



HAL
open science

Metodos Numericos por Cadenas de Markov (en español)

Xavier Guyon

► **To cite this version:**

Xavier Guyon. Metodos Numericos por Cadenas de Markov (en español). École thématique. Metodos Numericos por Cadenas de Markov, Mérida - Venezuela, 1999, pp.102. cel-00762829

HAL Id: cel-00762829

<https://cel.hal.science/cel-00762829>

Submitted on 8 Dec 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

MÉTODOS NUMÉRICOS POR CADENAS DE MARKOV

Xavier GUYON
Université Paris 1

y
Universidad de Los Andes, Mérida, Venezuela

Escuela Venezolana de Matemáticas
Septiembre 1999

Índice general

1. Cadenas de Markov con un número finito de estados	1
1.1. Cadenas de Markov	1
1.1.1. Definiciones y notaciones básicas	1
1.1.2. Ley invariante y ergodicidad.	2
1.1.3. Cadena irreducible: existencia y unicidad de la medida invariante	2
1.1.4. El Teorema de Perron-Frobenius	3
1.2. Resultados de ergodicidad	4
1.2.1. El teorema de Frobenius	4
1.2.2. Lema de contracción para P	5
1.2.3. Ergodicidad y acoplamiento de cadenas de Markov	6
1.3. Reversibilidad y medida invariante	7
1.4. Ergodicidad para un espacio de estados general	8
1.4.1. Espacio de estados discreto	8
1.4.2. Espacio de estados general	8
1.5. Ejercicios	9
2. Simulación usando Cadenas de Markov	15
2.1. El problema y el método	15
2.2. El muestreador de Gibbs	16
2.2.1. Muestrador de Gibbs	17
2.2.2. Muestrador de Gibbs con barrido aleatorio	19
2.2.3. Muestrador con barridos síncronos	20
2.2.4. Muestrador de Gibbs sobre $E = \mathbb{R}^n$ (o $(\mathbb{R}^d)^n$)	21
2.3. La dinámica de Metropolis-Hastings	22
2.3.1. Construcción de la transición de M.H.	22
2.3.2. π -reversibilidad de M.H. y ergodicidad	23
2.3.3. Ejemplos	24
2.3.4. Algunos resultados generales sobre las transiciones π -reversibles	26
2.3.5. Comparación de diferentes dinámicas de M.H.	27
2.3.6. Algoritmo de M.H. para un espacio de estados general [98]	29
2.4. Ejercicios	32
3. Cadena inhomogénea y recocido simulado	39
3.1. Coeficiente de contracción de Dobrushin	39
3.1.1. Preliminares	39

3.1.2.	Coeficiente de contracción y ergodicidad	40
3.2.	Optimización por recocido simulado	46
3.2.1.	Recocido simulado para la dinámica de Gibbs	46
3.2.2.	Recocido simulado para la dinámica de Metropolis	48
3.3.	Simulación y optimización con restricciones	51
3.3.1.	Simulación y optimización bajo restricciones para la dinámica del muestreador de Gibbs	52
3.3.2.	Simulación y optimización bajo restricción para la dinámica de Metropolis	54
3.4.	Ejercicios	54
4.	Diagnostico de convergencia de una cadena	61
4.1.	Fenmeno de Cutoff y test de parada de una cadena	63
4.1.1.	Convergencia abrupta para una cadena	63
4.1.2.	Ejemplo 2 (continuación) : tiempo de fusión y regla de parada [109]	65
4.2.	Simulación exacta por acoplamiento desde el pasado	66
4.2.1.	Acoplamiento desde el pasado	66
4.2.2.	Un nuevo germen para definir $\mathcal{S}_{-t} = \{f_{-t}(i), i \in I\}$	68
4.2.3.	Algoritmo de Monte-Carlo monótono	68
4.3.	Ejercicios	72
5.	Problemas inversos y reconstrucción bayesiana	77
5.1.	Modelo de Gibbs y campo de Markov en un conjunto finito	77
5.1.1.	Modelo de Gibbs	77
5.1.2.	Especificación condicional y propiedad de Markov	78
5.1.3.	Campo de Markov y campo de Gibbs	79
5.1.4.	Ejemplos	79
5.2.	Procesos puntuales de Gibbs	83
5.2.1.	Propiedades de Markov	84
5.2.2.	Un ejemplo: el proceso de interacción de reas	84
5.3.	Modelización de Gibbs y problemas inversos	85
5.3.1.	Formación de la imagen y reconstrucción bayesiana	85
5.4.	Algunos ejemplos	87
5.4.1.	Carácter markoviano de la distribución a posteriori $P(x y)$	87
5.4.2.	Segmentación de imágenes	88
5.4.3.	Detección de rupturas en una señal	89
5.4.4.	Reconocimiento de objetos	89
5.5.	Ejercicios	90

