n se distribuye como una exponencial con parámetro βn .

$E(t_i) = \frac{1}{\beta i}$, de modo que

$$E(T_n) = \frac{1}{\beta} \sum_{m=1}^n \frac{1}{m} \approx \frac{\log n}{\beta}$$

cuando $n \rightarrow \infty$. Se puede demostrar que $T_n \rightarrow \infty$.

Cálculo de la probabilidad de transición

Se trata de calcular la probabilidad de transición p_t a partir de las tasas de salto q . Para ello se usa la ecuación de Chapman-Kolmogorov considerando $k = i$ y calculando el siguiente límite

$$\lim_{h \rightarrow 0} \frac{p_{t+h}(i, j) - p_t(i, j)}{h} = p'_t(i, j)$$

donde

$$p_{t+h}(i, j) = \sum_k p_h(i, k) p_t(k, j).$$

Por definición de las tasas de salto

$$q(i, j) = \lim_{h \rightarrow 0} \frac{p_h(i, j)}{h}$$

para $i \neq j$, y operando, se obtiene la siguiente relación:

$$p'_t(i, j) = \sum_{k \neq i} q(i, k) p_t(k, j) - \lambda_i p_t(i, j). \quad (1)$$

Para simplificar esta expresión se define la matriz

$$Q(i, j) = \begin{cases} q(i, j) & \text{Si } j \neq i \\ -\lambda_i & \text{Si } j = i \end{cases}$$

A la matriz Q se le denomina, también, *generador infinitesimal* de la cadena. La suma de las filas de Q es 0, ya que $\lambda_i = \sum_{j \neq i} q(i, j)$. Los elementos fuera de la diagonal son no negativos y los de la diagonal son no positivos.

La ecuación (1) se puede escribir como

$$p'_t = Q p_t$$

Esto es una ecuación diferencial matricial cuya solución es

$$\begin{aligned} p_t &= e^{Qt} = \sum_{n=0}^{\infty} \frac{(Qt)^n}{n!} = \\ &= \sum_{n=0}^{\infty} Q^n \frac{t^n}{n!}. \end{aligned}$$

Se puede obtener también que

$$p'_t = p_t Q$$

es decir: $p_t Q = Q p_t$, lo cual no resulta siempre cierto al multiplicar matrices, y la razón de este hecho se encuentra en que $p_t = e^{Qt}$ se construye a partir de potencias de la matriz Q :

$$Q p_t = Q \exp(Qt) = \sum_{n=0}^{\infty} Q \frac{(Qt)^n}{n!} = \sum_{n=0}^{\infty} \frac{(Qt)^n}{n!} Q = \exp(Qt) \cdot Q = p_t Q$$

Ejemplo. *Proceso de Yule:* En este modelo cada individuo se dividía en dos con tasa β , de modo que $q(i, i + 1) = \beta i$.

Se demuestra que la probabilidad de transición es

$$p_t(1, j) = e^{-\beta t} (1 - e^{-\beta t})^{j-1}$$

para $j \geq 1$. Esto es la función de probabilidad de una distribución geométrica con probabilidad de éxito igual a $e^{-\beta t}$.

Se puede obtener del mismo modo $p_t(i, j)$, porque la cadena que empieza con i individuos es la suma de i copias de la cadena empezando en 1 individuo. De este modo se obtiene que

$$p_t(i, j) = \binom{j-1}{i-1} (e^{-\beta t})^i (1 - e^{-\beta t})^{j-i}$$

es decir, de la suma de i distribuciones geométricas resulta una binomial negativa.

De todas formas, no siempre se puede resolver explícitamente las ecuaciones de Kolmogorov para calcular las probabilidades de transición.

Ejemplo. Supongamos una cadena con dos estados : 0 y 1. Asumimos que $q(0, 1) = 1$ y que $q(1, 0) = 2$, entonces

$$Q = \begin{pmatrix} -1 & 1 \\ 2 & -2 \end{pmatrix}$$

ya que la suma es 0 por filas, por construcción de Q .

Para calcular e^{Qt} se diagonaliza la matriz, calculando previamente los autovectores y autovalores de Q .

