

Capítulo 7

Monte Carlo vía Cadenas de Markov

En este capítulo presentamos una herramienta de simulación genérica, conocida como Monte Carlo vía cadenas de Markov (MCMC), la cuál permite generar muestras (aproximadas) de cualquier distribución. Las técnicas de Monte Carlo vía cadenas de Markov permiten generar, de manera iterativa, observaciones de distribuciones multivariadas que difícilmente podrían simularse utilizando métodos como los estudiados hasta ahora. La idea básica es construir una cadena de Markov que sea fácil de simular y cuya distribución estacionaria corresponda a la distribución objetivo que nos interesa.

Las ideas de esta técnica han tenido su desarrollo histórico desde diferentes ramas de las matemáticas: de la probabilidad, de la mecánica estadística y de la estadística bayesiana. Cada enfoque tiene su utilidad conceptual y de implementación, y se enriquecen entre ellos. Es por ello que presentaremos una introducción a estos diferentes enfoques en el capítulo.

Antes de presentar los algoritmos, presentaremos el resultado que motiva y fundamenta matemáticamente el uso de los algoritmos MCMC.

Tenemos el siguiente resultado fundamental para Cadenas de Markov.

Proposición 7.1. *Sea $\{X_n\}_{n \geq 0}$ una cadena de Markov homogénea, irreducible y aperiódica con espacio de estados E y distribución estacionaria π , se tiene que cuando $n \rightarrow \infty$ ocurre que*

$$X_n \xrightarrow{d} X$$

donde X tiene distribución π y

$$\frac{1}{n} \sum_{i=1}^n g(X_n) \xrightarrow{c.s.} E(g(X)).$$

La primera afirmación del resultado es el teorema clásico de convergencia de cadenas de Markov. Sin embargo, la segunda afirmación requiere un estudio más detallado. En la teoría de sistemas dinámicos, se prueba que para ciertos sistemas ocurre que los promedios temporales son iguales a los promedios espaciales. Este resultado nos permitirá tomar promedios de valores temporales en una cadena de Markov que son claramente dependientes pero que al promediarlos se comportan como si fuesen independientes; de forma similar a la ley fuerte de los grandes números. Este resultado es conocido como el teorema ergódico para Cadenas de Markov.

7.1. Metropolis-Hastings

En esta sección estudiaremos el principal algoritmo de Monte Carlo vías cadenas de Markov. La idea de este algoritmo, como se mencionó anteriormente, es simular una cadena de Markov tal que su distribución estacionaria coincida con la distribución a simular.

Como motivación para el desarrollo este algoritmo, supongamos que queremos generar valores de una variable aleatoria X con soporte en $E = \{1, \dots, N\}$ y distribución $\pi = (\pi_1, \dots, \pi_N)$, donde $\pi_i = b_i/C$, $i \in E$, N es grande y C es la constante de normalización (difícil de calcular). Construiremos un método que solucione el problema.

Construimos una cadena de Markov $\{X_n\}_{n \geq 0}$ con espacio de estados E , cuya evolución depende de la matriz de transición $Q = \{q_{ij}\}$ de otra cadena de Markov irreducible. Dicha evolución está definida de la siguiente manera:

1. Cuando $X_n = i$, generamos una variable aleatoria Y tal que

$$P(Y = j) = q_{ij}$$

para todo $i, j \in E$.

2. Si $Y = j$, hacemos

$$X_{n+1} = \begin{cases} j & \text{con probabilidad } \alpha_{ij} \\ i & \text{con probabilidad } 1 - \alpha_{ij} \end{cases}$$

donde

$$\alpha_{ij} = \min \left\{ \frac{\pi_j q_{ji}}{\pi_i q_{ij}}, 1 \right\} = \min \left\{ \frac{b_j q_{ji}}{b_i q_{ij}}, 1 \right\}.$$

La evolución anterior se traduce en que el proceso $\{X_n\}_{n \geq 0}$ tiene como matriz de transición a $P = \{p_{ij}\}_{\{i, j \in E\}}$ donde

$$p_{ij} = \begin{cases} q_{ij} \alpha_{ij} & \text{si } i \neq j \\ q_{ii} + \sum_{k \neq i} q_{ik} (1 - \alpha_{ik}) = 1 - \sum_{k \neq i} q_{ik} \alpha_{ik} & \text{si } i = j. \end{cases}$$

A α_{ij} se le conoce como la **probabilidad de aceptación**.

El siguiente resultado prueba la eficacia de la construcción. Presentamos la prueba por ser simple.

Proposición 7.2. *La cadena de Markov $\{X_n\}_{n \geq 0}$ tiene como distribución estacionaria a π .*

Demostración. Demostraremos que la cadena de Markov con matriz de transición $P = \{p_{ij}\}_{i, j \in E}$ es reversible (con respecto al tiempo) y tiene como distribución estacionaria a π . Para ello basta verificar que se satisfacen las ecuaciones de balance local:

$$\pi_i p_{ij} = \pi_j p_{ji},$$

para toda $i, j \in E$. Esto es equivalente a que se cumpla:

$$\pi_i q_{ij} \alpha_{ij} = \pi_j q_{ji} \alpha_{ji}.$$

Para verificar esta última igualdad basta notar que si

$$\alpha_{ij} = \frac{\pi_j q_{ji}}{\pi_i q_{ij}}$$

entonces $\alpha_{ji} = 1$ y viceversa. □

Corolario 7.1. *Si la cadena de Markov $\{X_n\}_{n \geq 0}$ es irreducible y aperiódica, entonces π es la distribución límite.*

Este resultado es inmediato de los resultados de convergencia de cadenas de Markov. Las hipótesis de irreducibilidad y aperiodicidad se siguen directamente de que la matriz Q es irreducible y de la construcción de P .

Tenemos dos observaciones importantes:

- La elección de la cadena Q es arbitraria, siempre que satisfaga la condición de irreducibilidad.
- No fue necesario calcular a la constante C para definir a la cadena de Markov $\{X_n\}_{n \geq 0}$.

La idea anterior (y su prueba) es suficientemente robusta y se puede extender a cadenas de Markov con espacio de valores continuos (por ejemplo \mathbb{R} o \mathbb{R}^n) bajo algunas condiciones generales (que involucran la interpretación de irreducibilidad y convergencia de Cadenas de Markov en espacios de valores continuos). Por supuesto, hay muchos detalles técnicos involucrados, por lo cuál no abordaremos la teoría correspondiente y simplemente daremos como efectiva la construcción análoga de la cadena de Markov $\{X_n\}_{n \geq 0}$ con valores continuos.

A partir de ahora cuando escribamos a x y y como estados en E , estaremos pensando que x y y son números reales o vectores. Imaginemos que queremos generar una muestra de una función de densidad (que podría ser multidimensional) objetivo $f(x)$.

El algoritmo siguiente usa una construcción análoga a la de la cadena de Markov en espacio de estados discreto, y se denomina algoritmo **Metropolis-Hastings**.