Capítulo 1

Cadenas de Markov con un número finito de estados

Este capítulo constituye una introducción a las cadenas de Markov con *un número finito de estados*. Citemos algunos títulos de referencia: Aldous y Fill [3], Feller [30], Isaacson y Madsen [53], Kemeny y Snell [59] y Seneta [92]. Siempre señalaremos situaciones en las cuales el espacio de estados no es finito, por ejemplo \mathbb{R}^d .

1.1. Cadenas de Markov

1.1.1. Definiciones y notaciones básicas

Sea $E = \{1, 2, \dots, r\}$ un espacio de estados finito. Una *matriz de transición* en E es una matriz P , de dimensión $r \times r$, con coeficientes todos ≥ 0 , y donde la suma de cada fila vale 1:

$$P = (p_{ij})_{1 \leq i, j \leq r}, \quad \forall i, j \in E, \quad p_{ij} \geq 0 \quad \text{y} \quad \forall i \in E, \quad \sum_{j \in E} p_{ij} = 1$$

P también se llama matriz estocástica: cada fila de P es una distribución de probabilidad sobre E .

Una *cadena de Markov* sobre E con matriz de transición P es una sucesión $X = (X_0, X_1, X_2, \dots)$ de variables aleatorias con índices en \mathbb{N} y con valores en E tal que $\forall n \geq 0$, se verifica:

$$\begin{cases} P(X_{n+1} = i_{n+1} \mid X_l = i_l, 0 \leq l \leq n) = P(X_{n+1} = i_{n+1} \mid X_n = i_n) \\ P(X_{n+1} = j \mid X_n = i) = p_{ij} \end{cases}$$

La primera igualdad expresa la *propiedad de Markov*: la ley de X_{n+1} condicional al pasado $(X_n, X_{n-1}, \dots, X_1, X_0)$ depende únicamente del estado del último instante $X_n = i_n$: la i -ésima fila de P no es otra que la ley $(X_{n+1} \mid X_n = i)$. La segunda igualdad nos dice que estas transiciones son independientes de n ; se dice que la cadena es *homogénea*. Estudiaremos igualmente en este curso cadenas de Markov *no-homogéneas*, en las cuales la transición P_n en el tiempo n depende de n .

La fórmula de las probabilidades totales muestra que la ley de X está completamente caracterizada por P y por la ley inicial μ_0 de X_0 (denotada $X_0 \sim \mu_0$), las leyes finito-dimensionales están dadas por:

$$P(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = \mu_0(i_0) p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-1} i_n}$$

Recíprocamente, es fácil verificar que si la ley de X es de esta forma, X verifica la propiedad de Markov y es una cadena homogénea de matriz de transición P .

Por convención, denotaremos una probabilidad sobre E por un *vector fila* $1 \times r$. Si μ_n es la distribución de X_n , entonces la distribución de X_{n+1} es $\mu_{n+1} = \mu_n P$, producto por la izquierda del vector fila μ_n por la matriz P . En efecto,

$$P(X_{n+1} = j) = \sum_{i \in E} P(X_{n+1} = j \mid X_n = i) P(X_n = i)$$

En particular, P^n es la potencia n -ésima de P y se verifica: $\mu_n = \mu_0 P^n$.

1.1.2. Ley invariante y ergodicidad.

Una pregunta importante es la siguiente: ¿bajo qué condiciones existe una probabilidad π sobre E tal que $\mu_n \rightarrow \pi$ (por ejemplo para la norma l_1 sobre \mathbb{R}^r), independientemente de la distribución inicial? Si este es el caso, se dice que la cadena es *ergódica*. Otra pregunta es a que velocidad esta convergencia tiene lugar.

Siendo uno de los objetivos de este curso la simulación de leyes π , vamos a responder a esta pregunta reformulada de la siguiente manera: sea π una probabilidad sobre E (es suficiente conocer π salvo un factor multiplicativo):

$$\text{Determinar una transición } P \text{ tal que: } \forall \mu_0, \mu_0 P^n \rightarrow \pi.$$

En este problema disponemos de mucha libertad para escoger P . Así P se puede seleccionar bajo diferentes criterios, por ejemplo la facilidad de construcción (cadena reversible), la menor complejidad algorítmica, o aún más la velocidad de la convergencia. Desafortunadamente, en lo que concierne a la velocidad, las respuestas teóricas que tienen un interés práctico para la selección efectiva de P son más bien raras, salvo en casos muy particulares donde la velocidad puede ser factorizada con precisión.