Los autovalores son 0 y -3 y los autovectores forman la matriz

$$C = \begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix}.$$

Se tiene que

$$C^{-1} = \begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix}^{-1} = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & -\frac{1}{3} \end{pmatrix}$$

de este modo,

$$Q = C \cdot D \cdot C^{-1} = \begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & -3 \end{pmatrix} \begin{pmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & -\frac{1}{3} \end{pmatrix}$$

Se puede calcular, entonces,

$$\begin{aligned} p_t &= e^{Qt} = \sum_{n=0}^{\infty} Q^n \frac{t^n}{n!} = \sum_{n=0}^{\infty} \frac{C (Dt)^n C^{-1}}{n!} = \\ &= C \begin{pmatrix} 1 & 0 \\ 0 & e^{-3t} \end{pmatrix} C^{-1} \end{aligned}$$

ya que

$$\begin{aligned} \sum_{n=0}^{\infty} \frac{(Dt)^n}{n!} &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & \frac{(-3t)^1}{1!} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & \frac{(-3t)^2}{2!} \end{pmatrix} + \dots = \\ &= \begin{pmatrix} 1 & 0 \\ 0 & e^{-3t} \end{pmatrix} \end{aligned}$$

De este modo,

$$\begin{aligned} p_t &= C \begin{pmatrix} 1 & 0 \\ 0 & e^{-3t} \end{pmatrix} C^{-1} = \begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & e^{-3t} \end{pmatrix} \begin{pmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & -\frac{1}{3} \end{pmatrix} = \\ &= \begin{pmatrix} \frac{2}{3} + \frac{1}{3}e^{-3t} & \frac{1}{3} - \frac{1}{3}e^{-3t} \\ \frac{2}{3} - \frac{2}{3}e^{-3t} & \frac{1}{3} + \frac{2}{3}e^{-3t} \end{pmatrix} = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{2}{3} & \frac{1}{3} \end{pmatrix} + \begin{pmatrix} \frac{1}{3}e^{-3t} & \frac{1}{3}e^{-3t} \\ \frac{2}{3}e^{-3t} & \frac{2}{3}e^{-3t} \end{pmatrix}. \end{aligned}$$

Se puede observar que cuando $t \rightarrow \infty$ entonces

$$\lim_{t \rightarrow \infty} p_t = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{2}{3} & \frac{1}{3} \end{pmatrix} = \begin{pmatrix} \bar{\pi} \\ \bar{\pi} \end{pmatrix}$$

donde $\bar{\pi} = \left(\frac{2}{3} \quad \frac{1}{3} \right)$.

Probabilidades límite

Al igual que en el caso discreto interesa estudiar el comportamiento a largo plazo de la cadena. La teoría es equivalente a los procesos discretos: se supone que la cadena es irreducible, así, todos los estados comunican, es decir para todo i, j se tiene que existen i_1, i_2, \dots, i_j tal que $q(i, i_1), q(i_1, i_2), \dots, q(i_j, j)$ son todos estrictamente positivos. Así, se obtiene que si una cadena de Markov X_t es irreducible y tiene distribución estacionaria π , entonces

$$\lim_{t \rightarrow \infty} p_t(i, j) = \pi(j).$$

En las cadenas de Markov en tiempo discreto, la distribución estacionaria es solución de la ecuación $\pi p_t = \pi$. Como en tiempo continuo no existe un primer $t > 0$, se necesita

una condición más fuerte: se dice que π es una distribución estacionaria si $\pi p_t = \pi$ para todo $t > 0$.

Esta condición es complicada de comprobar porque implica conocer todas las p_t que no son, a menudo fáciles de calcular. Por ello, se hace en términos de la matriz Q , esto es,

$$Q(i, j) = \begin{cases} q(i, j) & \text{Si } j \neq i \\ -\lambda_i & \text{Si } j = i \end{cases}$$

donde $\lambda_i = \sum_{j \neq i} q(i, j)$ es la tasa total de transiciones a partir de i .

Se tiene que π es una distribución estacionaria si sólo si $\pi Q = 0$.

Esto se puede ver de manera intuitiva, sustituyendo en la condición anterior el valor de Q , de modo que la condición $\pi Q = 0$ queda como

$$\sum_{k \neq j} \pi(k)q(k, j) = \pi(j)\lambda_j$$

Si usáramos el símil de que $\pi(k)$ representa la cantidad de agua que hay en k , el término de la derecha representa la tasa a la que el agua deja j , mientras que la de la izquierda representa la tasa a la que el agua llega a j . Así, π será la distribución estacionaria si para cada j el flujo de agua que entra es igual al flujo de agua que sale. Se puede ver una demostración detallada en Durrett (1999) pág. 170.