Se define la evolución de la cadena a continuación, dado que la cadena se encuentra en el estado X_n :

Algoritmo 48 : Metropolis-Hastings

- 1: Generar $Y \sim q(X_n, y)$, es decir generamos Y con el kernel de transición a un valor y dado que estamos en el valor X_n (la función de densidad de y al siguiente paso dado el valor anterior).
- 2: Generar $U \sim U(0, 1)$ y hacemos

$$X_{n+1} = \begin{cases} Y & \text{si } U \leq \alpha(X_n, Y), \\ X_n & \text{en otro caso} \end{cases}$$

donde

$$\alpha(x, y) = \min \left\{ \frac{f(y)q(y, x)}{f(x)q(x, y)}, 1 \right\}.$$

De esta forma tenemos que, repitiendo los pasos del algoritmo anterior, obtenemos una sucesión de variables aleatorias dependientes X_1, X_2, \dots, X_n , donde X_m tiene aproximadamente la función de densidad

$f(x)$, para toda m suficientemente grande.

Observemos que en el fondo, este es un algoritmo dinámico de aceptación y rechazo. La eficiencia del algoritmo dependerá de la probabilidad de aceptación $\alpha(x, y)$.

Una cuestión importante en este algoritmo es la selección de matriz de transición Q . Existen diferentes alternativas para esta matriz, que dependen del problema a resolver. Las más usadas son:

- **Simetría:** Consiste en elegir q tal que $q(x, y) = q(y, x)$. En este caso tendríamos que la probabilidad de aceptación es:

$$\alpha(x, y) = \min \left\{ \frac{f(y)}{f(x)}, 1 \right\},$$

es decir, solamente depende de la función de densidad de la variable aleatoria a simular, entonces tenemos un método de aceptación-rechazo similar al método original de aceptación-rechazo.

- **Independencia:** Podemos elegir q tal que $q(x, y) = g(y)$, para alguna función de densidad g . Esto quiere decir que la transición no está dependiendo del estado actual, por lo que en cada paso de la evolución simplemente estamos sorteando un candidato (con densidad g) independiente. El candidato generado será aceptado con probabilidad

$$\alpha(x, y) = \min \left\{ \frac{f(y)g(x)}{f(x)g(y)}, 1 \right\}.$$

Este método es muy parecido al de aceptación-rechazo original por lo que es importante que la densidad propuesta g esté cercana a f . Una observación importante es que a diferencia del método de aceptación-rechazo original, éste produce muestras dependientes (por ejemplo si se rechaza una vez, tendremos dos variables consecutivas generadas con valores idénticos).

- **Caminata aleatoria:** Bajo esta elección, el candidato es obtenido con la siguiente expresión: $y = x + Z$, donde Z es una variable con densidad g simulable y simétrica (alrededor del origen). Así, la transición está dada por $q(x, y) = g(y - x)$, y la probabilidad de aceptación es:

$$\alpha(x, y) = \min \left\{ \frac{f(y)}{f(x)}, 1 \right\}.$$

Ahora veremos algunos ejemplos del uso del algoritmo para la simulación de vectores aleatorios dependientes.

Ejemplo 7.1.1. Sea (X, Y) un vector aleatorio cuya densidad es

$$f(x, y) = c * \exp(-((xy)^2 + x^2 + y^2 - 8x - 8y)/2),$$

para x e y reales.

Imaginemos que queremos estimar $\lambda = \mathbb{E}(X)$ por medio del estimador de Monte Carlo

$$\hat{\lambda} = \frac{\sum_{i=1}^M X_i}{M},$$

donde la muestra $\{(X_i, Y_i)\}_{i=1}^n$ que aproxima a la distribución f se genera usando una matriz de transición Q dada por un algoritmo de Metropolis-Hastings **caminata aleatoria**.

Elegiremos que el incremento $Z = (Z_1, Z_2)$ esté dado por variables aleatorias $Z_1, Z_2 \sim N(0, a^2)$ independientes para alguna $a > 0$. Algunas observaciones en la selección de la constante a son las siguientes:

- Si a es demasiado pequeña, los componentes serán muy correlacionados positivamente, por lo que necesitaremos una muestra grande para la convergencia.
- Si a es demasiado grande, muchas de las muestras serán rechazadas, por lo que el algoritmo es poco eficiente.

Entonces, nuestro algoritmo para generar números provenientes de f tiene la forma explícita siguiente.

Algoritmo 49 : Metropolis-Hastings caminata aleatoria

- 1: Iniciamos en un valor arbitrario (X_0, Y_0) , $n = 0$.
 - 2: Dado el vector (X_n, Y_n) , generamos $Z = (Z_1, Z_2)$ donde $Z_1, Z_2 \sim N(0, 1)$ independientes. Hacemos $C = (X_n, Y_n) + a * Z$.
 - 3: Calculamos $\alpha((X_n, Y_n), C) = \min\{\frac{f(C)}{f(X_n, Y_n)}, 1\}$, donde f es la función de densidad objetivo.
 - 4: Generamos $U \sim U(0, 1)$. Si $U < \alpha((X_n, Y_n), C)$, hacemos $(X_{n+1}, Y_{n+1}) = C$ y $n = n + 1$. Volvemos al paso 2.
 - 5: Paramos cuando $n = N$, un valor dado N de iteraciones totales.
-

A continuación presentamos un ejemplo concreto en donde estimar la distribución estacionaria nos puede ayudar a estimar la solución de un problema de optimización relacionado con una evolución markoviana de un sistema a través del tiempo.

Ejemplo 7.1.2. *Una persona tiene el proyecto de poner una pensión para autos con K cajones. Supongamos que Y_n es el número de personas que acuden a rentar un cajón el n -ésimo día (el contrato es por día) y que $\{Y_n\}_{n \geq 0}$ son v.a.i.i.d. $Poisson(\lambda)$. Si hay cajones disponibles se consideran como contratos nuevos, en otro caso cada cliente toma su lugar ya asignado. La probabilidad de que una persona desocupe el cajón al día siguiente es p (los cajones se regresan siempre al inicio del día). Sea X_n el número de cajones en renta al final del n -ésimo día y Z_n el número de nuevos contratos durante el n -ésimo día. Cada nuevo contrato genera una ganancia de g pesos y un ingreso de r pesos por día (o fracción del día) y cada cajón se puede comprar sólo al principio con un costo de c pesos por cajón y por día. El objetivo es encontrar k que maximice la ganancia.*

Notemos que

$$X_{n+1} = \min\{k, Bin(X_n, 1 - p) + Y_{n+1}\},$$

y

$$Z_{n+1} = X_{n+1} - Bin(X_n, 1 - p).$$

Además $\{X_n\}_{n \geq 0}$ es una cadena de Markov homogénea a tiempo discreto con probabilidades de transición

$$\mathbb{P}(X_{n+1} = j | X_n = i) = \sum_{l=1}^{\min\{i, j\}} \binom{i}{l} (1-p)^l p^{i-l} \frac{\lambda^{j-l} e^{-\lambda}}{(j-l)!},$$

con $0 \leq i \leq k$ y $0 \leq j \leq k - 1$. Si $j = k$ entonces $p_{ik} = 1 - \sum_{j=0}^{k-1} p_{ij}$.