Si una convergencia de este tipo tiene lugar, $\mu_0 P^{n+1} \rightarrow \pi P$, pero también se tiene que $\mu_0 P^{n+1} \rightarrow \pi$. π verifica entonces necesariamente la condición:

$$\pi = \pi P$$

Una probabilidad π , que verifica lo anterior, es una *medida invariante* para P . En efecto, si X_0 tiene distribución π , X_n también tiene distribución π para todo $n \geq 0$. Por otra parte, si la distribución de la cadena es invariante por traslaciones, X es estacionaria. Veremos que si P es irreducible, tal medida invariante existe y es única (caso E finito). Pero esto no es suficiente para garantizar la ergodicidad.

1.1.3. Cadena irreducible: existencia y unicidad de la medida invariante

La transición P es *irreducible* si, para todo par $(i, j) \in E^2$, existe un entero $n(i, j) \geq 1$ t.q. $P^{n(i, j)}(i, j) > 0$. En la práctica, estableceremos la irreducibilidad al proponer, para todo (i, j) , un camino $\mathcal{C}(i, j)$ de i hacia j ($i \mapsto j$), de longitud $n(i, j)$ realizable con una probabilidad > 0 :

$$\mathcal{C}(i, j): i = i_1 \mapsto i_2 \mapsto i_3 \mapsto \cdots \mapsto i_{n(i, j)} = j$$

con $p_{i_l, i_{l+1}} > 0$ para $l = 1, n(i, j) - 1$. Se dice entonces que el estado i se comunica con el estado j .

Una condición de irreducibilidad fuerte se cumple cuando se puede seleccionar un mismo $k = n(i, j)$ común a todos los $(i, j) \in E^2$: todos los estados se comunican a través de caminos de la misma longitud k . Se dice entonces que la transición es *regular* (o que P es *primitiva*), P^k tiene todos sus términos > 0 . Veremos que para E finito: P regular $\Leftrightarrow P$ ergódica.

Notación : No importa que A sea un vector o una matriz, escribiremos $A > 0$ (resp. $A \geq 0$) si todos sus coeficientes son > 0 (resp. ≥ 0).

Proposición 1.1 *Si P es irreducible y E es finito, existe una nica medida invariante π tal que $\pi > 0$, es decir: $\forall i \in E, \pi_i > 0$.*

Demostración ([91]):

1. Si $\mathbf{1} = {}^t(1, 1, \dots, 1)$ es el vector unidad de \mathbb{R}^r , $P\mathbf{1} = \mathbf{1}$. Existe $v \neq 0$ t.q. $vP = v$.
2. Llamemos $|v|$ al vector de coordenadas $|v_i|$. Demostremos que $|v|P = |v|$. Puesto que siempre se verifica $v_j = \sum_i v_i P_{ij}$, $|v_j| \leq \sum_i |v_i| P_{ij}$, supongamos que para un estado j_0 , esta desigualdad es estricta. Dado que P es estocástica, $\sum_i |v_i| = \sum_{ij} |v_i| P_{ij} > \sum_j |v_j|$ lo que no es posible. Así para todo j , $|v_j| = \sum_i |v_i| P_{ij}$ y $|v|$ es autovector por la izquierda para la matriz P asociado al autovalor 1. Dado que $\sum_j |v_j| \neq 0$, existe una medida invariante π asociada a P , $\pi_i = \frac{|v_i|}{\sum_j |v_j|}$ (para probar la existencia de π , hay que señalar que solamente hemos usado la propiedad “ E finito”).
3. Mostremos que cada $\pi_i > 0$. Puesto que P es irreducible, $\exists l$ t.q. $A = (I + P)^l > 0$ (tomar $l = \sup_{i,j} n(i, j)$). Entonces $0 < \pi A = 2^l \pi$, es decir $\pi > 0$.
4. Si $uP = u$, verifiquemos que u tiene todas sus coordenadas del mismo signo. Hemos visto que para un tal u , $|u|P = |u|$, y entonces $u^+ = \frac{1}{2}(u + |u|)$ es también autovector de P asociado a 1. Se presentan dos casos: (i) $u^+ = 0$ y entonces $u \leq 0$; (ii) $u^+ \neq 0$. Del punto 3 se desprende, $u^+ > 0$ y por consiguiente $u > 0$.
5. Sean π y π' dos medidas invariantes; $u = \pi - \pi'$ es el autovector asociado a 1, y tal que $\sum_i u_i = 0$. Como consecuencia del punto 4, esto no es posible a no ser que $u = 0$: hay, entonces, unicidad de la medida invariante. \square