Otra manera de verlo es que la probabilidad límite no debe cambiar con el tiempo así como

$$p'_t = p_t Q$$

entonces

$$\pi Q = 0,$$

es decir, la derivada es igual a 0 y p_t es constante, y en este caso, π es un autovector de Q con autovalor 0.

Proposición. Si se verifica que

$$\pi(k)q(k, j) = \pi(j)q(j, k) \tag{2}$$

para todo j, k , entonces π es una distribución estacionaria.

Se demuestra sumando sobre todo $k \neq j$ y reordenando el resultado.

Procesos de nacimiento y muerte.

En este caso, el espacio de estados es $S = \{0, 1, 2, \dots, N\}$ y los cambios en los estados siempre son desde n hasta $n + 1$, ó desde n hasta $n - 1$. De manera intuitiva se puede considerar el estado del sistema como el tamaño de la población que se puede incrementar en 1 por un nacimiento, o decrecer en 1 por una muerte.

Para describir la cadena, asumimos que las tasas de nacimiento son λ_n , $n = 0, 1, 2, \dots, N - 1$ y las tasas de muerte son μ_n , $n = 1, 2, 3, \dots, N$

Si la población es n , entonces los nuevos individuos aparecen con tasa λ_n y los individuos dejan la población con tasa μ_n (de modo que si la población es 0 entonces no puede haber muertes, de modo que $\mu_0 = 0$). Así,

$$\begin{aligned} q(n, n + 1) &= \lambda_n && \text{para } 0 \leq n < N \\ q(n, n - 1) &= \mu_n && \text{para } 0 < n \leq N \end{aligned}$$

Aplicamos la condición

$$\pi(k)q(k, j) = \pi(j)q(j, k)$$

entonces

$$\begin{aligned} \pi(n)q(n, n - 1) &= \pi(n - 1)q(n - 1, n) \Rightarrow \\ \pi(n)\mu_n &= \pi(n - 1)\lambda_{n-1} \Rightarrow \\ \pi(n) &= \pi(n - 1) \frac{\lambda_{n-1}}{\mu_n}. \end{aligned}$$

Si se vuelve a usar esta ecuación de la misma manera, se tiene que

$$\pi(n - 1) = \pi(n - 2) \frac{\lambda_{n-2}}{\mu_{n-1}}$$

esto es,

$$\pi(n) = \frac{\lambda_{n-1}}{\mu_n} \frac{\lambda_{n-2}}{\mu_{n-1}} \pi(n - 2)$$

e iterando de la misma manera, se obtiene que

$$\pi(n) = \frac{\lambda_{n-1} \cdot \lambda_{n-2} \cdots \lambda_0}{\mu_n \cdot \mu_{n-1} \cdots \mu_1} \pi(0).$$

Se puede observar que si el espacio de estados es $S = \{0, 1, 2, \dots, N\}$, entonces $\mu_0 = 0$ y $\lambda_n = 0$, de modo que esos términos no pueden aparecer en la fórmula anterior.

Ejemplo. Supongamos una peluquería que corta el pelo a un ritmo o tasa de 3 personas por hora, esto es, cada corte de pelo requiere una cantidad de tiempo que se distribuye como una exponencial de media 20 minutos. Supongamos que los clientes llegan según un proceso de Poisson de tasa 2, pero que se marchan si las dos sillas de la salita de espera están ocupadas. Se trata de calcular la fracción de tiempo que las dos sillas están ocupadas. También, interesa averiguar cuantos clientes son atendidos por hora, a largo plazo.

Se define el estado como el número de clientes en el sistema, de modo que $S = \{0, 1, 2, 3\}$. A partir del enunciado se deduce que

$$\begin{aligned} \mu_i &= q(i, i-1) = 3 && \text{para } i = 1, 2, 3 \\ \lambda_i &= q(i, i+1) = 2 && \text{para } i = 0, 1, 2 \end{aligned}$$