Para maximizar la ganancia promedio por día (a largo plazo), tenemos que maximizar

$$M(k) = g\mathbb{E}[Z(k)] + r\mathbb{E}[X(k)] - ck \quad (7.1)$$

donde

$$\mathbb{E}[Z(k)] = \lim_{n \rightarrow \infty} \mathbb{E}[Z_n(k)] = \lim_{n \rightarrow \infty} \mathbb{E}[X_n(k)] - \mathbb{E}[\text{Bin}(X_{n-1}(k), 1 - p)] = p\mathbb{E}[X(k)].$$

Entonces podemos reescribir a la función objetivo (7.1) como

$$M(k) = (g + r)\mathbb{E}[X(k)] - sk,$$

donde

$$\mathbb{E}[X(k)] = \sum_{x=0}^k x\pi_x$$

y π es la distribución estacionaria de la cadena $\{X_n\}_{n \geq 0}$. Si usamos la media muestral tendremos un estimador para k .

7.2. Muestreo de Gibbs

En esta sección estudiaremos un caso particular del algoritmo de Metropolis-Hastings que es muy utilizado para generar muestras de vectores aleatorios y ha sido extensamente usado en estadística bayesiana. Este método es conocido como el muestreo de Gibbs. La principal característica del muestreo de Gibbs es que la Cadena de Markov subyacente es construida de una sucesión de distribuciones condicionales.

Supongamos que queremos obtener una muestra un vector aleatorio $X = (X_1, X_2, \dots, X_k)$ con función de densidad conjunta $f(X)$. Denotemos por $X_{[i]}$ al vector que se obtiene al considerar el vector X tras eliminar la i -ésima entrada. Consideremos la densidad condicional de X_i dado $X_{[i]}$, dada por

$$f(x_i|x_{[i]}) = \frac{f(x_i, x_{[i]})}{f(x_{[i]})} = \frac{f(x)}{\int_{-\infty}^{\infty} f(x)dx_i},$$

donde hemos extendido la notación de X, X_i y $X_{[i]}$ a vectores deterministas x, x_i y $x_{[i]}$.

Supongamos que las distribuciones condicionales son conocidas y fáciles de simular. Consideremos el algoritmo de Metropolis-Hastings eligiendo a la probabilidad de transición q de la siguiente manera particular: Elegimos una entrada i de manera equiprobable entre las coordenadas de X . Construimos Y como un vector idéntico a X salvo en la coordenada i . La coordenada i -ésima del vector Y se elige con la densidad de probabilidad:

$$q(x_i|x) := f(x_i|x_{[i]}).$$

Entonces, tendremos que

$$\alpha(x_i, x) = \min \left\{ 1, \frac{f(x_i)f(x_{[i]}|x_i)}{f(x_{[i]})f(x_i|x_{[i]})} \right\} = 1,$$

es decir, se acepta siempre al candidato; no hay rechazo.

Así, podemos escribir el algoritmo del Muestreo de Gibbs de forma explícita.

Ejemplo 7.2.1. *Distribución Normal Bivariada.*

Algoritmo 50 : Muestreo de Gibbs

- 1: Inicializar $X^0 = (X_1^{(0)}, \dots, X_k^{(0)})$.
 - 2: Elegimos un índice i de forma equiprobable en $1, \dots, k$.
 - 3: Generamos $Z \sim f(X_i^{(n)} | X_{[i]}^{(n)})$.
 - 4: Hacemos $X^{(n+1)}$ como la concatenación de $X_{[i]}^{(n)}$ y Z (poniendo Z como la i -ésima coordenada).
 - 5: Repetir los paseos anteriores.
-

Recordemos que un vector aleatorio (X, Y) continuo que toma valores en el plano \mathbb{R}^2 tiene distribución Normal Bivariada si su función de densidad está dada por

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\frac{(x-\mu_X)^2}{\sigma_X^2} - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} \right] \right\},$$

para x, y reales, donde ρ es el coeficiente de correlación. Consideremos el caso cuando $\sigma_X = \sigma_Y = 1$ y $\mu_X = \mu_Y = 0$, entonces tenemos que la función de densidad correspondiente es:

$$f_{X,Y}(x, y) = \frac{1}{2\pi} \exp \left\{ -\frac{1}{2(1-\rho^2)} [x^2 - 2\rho xy + y^2] \right\}.$$

Es fácil ver que $X, Y \sim N(0, 1)$, y también verificar que

$$X|Y \sim N(\rho Y, 1 - \rho^2)$$

y

$$Y|X \sim N(\rho X, 1 - \rho^2),$$

por lo que podemos simular las densidades condicionales utilizando cualquiera de los métodos presentados antes para simular una densidad normal.

Ejemplo 7.2.2. *Implementemos el algoritmo anterior en un programa.*

A continuación presentamos la implementación de tal código. La siguiente función implementa el Muestreo de Gibbs donde los vectores generados tienen aproximadamente la densidad de una normal bivariada con el coeficiente de correlación deseado.

```
Gibbs<-function(k, rho){  
  
  A=matrix(0,k,2)  
  x=rnorm(1) #Valores iniciales  
  y=rnorm(1) #Valores iniciales  
  A[1,1]=x*sqrt(1-rho^2)+rho*y  
  #Generamos otro parametro con la condicional en Xo  
  A[1,2]<-x*sqrt(1-rho^2)+rho*A[1,1]  
  #Generamos el otro con la misma condicional pero ahora en A[1,1]  
  for (i in 2:k){ #Repetimos k veces  
    A[i,1]<-rnorm(1)*sqrt(1-rho^2)+rho*A[i-1,2]
```

```

    A[i,2]<-rnorm(1)*sqrt(1-rho^2)+rho*A[i,1]
  }
  return (A)
}

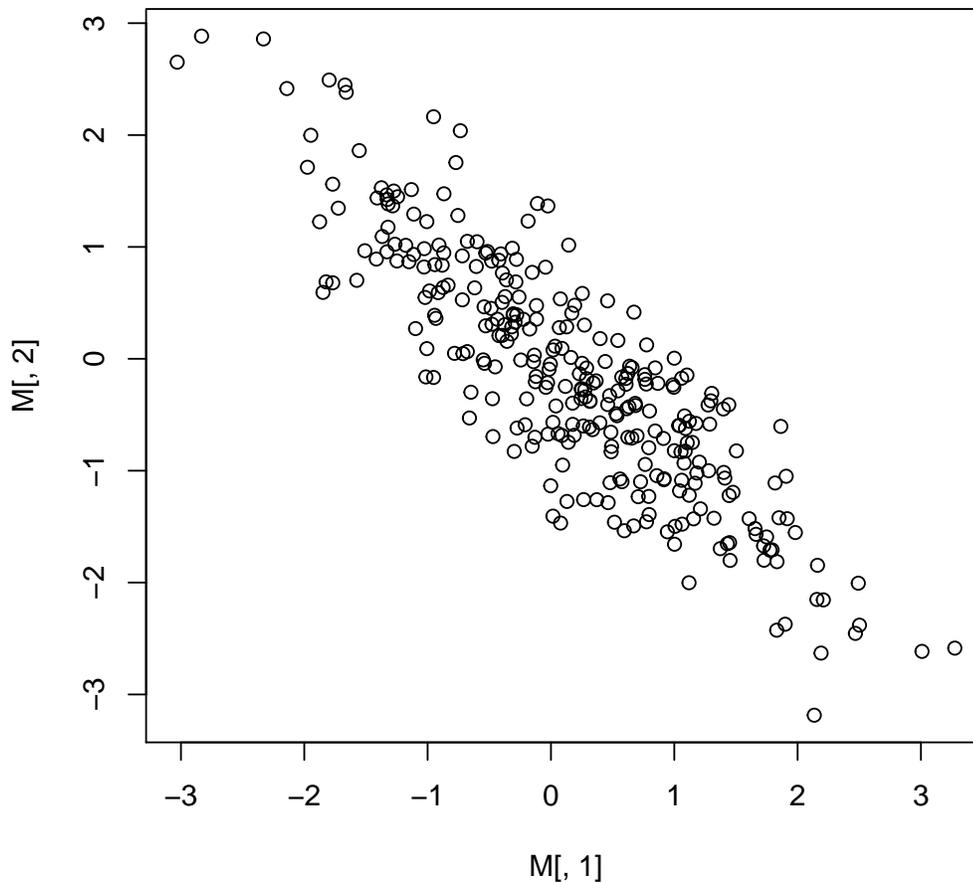
```