1.1.4. El Teorema de Perron-Frobenius

El resultado precedente es en parte una consecuencia del teorema siguiente, útil para el estudio del comportamiento asintótico de sistemas dinámicos lineales y con coeficientes ≥ 0 ([36]; [57]; [69]; [92]).

Teorema 1.1 (*Perron-Frobenius*) *Sea A una matriz $r \times r$ con coeficientes ≥ 0 , $A \neq 0$.*

- (1) *Existe $\lambda_0 > 0$ y $u_0 \geq 0$ no nulo t.q. $Au_0 = \lambda_0 u_0$. Entonces*
- (2) *Si λ es otro autovalor (en \mathbb{C}) de A , $|\lambda| \leq \lambda_0$. Adems si $|\lambda| = \lambda_0$, $\eta = \lambda \lambda_0^{-1}$ es una raz de la unidad y para $p \in \mathbb{N}$, $\eta^p \lambda_0$ es un autovalor de A .*
- (3) *Si existe k t.q. $A^k > 0$, λ_0 es de multiplicidad 1 y se puede escoger $u_0 > 0$. Todo otro autovalor verifica $|\lambda| < \lambda_0$.*

Demostración: Nos limitaremos al caso $A > 0$. El punto (3) será una consecuencia fácil pues sólo hay que señalar que (λ^k, u) es un par (autovalor, autovector) de A^k cuando (λ, u) lo es para A .

◦ *Autovalor/autovector de Frobenius*: Sea \mathcal{P} el conjunto de probabilidades en E . Definamos la función continua $h : \mathcal{P} \rightarrow \mathbb{R}^+$:

$$h(m) = \inf_{i=1,r} \frac{Am(i)}{m(i)}.$$

Como \mathcal{P} es un subconjunto compacto de \mathbb{R}^r , existe un elemento máximo λ_0 de $h(\mathcal{P})$ y una probabilidad m_0 t.q. $h(m_0) = \lambda_0$. Mostremos que m_0 es un λ_0 -autovector de A . Por definición de h , $Am_0(i) \geq \lambda_0 m_0(i)$ para todo i . Si la desigualdad es estricta para algún i , siendo $A > 0$, $A(Am_0 - \lambda_0 m_0) > 0$, es decir $h(Am_0) > \lambda_0$, lo que contradice la definición de λ_0 . Así, $Am_0 = \lambda_0 m_0$ y $m_0 > 0$.

◦◦ *El espacio propio de Frobenius tiene dimensión 1*: Sea u un λ_0 -autovector de A (si u es complejo, \bar{u} es también autovector y tomaremos $\mathcal{R}(u)$ y $\mathcal{I}m(u)$ como autovectores reales). Sea $c = \inf_{i=1,r} \frac{u(i)}{m_0(i)}$. Se tiene para todo i : $u(i) \geq cm_0(i)$ y es fácil ver que si, para algún índice, la desigualdad es estricta, esto contradice la definición de c . En consecuencia, $u = cm_0$ y el espacio propio de Frobenius es de dimensión 1.

◦◦◦ Sólo resta ver que todo otro autovalor λ verifica $|\lambda| < \lambda_0$. Sea $v \neq 0$ un λ -autovector y $\delta = \inf_i A_{ii}$, $\delta > 0$. Puesto que $(A - \delta I)v = (\lambda - \delta)v$, se tiene para todo i :

$$|\lambda - \delta| |v(i)| \leq \sum_j (A_{ij} - \delta I_{ij}) |v(j)|$$

sea además: $\forall i, (\delta + |\lambda - \delta|) |v(i)| \leq \sum_j A_{ij} |v(j)|$. Volviendo a la definición de λ_0 , se obtiene:

$$\lambda_0 \geq (\delta + |\lambda - \delta|)$$

Si $\lambda \neq \lambda_0$, esto implica $|\lambda| < \lambda_0$ (más precisamente, λ está dentro de la bola abierta con centro en δ , $0 < \delta < \lambda_0$ y de radio $\lambda_0 - \delta$). □

Volvamos al contexto de una cadena de Markov. Si $P = {}^t A$ es una matriz estocástica, $\lambda_0 = 1$. En efecto, es suficiente escribir $Au_0 = \lambda_0 u_0$ y sumar las coordenadas de cada uno de esos vectores: las columnas de A suman 1, y puesto que $\sum_i u_{0i} \neq 0$, $1 = \lambda_0$. Un resultado importante es que todo autovalor λ de P verifica $|\lambda| \leq \lambda_0$, esta desigualdad es estricta si $\lambda \neq \lambda_0$ y esto pasa si P es regular.