Aplicamos la condición (2) y se obtiene

$$\lambda_{n-1}\pi(n-1) = \mu_n\pi(n) \Rightarrow$$

$$2\pi(0) = 3\pi(1)$$

$$2\pi(1) = 3\pi(2)$$

$$2\pi(2) = 3\pi(3)$$

Fijamos, por ejemplo, $\pi(0) = c$ y resolviendo el sistema anterior, se obtiene

$$\begin{aligned} \pi(1) &= \frac{2c}{3}, \\ \pi(2) &= \frac{2}{3}\pi(1) = \frac{2}{3} \cdot \frac{2c}{3} = \frac{4c}{9} \\ \pi(3) &= \frac{2}{3}\pi(2) = \frac{2}{3} \cdot \frac{4c}{9} = \frac{8c}{27} \end{aligned}$$

La suma de todas ellas debe ser igual a 1, luego

$$\begin{aligned}\pi(0) + \pi(1) + \pi(2) + \pi(3) &= 1 \Rightarrow \\ c + \frac{2c}{3} + \frac{4c}{9} + \frac{8c}{27} &= 1 \Rightarrow c = \frac{27}{65}\end{aligned}$$

luego

$$\begin{aligned}\pi(0) &= \frac{27}{65} \\ \pi(1) &= \frac{18}{65} \\ \pi(2) &= \frac{12}{65} \\ \pi(3) &= \frac{8}{65}\end{aligned}$$

Esto significa que un $\frac{8}{65}$ del tiempo del tiempo ambas cadenas están llenas de modo que se pierde esa fracción de clientes, y así un $\frac{57}{65}$, es decir un 87.7% de los clientes es atendido. Como la tasa de llegada original es de 2 personas entonces se corta el pelo a una media de $2 \cdot \frac{57}{65} = \frac{114}{65} = 1,754$ personas por hora.

Ejemplo. Supongamos una fábrica con tres máquinas en uso y un técnico para repararlas. Cada máquina trabaja durante un tiempo exponencial con media 60 días antes de una avería, pero cada avería requiere un tiempo de reparación exponencial con media 4 días. Se trata de calcular la fracción de tiempo en que las tres máquinas trabajan a largo plazo.

Denominamos X_t el número de máquinas que está funcionando, dado que sólo hay un técnico para repararlas se tiene que

$$q(i, i + 1) = \frac{1}{4}$$

donde $i = 0, 1, 2$.

Por otro lado, la tasa de fallo es proporcional al número de máquinas trabajando, de modo que

$$q(i, i - 1) = \frac{i}{60}$$

donde $i = 1, 2, 3$.

Fijamos $\pi(0) = c$ de modo que, como

$$\lambda_{n-1}\pi(n-1) = \mu_n\pi(n),$$

$$\begin{aligned}\lambda_0\pi(0) &= \mu_1\pi(1) \implies \\ \pi(1) &= \frac{\lambda_0}{\mu_1}\pi(0) = \frac{\frac{1}{4}}{\frac{1}{60}} \cdot c = 15c\end{aligned}$$

$$\begin{aligned}\lambda_1\pi(1) &= \mu_2\pi(2) \implies \\ \pi(2) &= \frac{\lambda_1}{\mu_2}\pi(1) = \frac{\frac{1}{4}}{\frac{2}{60}} \cdot 15c = \frac{225c}{2}\end{aligned}$$

$$\begin{aligned}\lambda_2\pi(2) &= \mu_3\pi(3) \implies \\ \pi(3) &= \frac{\lambda_2}{\mu_3}\pi(2) = \frac{\frac{1}{4}}{\frac{3}{60}} \cdot \frac{225c}{2} = \frac{1225c}{2}\end{aligned}$$

La suma de todas ellas debe ser igual a 1, luego

$$\begin{aligned}\pi(0) + \pi(1) + \pi(2) + \pi(3) &= 1 \implies \\ c + 15c + \frac{225c}{2} + \frac{1225c}{2} &= 1 \implies c = \frac{1}{741}\end{aligned}$$

luego

$$\begin{aligned}\pi(0) &= \frac{1}{741} \\ \pi(1) &= \frac{15}{741} \\ \pi(2) &= \frac{75}{494} \\ \pi(3) &= \frac{1225}{1482}\end{aligned}$$

luego, a largo plazo, el porcentaje de las máquinas que están trabajando es el $\frac{1225}{1482} \cdot 100 = 82,66\%$ del tiempo.

Martingalas

Nociones sobre esperanzas condicionadas.

Supongamos que Y es una *v.a.*, si no se sabe nada acerca de la distribución de sus valores la mejor predicción es utilizar su esperanza: $E(Y)$. La noción de esperanza condicionada trata el problema de predecir los valores de una variable aleatoria sin que se tenga información completa sobre la misma.