Podemos graficar los valores obtenidos.

```

M=Gibbs(300,-.85)
plot( M[,1],M[,2])

```



Un ejercicio sugerido es utilizar nuestro programa para resolver un pequeño problema de estimación: Supongamos que deseamos estimar el valor de la probabilidad

$$\mathbb{P}(X^2 + Y^2 > 1.5)$$

cuando (X, Y) tiene distribución normal estándar bivariada con coeficiente de correlación $\rho = -0.85$. Considere que el programa contemple un tamaño de muestra n y un período de calentamiento k .

Ejemplo 7.2.3. *Distribución Autoexponencial.*

Este modelo fue introducido por Besag (1974) en el área de estadística espacial. Supongamos que

$$f(y_1, y_2, y_3) \propto \exp\{-(y_1 + y_2 + y_3 + \theta_{12}y_1y_2 + \theta_{13}y_1y_3 + \theta_{23}y_2y_3)\},$$

para valores $y_1, y_2, y_3 \in [0, \infty)$, en donde θ_{12} , θ_{13} y θ_{23} son conocidos (recordemos que la notación $f \propto g$ indica que f y g difieren en una constante positiva, es decir, son directamente proporcionales).

Una propuesta inicial para simular observaciones de esta distribución es usando la relación

$$f(y_1, y_2, y_3) = f(y_1)f(y_2|y_1)f(y_3|y_1, y_2).$$

Para poder usar esta representación calculemos la marginal de (y_1, y_2) ,

$$\begin{aligned} f(y_1, y_2) &\propto \int_0^{\infty} f(y_1, y_2, y_3) dy_3 \\ &\propto \exp\{-(y_1 + y_2 + \theta_{12}y_1y_2)\} \int_0^{\infty} \exp\{-y_3(1 + \theta_{23}y_2 + \theta_{31}y_1)\} dy_3 \\ &\propto \frac{\exp\{-(y_1 + y_2 + \theta_{12}y_1y_2)\}}{1 + \theta_{23}y_2 + \theta_{31}y_1}. \end{aligned}$$

Entonces, la marginal de y_1 está dada por

$$f(y_1) \propto e^{-y_1} \int_0^{\infty} \frac{\exp\{-(y_2 + \theta_{12}y_1y_2)\}}{1 + \theta_{23}y_2 + \theta_{31}y_1} dy_2,$$

cuya solución analítica se dificulta. Sin embargo, es fácil mostrar que las densidades condicionales están dadas por

$$\begin{aligned} f(y_1|y_2, y_3) &= \exp(y_1|1 + \theta_{12}y_2 + \theta_{13}y_3), \\ f(y_2|y_1, y_3) &= \exp(y_2|1 + \theta_{12}y_1 + \theta_{23}y_3), \\ f(y_3|y_1, y_2) &= \exp(y_3|1 + \theta_{13}y_1 + \theta_{23}y_2), \end{aligned}$$

donde estamos utilizando la notación $\exp(x|\lambda)$ para denotar a la densidad dada por $f(x) = \lambda e^{-\lambda x}$. Así que estas densidades son muy sencillas de simular y por lo tanto un muestreo de Gibbs es fácil de implementar. Un ejercicio sugerido es la implementación del algoritmo de muestreo de Gibbs para simular observaciones provenientes de la distribución Autoexponencial y graficarlas.

7.2.1. Aplicación en estadística bayesiana

Sea $X = (X_1, \dots, X_n)$ una muestra de una distribución Poisson con un punto de cambio aleatorio, es decir, supongamos que el punto de cambio se presenta en un valor $m \in \{1, \dots, n\} = J_n$ y ocurre que, condicional al valor que toma m , tenemos que

$$X_i \sim \text{Poisson}(\lambda)$$

para $i \in \{1, \dots, m\}$ y

$$X_i \sim \text{Poisson}(\tau)$$

para $i \in \{m+1, \dots, n\}$. Por otro lado, supongamos que λ, τ y m tienen distribuciones a priori (es decir, no condicionales) independientes donde $\lambda \sim \text{Gamma}(\alpha, \beta)$, $\tau \sim \text{Gamma}(\sigma, \delta)$ y $m \sim \text{Unif}\{1, \dots, n\}$, con α, β, σ y δ constantes conocidas.

La interpretación usual es la siguiente. Tenemos un conjunto de parámetros con cierta distribución a priori, que representa nuestro conocimiento previo de la distribución de tales parámetros. Esa información puede provenir de experiencia estadística anterior o de la información que nos proporcione un experto en el fenómeno. Por otro lado, recibimos datos, provenientes de información factual (es decir, trabajo de campo, experimentos en un laboratorio, etc). Nos gustaría recalcular la distribución de los parámetros, dados los datos que obtuvimos; esto servirá para tomar en cuenta la información inicial dada por la distribución a priori, contemplando la certeza que nos dieron los datos recibidos.