1.2. Resultados de ergodicidad

1.2.1. El teorema de Frobenius

El teorema de Frobenius junto con la reducción a la forma canónica de Jordan de P permiten obtener fácilmente la condición necesaria y suficiente de ergodicidad siguiente,

Proposición 1.2 *Supongamos que P es irreducible. Entonces P es ergdica si y slo si P es regular.*

Demostración :

⇒ Puesto que P es irreducible, P admite una única medida invariante $\pi > 0$ tal que para toda distribución inicial μ_0 , $\mu_0 P^n \rightarrow \pi$. En particular $P^n \rightarrow \Pi$ donde $\Pi > 0$ es la matriz tal que cada línea es π . Deducimos entonces que para un k bastante grande, $P^k > 0$: P es regular.

⇐ Como P es irreducible, llamemos π a su medida invariante.

Usaremos la reducción de Jordan de P sobre \mathbb{C} y a la izquierda ([44], §35). Sean $\lambda_0 = 1, \lambda_1, \dots, \lambda_s$ los autovalores de P , de multiplicidades respectivas $1, r_1, \dots, r_s$. Asociada a estos autovectores, se puede construir una base de Jordan de \mathbb{R}^r , el l -ésimo espacio $\mathcal{E} = \mathcal{E}_l$, asociado a $\lambda = \lambda_l$ (de multiplicidad $r = r_l$) generado por los vectores e_1, e_2, \dots, e_r tales que: $e_1 P = \lambda e_1$ y $e_i P = e_{i-1} + \lambda e_i$ para $i = 2, r$. Para $l = 0$, se tomará π como autovector. Cada espacio \mathcal{E}_l es estable para P . Ahora es fácil verificar que si $u \in \mathcal{E}_l$, $u P^n = \lambda_l^n Q_l(n)$, donde $Q_l(n)$ es un vector de \mathcal{E}_l con coordenadas polinomiales en n de grados $\leq r_l - 1$.

Sea ahora μ_0 una distribución inicial, y $\mu_0 = \alpha\pi + \sum_{l=1,r} v_l$, donde $v_l \in \mathcal{E}_l$, su descomposición en la base de Jordan. Se verifica:

$$\mu_0 P^n = \alpha\pi + \sum_{l=1,s} \lambda_l^n Q_l(n)$$

Como consecuencia del punto (3) del teorema de Perron-Frobenius, $|\lambda_l| < 1$ si $l \geq 1$. Entonces $\mu_0 P^n \xrightarrow{n \rightarrow \infty} \alpha\pi$. Porque $\mu_0 P^n$ es una probabilidad, su límite también, sea $\alpha \sum_i \pi_i = 1 = \alpha$. Así la ergodicidad de P queda demostrada. \square

Comentario : Llamemos $\rho = \rho(P) = \sup\{|\lambda|, \lambda \text{ autovalor de } P, \lambda \neq 1\}$, y $p + 1$ el mayor orden de multiplicidad de los autovalores de módulo ρ . Si P es ergódica, $\rho < 1$, y (??) da:

$$\|\mu_0 P^n - \pi\|_1 \leq C n^p \rho^n \quad (1.1)$$

Para que esta fórmula sea útil, se deben controlar: ρ , esto es el autovalor subdominante; p su orden de multiplicidad; y C que depende de μ_0 y de la expresión explícita de la base de Jordan. Salvo casos muy particulares, el control explícito (1.1) se torna imposible, lo que limita su interés práctico.

Otros controles, más *burdos pero explícitos*, existen. Su debilidad está ligada al hecho de que la acotación geométrica en δ^n , que ellos otorgan, se cumple sólo para un δ muy cercano de 1. Describamos dos de estos controles.