Supongamos que X e Y son dos *v.a.* discretas cuya función de probabilidad conjunta es

$$f(x, y) = P(X = x, Y = y)$$

y las funciones de probabilidad marginales son

$$\begin{aligned} f_X(x) &= \sum_y f(x, y) \\ f_Y(y) &= \sum_x f(x, y) \end{aligned}$$

Si conocemos un valor de $X = x$ se trata de dar el mejor valor posible de Y , esto es, $E(Y | X = x)$ que se puede expresar como

$$\begin{aligned} E(Y | X = x) &= E(Y | X)(x) = \sum_y yP(Y = y | X = x) = \\ &= \sum_y y \frac{P(X = x, Y = y)}{P(X = x)} = \\ &= \frac{\sum_y yf(x, y)}{f_X(x)} \end{aligned}$$

y esto está bien definido si $f_X(x) > 0$.

Esto se puede generalizar de manera inmediata a varias variables: Si X_1, X_2, \dots, X_n, Y son variables aleatorias discretas con función de probabilidad conjunta

$$f(x_1, x_2, \dots, x_n, y) = P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n, Y = y)$$

y la marginal

$$g(x_1, x_2, \dots, x_n) = \sum_y f(x_1, x_2, \dots, x_n, y)$$

de modo que la esperanza condicionada de Y dados X_1, X_2, \dots, X_n son

$$E(Y | X_1, X_2, \dots, X_n)(x_1, x_2, \dots, x_n) = \frac{\sum_y y f(x_1, x_2, \dots, x_n, y)}{g(x_1, x_2, \dots, x_n)}$$

Se puede considerar que $E(Y | X_1, X_2, \dots, X_n)$ es la mejor predicción de del valor de Y dado que se conocen los valores de X_1, X_2, \dots, X_n . En realidad, esto es lo que se busca en los métodos de Regresión, considerando, generalmente funciones lineales.

Todo lo anterior se puede definir, igualmente, para *v.a.* continuas sustituyendo sumatorios por integrales:

$$E(Y | X_1, X_2, \dots, X_n)(x_1, x_2, \dots, x_n) = \frac{\int_{-\infty}^{\infty} y f(x_1, x_2, \dots, x_n, y) dy}{f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)}$$

La esperanza condicionada tiene dos propiedades características:

(i) El valor de la variable $E(Y | X_1, X_2, \dots, X_n)$ depende sólo de los valores de X_1, X_2, \dots, X_n , es decir podemos escribir esta esperanza como una función exclusivamente de ellos:

$$E(Y | X_1, X_2, \dots, X_n) = \phi(X_1, X_2, \dots, X_n)$$

para alguna función dada ϕ . Si una *v.a.* Z se puede expresar como una función de una serie de *v.a.* X_1, X_2, \dots, X_n se dice que es *medible* respecto de X_1, X_2, \dots, X_n .

(ii) Supongamos que A es un suceso que depende sólo de X_1, X_2, \dots, X_n . Por ejemplo, A podría ser

$$A = \{a_1 \leq X_1 \leq b_1, \dots, a_n \leq X_n \leq b_n\}.$$

Denotamos por I_A la función indicatriz del suceso A (es decir, es una *v.a.* tal que toma el valor 1 si A aparece y 0 en caso contrario), entonces se verifica que

$$E(Y \cdot I_A) = E(E(Y | X_1, X_2, \dots, X_n) \cdot I_A)$$

La notación se puede hacer más compacta de la manera siguiente: Si se tiene una sucesión de *v.a.* X_1, X_2, \dots, X_n se escribe \mathcal{F}_n como la información contenida por las variables anteriores. De este modo se pone

$$E(Y | X_1, X_2, \dots, X_n) = E(Y | \mathcal{F}_n)$$

Algunas propiedades.