La densidad condicional anterior se puede derivar a través de la fórmula de Bayes. En nuestro caso toma la forma específica:

$$f(\lambda, \tau, m|X) = \frac{f(X|\lambda, \tau, m)f(\lambda, \tau, m)}{f(X)}.$$

Podemos verificar, de forma relativamente sencilla, que la densidad conjunta a posteriori (es decir condicional a los datos X) es

$$f(\lambda, \tau, m|X) \propto \lambda^{\alpha+y_m-1} \tau^{\sigma+z_m-1} \exp\{-(\beta+m)\lambda\} \exp\{-(\delta+n-m)\tau\},$$

donde

$$y_m = \sum_{i=1}^m X_i$$

y

$$z_m = \sum_{i=m+1}^n X_i,$$

expresión que no parece fácil de simular. Para hacer uso del muestreo de Gibbs tenemos que encontrar todas las condicionales a posteriori correspondientes, las cuales están dadas por:

$$\lambda|\tau, m, X \sim \text{Gamma}(\alpha + y_m, \beta + m),$$

$$\tau|\lambda, m, X \sim \text{Gamma}(\sigma + z_m, \delta + n - m)$$

y

$$f(m|\lambda, \tau, X) = \frac{\lambda^{\alpha+y_m-1} \tau^{\sigma+z_m-1} \exp\{-(\beta+m)\lambda\} \exp\{-(\delta+n-m)\tau\}}{\sum_{i=1}^m \lambda^{\alpha+y_i-1} \tau^{\sigma+z_i-1} \exp\{-(\beta+i)\lambda\} \exp\{-(\delta+n-i)\tau\}},$$

para toda $m \in \{1, \dots, n\}$. Con lo anterior, somos capaces de programar un muestreo de Gibbs que simule a la distribución conjunta a posteriori $f(\lambda, \tau, m|X)$.

El algoritmo es el siguiente.

Algoritmo 51 : Muestreo de Gibbs 1

- 1: Inicializar τ, λ, m .
 - 2: Elegir τ o λ o m , cada uno con probabilidad $1/3$.
 - 3: Dado que elegimos el parámetro θ , simularlo de nuevo utilizando la densidad condicional $f(\theta | (\lambda, \tau, \theta) \setminus \{\theta\})$.
 - 4: Repetir desde el paso 2.
-

Este algoritmo nos permite hacer lo siguiente: dado un conjunto de datos X , podríamos estimar los valores λ, τ y μ . Este es un ejercicio del capítulo y es un ejemplo típico de estimación utilizando el muestreo de Gibbs.

7.2.2. El muestreo de Gibbs dentro del contexto de la Física

Muchas de las ideas del muestreo de Gibbs tienen sus orígenes en la mecánica estadística. Para mostrar la versatilidad de tales ideas y profundizar en su comprensión presentamos en esta sección un breve esbozo de los conceptos utilizados y la forma específica que toma el muestreo de Gibbs en este contexto. Este lenguaje es utilizado en también en otras áreas como la estadística bayesiana, donde se utilizan ideas similares.

Campos de Markov

Sea S un conjunto finito o numerable con elementos a los que llamaremos sitios y Λ un conjunto finito al cual nos referiremos como espacio fase. Un campo aleatorio con sitios S y espacio fase Λ es una colección $\{X(s)_{s \in S}\}$ de variables aleatorias con valores en Λ .

Alternativamente, un campo aleatorio es una variable aleatoria que toma sus valores en el espacio Λ^S . Al espacio Λ^S le llamaremos espacio de configuraciones. Para una configuración fija x y dado un subconjunto $A \subset S$, definimos:

$$x(A) = (x(s), s \in A)$$

como la restricción de la configuración x al subconjunto A . Por comodidad, algunas veces escribiremos a toda la configuración x como la concatenación de la configuración x restringida a A con la configuración x restringida al complemento de A , es decir

$$x = (x(A), x(S \setminus A)).$$

Veamos un ejemplo. Consideremos $S = \mathbb{Z}_5 \times \mathbb{Z}_5$ y $\Lambda = \{0, 1\}$. En este caso, una configuración posible puede representarse gráficamente de la siguiente manera:

$$\begin{array}{ccccc} + & - & - & - & + \\ - & + & - & + & - \\ + & + & + & - & + \\ - & + & - & + & + \\ + & + & + & - & + \end{array}$$

Ahora procederemos a definir la dependencia posible entre las variables que conforman al campo aleatorio. Para ello definimos un sistema de vecindades en el conjunto de sitios S .

Definición 7.1. Un sistema de vecindades de S es una familia $\{\mathcal{N}_s\}_{s \in S}$ de subconjuntos de S tales que para toda $s \in S$:

1. $s \notin \mathcal{N}_s$,
2. Si $t \in \mathcal{N}_s$ entonces $s \in \mathcal{N}_t$.

Denotaremos por \mathcal{N}_s a la vecindad de s . Y definiremos como $\widetilde{\mathcal{N}}_s = \mathcal{N}_s \cup \{s\}$ a la vecindad no agujereada de s . Cuando S consiste en un malla cuadrada (finita o infinita) un ejemplo clásico de vecindades son las α (vecindad de von Neumann) y β (vecindad de Moore):

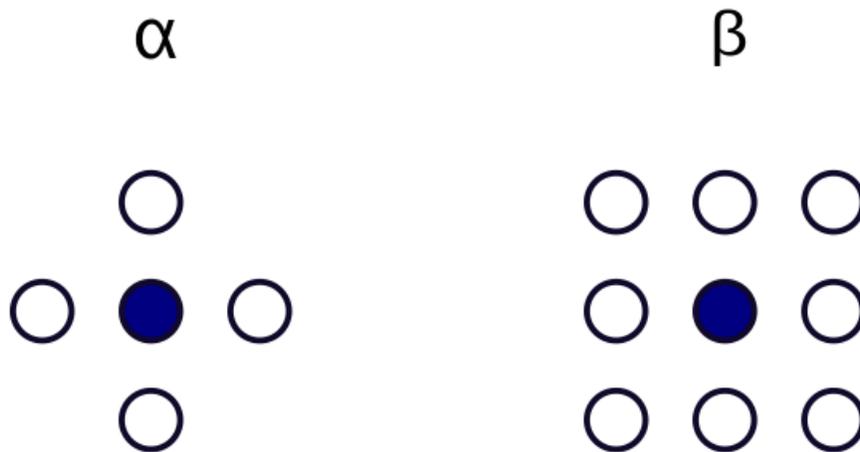


Figura 7.1: Tipos de vecindades.

Ahora explicaremos cómo se puede definir a la propiedad de Markov en este contexto.

Notemos que si tenemos una cadena de Markov $\{X_n\}_{n \geq 0}$ se puede probar (utilizando la propiedad de Markov y el teorema de Bayes) que para todo n natural ocurre lo siguiente:

$$(X_1, \dots, X_{n-1})|_{X_n} \perp (X_{n+1}, \dots, X_{n+m})|_{X_n},$$

es decir, que condicionalmente al valor de X_n , los vectores (X_1, \dots, X_{n-1}) y $(X_{n+1}, \dots, X_{n+m})$ son independientes. Esta propiedad, que es de hecho equivalente a la propiedad de Markov, nos da una intuición espacial: dado un sitio n , si conocemos el valor de la variable X_n , ocurre lo que queda a la izquierda es independiente a lo que queda a la derecha de tal posición.

Siguiendo tal intuición espacial hacemos una extensión para definir la propiedad de Markov en campos aleatorios.

Definición 7.2. Decimos que X es un campo aleatorio de Markov con respecto al sistema de vecindades $\{\mathcal{N}_s\}_{s \in S}$, si para todos los sitios de S las variables aleatorias $X(s)$ y $X(S \setminus \widetilde{\mathcal{N}}_s)$ son independientes

condicionadas a $X(\mathcal{N}_s)$, es decir:

$$\mathbb{P}(X(s) = x(s) | X(S \setminus s) = x(S \setminus s)) = \mathbb{P}(X(s) = x(s) | X(\mathcal{N}_s) = x(\mathcal{N}_s)),$$

para todo $s \in S$ y $x \in \Lambda^S$.

El concepto análogo a las probabilidades de transición para cadenas de Markov está dado por las especificaciones locales para campos de Markov.

Definición 7.3. *La especificación local de un campo aleatorio de Markov en el estado s es la función $\pi^s : \Lambda^S \rightarrow [0, 1]$ definida por:*

$$\pi^s(x) = \mathbb{P}(X(s) = x(s) | X(\mathcal{N}_s) = x(\mathcal{N}_s))$$

La familia $\{\pi^s\}_{s \in S}$ es llamada la especificación local del campo aleatorio de Markov.

Campos de Gibbs

En esta sección definiremos a los campos de Gibbs, que son un tipo de campos aleatorios con una forma específica para su distribución conjunta. Su importancia radica en que, bajo ciertas condiciones generales, un campo de Markov es un campo de Gibbs y viceversa. La forma específica de los campos de Gibbs nos permite hacer cálculos explícitos, por lo que nos da una manera de implementar el muestreo de Gibbs en una forma detallada.

Comenzaremos con los conceptos utilizados para caracterizar la dependencia de las variables aleatorias de un campo de Gibbs (que definiremos después). Continuamos en el contexto de un campo aleatorio con espacio de sitios S y espacio fase Λ .

Definición 7.4. *Un subconjunto $C \subset S$ es un clique si y solo si:*

1. $C = \{s\}$ ó
2. Para todo par de sitios s, t tales que $s, t \in C$ ocurre que s y t son vecinos.

Notemos que para las vecindades α y β presentadas anteriormente tenemos que los posibles cliques, sin contar rotaciones, son de la forma:

Definimos ahora al potencial para un campo de Gibbs, que nos es una colección de funciones en donde cada función asociada a un clique C depende exclusivamente de los valores de la configuración en tal clique C .

Definición 7.5. *Un potencial de Gibbs en el espacio Λ^S relacionado a un sistema de vecindades $\{\mathcal{N}_s\}_{s \in S}$ es una colección de funciones $\{V_C\}_{C \subset S}$, $V_C : \Lambda^S \rightarrow \mathbb{R} \cup \{\infty\}$ tales que:*

1. $V_C \equiv 0$ si C no es un clique.
2. $\forall x, x' \in \Lambda^S$ y $\forall C \subset S$ si $x(C) = x'(C) \implies V_C(x) = V_C(x')$

A través del potencial de Gibbs, podemos definir a la energía de una configuración x como la suma de todos las funciones potenciales (sumando sobre todos los cliques) de la configuración x .

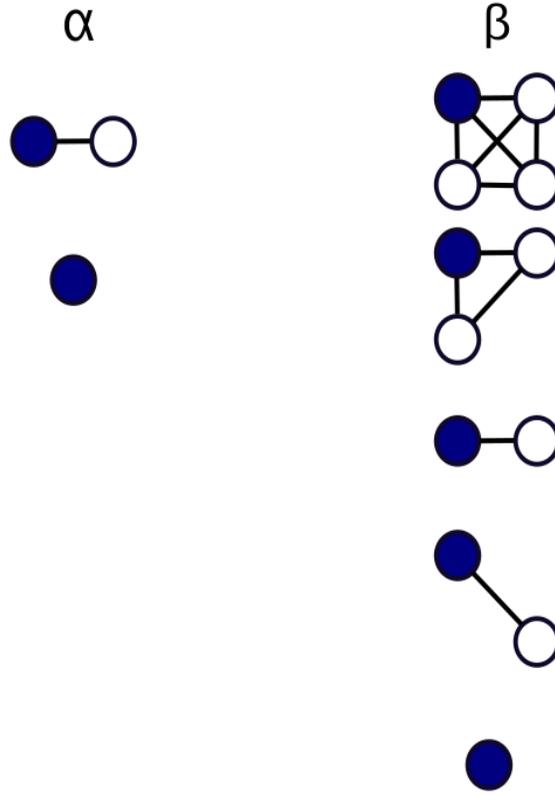


Figura 7.2: Todos los cliques de cada vecindad (salvo rotaciones).

Definición 7.6. La función de energía $\varepsilon : \Lambda^S \rightarrow \mathbb{R} \cup \{\infty\}$ está dada por

$$\varepsilon(x) = \sum_C V_C(x).$$

Tras definir estos conceptos, podemos definir a una distribución de Gibbs. Un campo aleatorio tiene una distribución de Gibbs si tiene la forma:

$$\pi_T(x) = \frac{1}{Z_T} e^{-\frac{1}{T}\varepsilon(x)},$$

donde Z_T es una constante de normalización llamada función de partición, que hace que $\{\pi_T(x)\}_{x \in \Lambda^S}$ sea una función de masa de probabilidades. La función de energía proviene de un potencial de Gibbs, para un sistema de cliques. Al parámetro $T > 0$ se le conoce como la temperatura del sistema, y tiene la siguiente intuición (correspondiente a la física): cuando T crece a infinito, la distribución de Gibbs se acerca a una distribución uniforme sobre el espacio de configuraciones Λ^S . En el caso en que T decrece a 0, la distribución de Gibbs tiende a la distribución de probabilidad de una variable aleatoria que toma con probabilidad 1 a la configuración x^* que maximiza $\pi_T(x)$.

En cuanto a la relación entre campos de Gibbs y campos de Markov nos contentaremos en enunciar el siguiente resultado, que nos indica que los campos de Gibbs son de hecho campos de Markov. El inverso es cierto: bajo ciertas condiciones muy generales, un campo de Markov puede expresarse como un campo de Gibbs.

Teorema 7.1. Si X es una v.a. con distribución π donde su función de energía proviene de un potencial de Gibbs $\{V_C\}_{C \subset S}$ relativo al sistema de vecindades de $\{\mathcal{N}_s\}_{s \in S}$ entonces X es un campo aleatorio de Markov respecto a las mismas vecindades $\{\mathcal{N}_s\}_{s \in S}$ y sus especificaciones están dadas por:

$$\pi^s(x) = \frac{e^{-\sum_{C \ni s} V_C(x)}}{e^{-\sum_{\lambda \in \Lambda} V_C(\lambda, x(S \setminus s))}}.$$

La notación $\sum_{C \ni s}$ significa la suma sobre todos los conjuntos C tales que contengan al estado s .

Veamos ahora un ejemplo específico de todos los conceptos presentados. Consideremos el modelo de Ising, un ejemplo clásico de la física estadística en donde se modela el ferromagnetismo. En este caso podemos considerar $S = \mathbb{Z}_m^2$ un toro discreto, y un espacio fase $\{+, -\}$ donde los valores del espacio fase representan si un sitio s está imantado positiva o negativamente.