1. Si el suceso A es toda la muestra entera se tiene, entonces, que

$$E(E(Y | \mathcal{F}_n)) = E(Y)$$

2. Si a y b son constantes, entonces

$$E(aY_1 + bY_2 | \mathcal{F}_n) = aE(Y_1 | \mathcal{F}_n) + bE(Y_2 | \mathcal{F}_n)$$

3. Si Y es ya una función de X_1, X_2, \dots, X_n entonces

$$E(Y | \mathcal{F}_n) = Y$$

4. Para cualquier Y , si $m < n$, entonces

$$E(E(Y | \mathcal{F}_n) | \mathcal{F}_m) = E(Y | \mathcal{F}_m)$$

5. Si Y es independiente de X_1, X_2, \dots, X_n entonces la información proporcionada por las anteriores variables no es relevante, de modo que

$$E(Y | \mathcal{F}_n) = E(Y)$$

6. Si Y es una *v.a.* cualquiera y Z es una *v.a.* que es medible respecto de X_1, X_2, \dots, X_n entonces

$$E(Y \cdot Z | \mathcal{F}_n) = Z \cdot E(Y | \mathcal{F}_n)$$

Definición de martingala

Una martingala es un modelo para un juego equilibrado. Sea X_0, X_1, \dots una sucesión de *v.a.* y se denota por \mathcal{F}_n a la información contenida en X_0, X_1, \dots, X_n . Se dice que una sucesión de *v.a.* M_0, M_1, M_2, \dots tal que $E(|M_i|) < \infty$ es una *martingala* con respecto a \mathcal{F}_n si

(i) Cada M_n es medible con respecto a X_0, X_1, \dots, X_n

(ii) Para todo $m < n$

$$E(M_n | \mathcal{F}_m) = M_m$$

o de modo equivalente

$$E(M_n - M_m | \mathcal{F}_m) = 0$$

La condición $E(|M_i|) < \infty$ se establece para que estén bien definidas las esperanzas condicionadas.

Para comprobar la condición (ii) basta probar que

$$E(M_{n+1} | \mathcal{F}_n) = M_n$$

ya que si esto se cumple, entonces aplicando la propiedad 4

$$\begin{aligned} E(M_{n+2} | \mathcal{F}_n) &= E(E(M_{n+2} | \mathcal{F}_{n+1}) | \mathcal{F}_n) = \\ &= E(M_{n+1} | \mathcal{F}_n) = M_n \end{aligned}$$

y así sucesivamente.

Ejemplo. Sean X_1, X_2, \dots variables aleatorias independientes cada una con media μ . Definimos $S_0 = 0$, y para $n > 0$ se definen

$$S_n = X_1 + \dots + X_n$$

entonces $M_n = S_n - n\mu$ es una martingala con respecto a \mathcal{F}_n , la información contenida en las variables X_0, X_1, \dots, X_n :

$$\begin{aligned} E(M_{n+1} | \mathcal{F}_n) &= E(S_{n+1} - (n+1)\mu | \mathcal{F}_n) = \\ &= E(S_{n+1} | \mathcal{F}_n) - (n+1)\mu \stackrel{(*)}{=} (S_n + \mu) - (n+1)\mu = \\ &= S_n - n\mu = M_n \end{aligned}$$

(*) ya que

$$\begin{aligned} E(S_{n+1} | \mathcal{F}_n) &= E(S_n | \mathcal{F}_n) + E(X_{n+1} | \mathcal{F}_n) = \\ &= S_n + E(X_{n+1}) = S_n + \mu \end{aligned}$$

porque $S_n = X_1 + \dots + X_n$ es medible con respecto a \mathcal{F}_n por la condición (i) de la definición de martingalas, y X_{n+1} es independiente respecto de X_1, \dots, X_n .

Por otro lado, si $\mu = 0$, S_n es una martingala con respecto a \mathcal{F}_n .

Ejemplo (martingala de Casanova). Se trata de definir la estrategia de ganancia en un juego equilibrado por medio de una martingala. El problema se lo planteó Giacomo Casanova en 1754, durante el carnaval de Venecia.

Supongamos que X_1, X_2, \dots son variables aleatorias independientes cada una con

$$P(X_i = 1) = P(X_i = -1) = \frac{1}{2}$$

Se puede considerar que las variables X_i son el resultado de un juego donde se pierde 1 unidad monetaria o se gana con la misma probabilidad. Una estrategia posible es la de ir doblando las apuestas hasta que se gane. En ese momento se para.

Denotamos por W_n las ganancias o pérdidas después de n juegos siguiendo esta estrategia (así $W_0 = 0$). Siempre que se gane termina el juego y así, también, las pérdidas o ganancias, y éstas dejan de ir cambiando:

$$P(W_{n+1} = 1 \mid W_n = 1) = 1.$$

Supongamos, ahora, que en las primeras n tiradas se ha perdido. Después de cada tirada se ha doblado la apuesta, de modo que se han perdido las siguientes unidades monetarias:

$$1 + 2 + 4 + \dots + 2^{n-1} = 2^n - 1,$$

es decir, $W_n = -(2^n - 1)$. En este momento se doblaría la apuesta de nuevo apostando 2^n unidades en el siguiente juego. De este modo,

$$\begin{aligned} P(W_{n+1} = 1 \mid W_n = -(2^n - 1)) &= \frac{1}{2} \\ P(W_{n+1} = -(2^{n+1} - 1) \mid W_n = -(2^n - 1)) &= \frac{1}{2} \end{aligned}$$

Por ello,

$$E(W_{n+1} | \mathcal{F}_n) = 1 \cdot \frac{1}{2} - (2^{n+1} - 1) \cdot \frac{1}{2} = 1 - 2^n = -(2^n - 1) = W_n$$

y, así, W_n es una martingala con respecto a \mathcal{F}_n .

Movimiento Browniano

Introducción

En 1827 el botánico R. Brown observó el comportamiento aleatorio de los granos de polen en una suspensión de agua; a este fenómeno lo denominó movimiento Browniano. 80 años más tarde A. Einstein desarrolló sus propiedades matemáticas de manera independiente de Brown.

Los procesos Brownianos reúnen tres procesos en uno: un proceso gaussiano, un proceso de Markov continuo y un proceso de incrementos independientes. Por ello, se pueden usar las técnicas de los tres procesos anteriores en su estudio. Es el caso más importante de los procesos en tiempo continuo y trayectorias continuas.

Se puede definir de la siguiente manera:

Definición

$B(t)$ es un movimiento Browniano unidimensional con varianza σ^2 si cumple las siguientes condiciones:

(i) $B(0) = 0$

(ii) Tiene incrementos independientes: para $0 = t_0 < t_1 < \dots < t_n$

$$B(t_1) - B(t_0), \dots, B(t_n) - B(t_{n-1})$$

son independientes.

(iii) Tiene incrementos estacionarios: la distribución de $B_t - B_s$ sólo depende de $t - s$.

(iv) $B(t)$ se distribuye como una normal $N(0, \sigma^2 t)$.

(v) $B(t)$ es continua con respecto a t .

Intuitivamente se puede imaginar que los granos de polen se arremolinan entre sí de manera continua, y que colisionan con las moléculas de agua según unos tiempos determinados por un proceso de Poisson con tasa λ . El desplazamiento de un grano de polen en una unidad de tiempo es la suma de un gran número de pequeños efectos aleatorios independientes. Se puede aplicar, entonces, el Teorema Central del Límite de modo que el cambio en el desplazamiento se aproxima por una normal de media 0 y varianza σ^2 .

Nota:

Si X_1, \dots, X_n son v.a. independientes con media μ y varianza común $\sigma^2 < \infty$, la v.a. Z definida como

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

es una v.a. cuya función de densidad se aproxima a la distribución normal cuando n es grande:

$$Z \sim N(0, 1)$$

esto es, para n grande

$$\frac{X_1 + \dots + X_n}{n} = \bar{X} \simeq N\left(\mu, \frac{\sigma}{\sqrt{n}}\right).$$

Como se supone que hay un gran número de moléculas ($> 10^{20}$), es razonable pensar que después de la colisión, no se vuelve a observar dicha partícula, lo cual hace que $X(t)$ tenga incrementos independientes.

El movimiento de las partículas es continuo y si se toma como punto de partida el origen, $X(0) = 0$, entonces el desplazamiento horizontal sigue un movimiento Browniano con varianza igual a σ^2 .

El movimiento Browniano como límite de un paseo aleatorio

El movimiento Browniano es, en realidad, una versión continua de un camino aleatorio.

Supongamos un camino aleatorio simétrico

$$S_n = S_{n-1} + X_n$$

$$S_0 = a$$

donde las v.a X_i son independientes con distribución

$$P(X_i = \pm 1) = \frac{1}{2},$$

de modo que, como ya se vió, $E(X_i) = 0$ y $Var(X_i) = 1$.