Sabemos que, físicamente, un cuerpo tiende a imantarse con el mismo signo que los cuerpos que tiene cercanos. Eso lo podemos expresar de la siguiente manera: consideremos un sistema de vecindades α y cliques dados por tales vecindades. Consideremos que los únicos cliques con función potencial distinto de cero tienen la forma $\{s, t\}$ donde s y t difieren exactamente en una coordenada, por $+1$ o -1 y el potencial está dado por:

$$V_{\{s,t\}} = -x(s)x(t).$$

Así, la energía está dada por

$$\varepsilon(x) = - \sum_{\{s,t\} \text{ clique}} x(s)x(t)$$

y la medida de Gibbs inducida es la siguiente:

$$\pi_T(x) = \frac{1}{Z_T} e^{\frac{1}{T} \sum_{\{s,t\} \text{ clique}} x(s)x(t)}.$$

Notemos que esta medida codifica el fenómeno que queremos capturar. La configuración

$$\begin{array}{ccccc} + & + & + & + & + \\ + & + & + & + & + \\ + & + & + & + & + \\ + & + & + & + & + \\ + & + & + & + & + \end{array}$$

tiene todos los vecinos con fases del mismo signo, por lo que su energía será pequeña. Eso implica que tal configuración tiene gran probabilidad. Por otro lado, la configuración

$$\begin{array}{ccccc} + & - & + & - & + \\ - & + & - & + & - \\ + & - & + & - & + \\ - & + & - & + & - \\ + & - & + & - & + \end{array}$$

tiene muchos más vecinos con fases de diferente signo, por lo que su energía será mayor y por lo tanto su probabilidad será menor que la de la configuración de arriba.

Utilizando el Teorema 7.1 tenemos que las especificaciones locales para nuestro ejemplo están dadas por

$$\pi^s(x) = \frac{g(x(s))}{g(x(s)) + g((1 - x(s)))},$$

donde

$$g(i) = e^{-ix(s+\mathbf{e}_1)+ix(s-\mathbf{e}_1)x(s)+ix(s+\mathbf{e}_2)+ix(s-\mathbf{e}_2)} \quad (7.2)$$

para valores $i = +1, -1$, donde \mathbf{e}_1 y \mathbf{e}_2 representan a los vectores canónicos.

Utilizando tales especificaciones locales (funciones de transición), podemos utilizar el muestreo de Gibbs para simular a la distribución estacionaria de un campo aleatorio de Ising en el toro finito discreto. El algoritmo es el siguiente:

Algoritmo 52 : Modelo de Ising

- 1: Comenzar con una configuración arbitraria $x_0 \in \{+, -\}^{\mathbb{Z}_m^2}$.
 - 2: Elegir al azar y de forma equiprobable un sitio s in \mathbb{Z}_m^2 .
 - 3: Actualizar el valor de la configuración en el sitio s : $x(s)$ toma el valor $+$ con probabilidad $\pi^s(1)$, y el valor $-$ con probabilidad $\pi^s(-1)$, donde π^s está definida por (7.2).
 - 4: Volvemos al paso 2.
-

La implementación de tal algoritmo nos ayudaría a estudiar la evolución del magnetismo en este modelo simple.

7.3. Muestreo de Slice

En esta sección analizaremos otra forma particular del muestreo de Metropolis-Hastings, que tiene una interpretación geométrica muy directa.

Supongamos que queremos simular valores de una variable aleatoria X con función de densidad $f(x)$. Consideremos al conjunto

$$B = \{(x, y) : 0 < y < f(x)\}.$$

La idea básica es la siguiente: si conseguimos generar un vector (X, Y) que tenga distribución uniforme sobre el conjunto B , entonces la marginal X tendrá la densidad deseada. Esto se puede verificar fácilmente:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy = \int_0^{f(x)} c dy = cf(x),$$

por lo que X tiene a f como densidad (y además la constante de normalización c es igual a 1).

Sin embargo, generar un vector uniforme (X, Y) en el conjunto B no necesariamente es sencillo. Por lo cual se propone el siguiente algoritmo, que como veremos después, es un algoritmo similar al muestreo de Gibbs:

Algoritmo 53 : Muestreo Slice

- 1: Comenzar con un valor arbitrario $X = x$, tal que $f(x) > 0$.
 - 2: Dado $X = x$, elegir Y como un valor uniforme entre 0 y $f(x)$.
 - 3: Dado $Y = y$, elegir X de manera uniforme en el conjunto S , donde $S = \{x : y < f(x)\}$.
 - 4: Volver al paso 2.
-

El nombre del algoritmo Slice (rebanar, en inglés) toma su nombre de la interpretación geométrica del algoritmo: Dado el valor $X = x$, se elige el valor Y como un nivel uniforme en la línea que une al punto

$(x, 0)$ con el punto $(f(x), 0)$, es decir, como un valor uniforme en una rebanada vertical. Dado el valor $Y = y$ se elige al punto X como un valor uniforme en el segmento (o unión de segmentos) de los valores x que al evaluarlos llegan a la línea de nivel y (en el conjunto $\{x : f(x) = y\}$), es decir, como un valor uniforme en una rebanada horizontal.

Este algoritmo tiene particularidades a observar. Por un lado, a diferencia del algoritmo de Metropolis, por ejemplo con el esquema de caminata aleatoria, hay una autoregulación de los valores que va tomando X : la misma densidad f regula qué tan probable es saltar al valor siguiente, haciendo improbable ir a un valor donde la densidad es pequeña y haciendo más probable ir a valores donde la densidad es grande. Por otro lado, los valores consecutivos generados por el algoritmo tienen alta dependencia.

A continuación probaremos la eficacia del algoritmo. Para ello basta verificar que este algoritmo es una variación del muestreo de Gibbs, donde la distribución invariante es uniforme sobre el conjunto B .

Notemos que debido a que las actualizaciones son deterministas (primero en X , luego en Y , luego en X , luego en Y , etc.) tenemos un esquema de Metropolis-Hastings, con matriz de transición:

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

Ahora, para verificar que en efecto tenemos como distribución estacionaria a un vector (X, Y) uniforme sobre el conjunto B , nos basta ver que las transiciones dadas por el algoritmo coinciden con las densidades marginales $f_{Y|X}$ y $f_{X|Y}$ de tal vector (y por lo tanto la probabilidad de aceptación es igual a 1). Lo probamos a continuación.

Proposición 7.3. *Sea (X, Y) con distribución uniforme sobre el conjunto $B = \{(x, y) : 0 < y < f(x)\}$. Entonces ocurre que*

1. $Y|_{X=x}$ tiene distribución uniforme en el intervalo $(0, f(x))$.
2. $X|_{Y=y}$ tiene distribución uniforme sobre el conjunto S , donde $S = \{x : y < f(x)\}$.

Demostración. 1) Notemos que para x tal que $f(x) > 0$ ocurre que

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)} = \frac{1_B(x, y)}{f(x)} = \frac{1_{\{0 < y < f(x)\}}(y)}{f(x)},$$

donde utilizamos lo que presentamos al principio de esta sección: la marginal de X es la función $f(x)$.