Si consideramos como unidad de tiempo 1, y la subdividimos en N periodos de longitud Δt , entonces

$$N\Delta t = 1$$

Se puede considerar

$$S_{n\Delta t}^N = S_{(n-1)\Delta t}^N + \sqrt{\Delta t}X_n$$

$$S_0^N = a$$

donde el superíndice N se pone para indicar la dependencia en el salto temporal $\Delta t = \frac{1}{N}$.

Así el tiempo va dando saltos de Δt en Δt , aunque los incrementos del espacio lo hacen de $\sqrt{\Delta t}$ en $\sqrt{\Delta t}$, ya que

$$\begin{aligned} Var(S_{tiempo1}^N) &= Var(S_{N\Delta t}^N) = \sum_{n=1}^N Var(S_{n\Delta t}^N - S_{(n-1)\Delta t}^N) = \\ &= N \cdot Var(\sqrt{\Delta t}X_i) = N \cdot \Delta t = 1. \end{aligned}$$

Si se hace $\Delta t \rightarrow 0$, se puede aplicar el Teorema Central del Límite, entonces para $n > m$

$$S_{n\Delta t}^N - S_{m\Delta t}^N \sim N(0; \sigma^2 = (n - m)\Delta t)$$

donde $(n - m)\Delta t$ es el tiempo transcurrido. Así, se obtiene un movimiento Browniano. Se puede demostrar que un camino aleatorio converge (en ley) a un movimiento Browniano, y aplicar, así, sus propiedades.

Algunas propiedades

Supongamos que $B(t)$ es un movimiento Browniano que comienza en $B(0) = a$, entonces

(i) $B^*(t) = b + B(t)$ es un movimiento Browniano que comienza en $a + b$.

(ii) $B^*(t) = -B(t)$ es un movimiento Browniano que comienza en $-a$.

(iii) $B^*(t) = \frac{1}{\sqrt{c}}B(ct)$ es un movimiento Browniano que comienza en $\frac{a}{\sqrt{c}}$

(iv) $B^*(t) = B(t + h) - B(h)$, para un valor de h fijado, es un movimiento Browniano que comienza en 0.

(v) Aunque sean continuas las trayectorias de un movimiento Browniano, no son derivables, esto es, son trayectorias *no suaves*.

Ejemplo de aplicación en cálculo financiero

Una posible situación financiera sencilla es aquella en que se analiza un activo con un periodo fijo de tiempo (de 0 a T sin negociación en tiempos intermedios) y una rentabilidad segura R efectiva para ese periodo de tiempo, entonces, si se denota como S_t el valor del activo en el tiempo t se tiene que

$$\begin{aligned}\frac{S_T - S_0}{S_0} &= R \implies \\ S_T &= (1 + R)S_0.\end{aligned}$$

En la realidad, sin embargo, la rentabilidad es estocástica (no determinística) y varía de manera continua y constante (la negociación es continua y su ley cambia con el tiempo).

Supongamos:

1. El tiempo se considera continuo y la rentabilidad se supone determinística y variable.

Entonces, la ecuación anterior toma la forma de la ecuación diferencial

$$dS(t) = S(t)R(t)dt$$

y si se añade la condición inicial $S(0) = S_0$, entonces

$$S(t) = S_0 \exp \left[\int_0^t R(v) dv \right]$$

2. El tiempo se considera discreto y la rentabilidad se supone estocástica y markoviana, R_n , de modo que sólo depende del tiempo n y del nivel S_n . En este caso, ΔS_n se puede modelizar como una cadena de Markov.

3. El tiempo se considera continuo y la rentabilidad se supone estocástica. Entonces, si X_t es el nivel de un activo en el tiempo t , el salto se puede denotar como dX_t que es una *v.a.* cuya distribución depende de X_t y de t .

La situación más simple e importante es aquella en la que dX_t es una *v. a.* normal

$$dX_t \sim N(m, v)$$

donde

$$\begin{aligned} m &= a(t, X_t)dt \\ v &= \sigma^2(t, X_t)dt \end{aligned}$$

donde, de manera implícita, se supone que la ley del salto dX_t sólo depende de t y de X_t , lo cual implica que se trata de un proceso markoviano.

En análisis financiero se modelizan las variables de forma continua y estocástica, de modo que la ecuación general es

$$dX_t = a(t, X_t)dt + \sigma(t, X_t)dB_t$$

donde B_t es un movimiento Browniano.

Esta ecuación entra ya en el capítulo de integrales estocásticas, o integrales de Itô, y su resolución queda fuera del alcance del curso.