2) Definamos $S = \{x : y < f(x)\}$. Calculemos primero la densidad de Y :

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx = \int_S 1 dx = \lambda(S)$$

donde $\lambda(S)$ denota a la longitud del conjunto S . Entonces, para $0 \leq y \leq \max_{x \in \mathbb{R}} f(x)$ ocurre que

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = \frac{1_B(x, y)}{\lambda(S)} = \frac{1_S(x)}{\lambda(S)},$$

que es lo que se quería demostrar. □

7.3.1. Implementación y calibración

Es importante aclarar que en la implementación de simulaciones Monte Carlo a través de cadenas de Markov pueden aparecer varios problemas. En esta sección discutimos algunos de estos problemas y las soluciones prácticas que se suelen utilizar para remediarlos.

Uno de los problemas es el siguiente. Aunque la existencia de una distribución estacionaria y la convergencia a ella pueden ser demostrados formalmente, las condiciones para que se cumpla la convergencia no son fáciles de demostrar en problemas concretos y que aún si pudiera hacerse, la convergencia puede ser muy lenta. Por ejemplo, en el caso del algoritmo de Metropolis-Hastings, la probabilidad de moverse a una nueva posición, puede ser muy pequeña y la cadena podría tardarse mucho en visitar todo el espacio.

Es entonces muy importante corroborar si efectivamente la cadena alcanza a la distribución estacionaria. Debemos dar un tiempo a que la cadena se estabilice, es decir, tomar un cierto número de iteraciones como un “período de calentamiento” y luego de este período analizar la convergencia y estabilidad de la cadena. Para determinar el período de calentamiento una sugerencia es hacer gráficas de los estados visitados por la cadena. También se pueden graficar funciones de interés (por ejemplo la media muestral). Lo fundamental en este sentido es monitorear la convergencia de las características estadísticas de los datos. Por ejemplo, se pueden observar los coeficientes de correlación serial, para tener idea de la velocidad de convergencia (correlación alta implica convergencia lenta). Una propuesta para medir la correlación serial de orden k está dada por:

$$\rho_k = \frac{\sum_{j=1}^{n-k} (x_j - \bar{x})(x_{j+k} - \bar{x})}{\sum_{j=1}^n (x_j - \bar{x})^2}$$

El monitoreo consiste en verificar si el estimador anterior tiende a cero conforme k crece.

Un problema fundamental subyacente es que las observaciones generadas no son independientes (pues son construidos a través de la dinámica de una cadena de Markov) a pesar de que el objetivo es generar una muestra aleatoria.

Existen diversas sugerencias para resolver el problema de la dependencia de los datos. Gelfand & Smith (1990), sugieren hacer corridas simultaneas de n cadenas de forma independientemente hasta obtener la convergencia en cada una. El problema en esta metodología es que si la convergencia se alcanza después de m iteraciones, entonces para obtener una muestra de tamaño n , es necesario generar mn observaciones. Por otro lado, Geyer (1992), sugiere que una vez que se alcanza la convergencia, basta con generar n observaciones más y tomar estas como muestra, pues aunque no son independientes la teoría ergódica sustenta la inferencia. Otra propuesta es la de Raftery & Lewis (1992) quienes sugieren que, tras un período de calentamiento, se seleccione un elemento para la muestra periódicamente, es decir, cada k observaciones. Por su parte, Gelman & Rubin (1992) proponen que es recomendable correr, de manera simultánea, l cadenas independientes, con l no mayor a diez, y de cada una tomar n/l observaciones, ya sea de manera consecutiva o cada k pasos. Se recomienda realizar distintas metodologías como las sugeridas.

7.4. Ejercicios

1. Implementar el algoritmo propuesto para la distribución Autoexponencial.
2. Implementar un algoritmo de Metropolis-Hastings para generar valores provenientes de una variable aleatoria $X \sim N(0, 1)$. Utilice el método de la caminata aleatoria simétrica, eligiendo a una transición

Q de la forma:

$$Y = X_n + U,$$

donde $U \sim Unif[-a, a]$ para $a > 0$. Hacer un histograma con una muestra grande generada por el algoritmo.

3. Hagamos un algoritmo de Metropolis-Hastings para generar valores provenientes de una variable aleatoria con densidad:

$$f(t) = C \text{sen}(x)^2 \text{sen}(2t)^2 \phi(t),$$

donde $C > 0$ y $\phi(t)$ denota a la densidad de una normal estándar. Utilice el método de la caminata aleatoria simétrica, eligiendo a una transición Q de la forma:

$$Y = X_n + U,$$

donde $U \sim Unif[-a, a]$ para $a > 0$. Permita que el algoritmo tenga un período de calentamiento k , genere después una muestra de tamaño n grande y gráfiquela con un histograma.

4. Para una normal multivariada estándar con coeficiente de correlación $\rho = -0.85$, estimar el valor

$$\mathbb{P}(X^2 + Y^2 > 1.5)$$

utilizando el muestreo de Gibbs. Considere que el programa contemple un tamaño de muestra n y un período de calentamiento k .

5. En la sección del ejemplo de aplicación bayesiana, considere una muestra de datos dada por

$$X = \{7, 3, 6, 5, 3, 2, 0, 6, 4, 5, 2, 2, 4, 4, 3, 4, 3, 3, 5, 0\},$$

y $\alpha = 4$, $\beta = \frac{3}{2}$, $\sigma = 2$, $\delta = 1$. Implementar el algoritmo y estimar los parámetros λ , τ y m . Comparar con los parámetros reales que resultaron al hacer la simulación de la muestra: $\lambda = 4.2729$, $\tau = 2.9205$, $m = 14$.

6. Hacer un programa que simule y grafique el modelo de Ising en el toro $\mathbb{Z}_5 \times \mathbb{Z}_5$. Hacer que el programa imprima la configuración inicial, al paso 50, al paso 1000, y al paso 10,000. Qué pasa a largo plazo? Hay una o varias configuraciones finales deterministas o converge hacia una distribución aleatoria estacionaria?

7. Utilizar el muestreo de slice para simular una muestra proveniente de la densidad

$$f(x) = \frac{1}{2} \exp(-\sqrt{x}), \quad x > 0.$$

- a) Simular, mediante el muestreo de Slice, una colección de variables $\{X_{k+1}, \dots, X_{k+n}\}$, para algún período de calentamiento k , y n fijo. Calcular la distribución empírica:

$$\hat{F}(t) := \frac{1}{n} \sum_{i=1}^n 1_{\{X_i \leq t\}},$$

y graficarla.

- b) Probar que si X tiene la densidad f y se hace la transformación $Y = \sqrt{X}$ entonces $Y \sim \text{Gamma}(\frac{3}{2}, 1)$.
- c) Utilizar el inciso anterior para simular una muestra de v.a. $\{X_1, \dots, X_n\}$ con densidad f (utilizando el mismo n que en el primer inciso). Calcular la distribución empírica y graficarla.
- d) Comparar ambas aproximaciones con la densidad teórica. Qué tan buenas son?Cuál es mejor? Qué tanto difieren, alterando k y n ?