1.2.2. Lema de contracción para P

Proposición 1.3 (Kemeny y Snell, [59]) *Supongamos que P es regular: $\exists k \geq 1$, t.q. $\varepsilon = \inf_{i,j} P_{ij}^k > 0$. Ahora si escribimos $[x]$ la parte entera de x , π la medida invariante de P , y r el cardinal de E , se verifica:*

$$\|\mu_0 P^n - \pi\|_1 \leq r(1 - 2\varepsilon)^{\lfloor \frac{n}{k} \rfloor}$$

Demostración : Será suficiente establecer el resultado para $k = 1$, el caso regular $k \geq 1$ se deducirá trabajando con la matriz de transición $Q = P^k$. Supongamos que $\varepsilon = \inf_{i,j} P_{ij} > 0$. Comencemos por establecer el lema de contracción siguiente,

Lema 1.1 *Sea $x \in \mathbb{R}^r$, $M_0 = \max x_i$, $m_0 = \min x_i$, $M_1 = \max(Px)_i$ y $m_1 = \min(Px)_i$. Entonces P es una contracción en el sentido de que*

$$m_0 \leq m_1 \leq M_1 \leq M_0 \text{ y } 0 \leq (M_1 - m_1) \leq (1 - 2\varepsilon)(M_0 - m_0)$$

Demostración del lema: Sea l un índice t.q. $x_l = m_0$ y $x' \in \mathbb{R}^r$, con la misma coordenada l que x , las otras coordenadas de x' valen M_0 . Puesto que $P \geq 0$,

$$(Px)_i \leq (Px')_i = M_0 - (M_0 - m_0)p_{il} \leq M_0 - (M_0 - m_0)\varepsilon$$

Entonces $M_1 = \sup(Px)_i \leq M_0 - (M_0 - m_0)\varepsilon$. Haciendo el mismo razonamiento para el vector $-x$, se obtiene: $-m_1 \leq -m_0 - (-m_0 + M_0)\varepsilon$. El resultados enunciado se obtiene sumando estas dos desigualdades. \square

Sea ahora $x = {}^t \delta_j$ donde δ_j es la ley inicial concentrada en j . Para M_n y m_n definidas como antes, pero asociadas a P^n , se cumple:

$$m_n \leq P_{ij}^n = (P^n x)_i \leq M_n \text{ y } 0 < \varepsilon \leq m_1 \leq m_n \leq M_n \leq M_1 = 1$$

Las sucesiones (m_n) y (M_n) son adyacentes, de límite común π_j y verifican $m_n \leq \pi_j \leq M_n$. Se deduce que $|\pi_j - P_{ij}^n| \leq M_n - m_n \leq (1 - 2\varepsilon)^n$. \square

Veremos en el capítulo 3 que la utilización del coeficiente de contracción de Dobruschin permite obtener un mejor control:

$$\|\mu_0 P^n - \pi\|_1 \leq 2(1 - r\varepsilon)^n$$

1.2.3. Ergodicidad y acoplamiento de cadenas de Markov

Al contrario de los desarrollos precedentes, que son algebraicos, el que viene será puramente probabilista. Él se basa en la técnica de acoplamiento, concebida aquí en el contexto de las cadenas de Markov (Doebelin, [27]; Aldous, [3]; Lindvall, [68]; Thorisson, [97]). Comencemos por definir la noción general de acoplamiento de medidas. Luego construiremos un acoplamiento particular en el caso de las cadenas de Markov, acoplamiento que nos permitirá controlar la velocidad de la convergencia.

Desigualdad de Acoplamiento.

Sea (F, \mathcal{F}) un espacio medible, Q una medida sobre $(F^2, \mathcal{F}^{\otimes 2})$, μ y ν dos medidas sobre (F, \mathcal{F}) . Decimos que Q es un *acoplamiento* de μ y de ν si μ (resp. ν) es la primera (resp. la segunda) marginal de Q . Hay varias maneras de construir acoplamientos, desde el acoplamiento independiente $Q = \mu \otimes \nu$ hasta los acoplamientos con dependencias. Recordemos que la norma de la variación total para una medida se define por

$$\|\mu\|_{VT} = 2 \sup\{|\mu(A)| \mid A \in \mathcal{F}\} = \sup_{A \in \mathcal{F}} \mu(A) - \inf_{A \in \mathcal{F}} \mu(A)$$

Si $F = E$ es finito, esta norma coincide con la norma l^1 : $\|\mu\|_{VT} = \sum_{i \in F} |\mu_i| = \|\mu\|_1$. Se tiene el resultado fundamental siguiente:

Lema 1.2 *Si Q es un acoplamiento de μ y de ν , y si Δ es la diagonal de F^2 , se verifica:*

$$\|\mu - \nu\|_{VT} \leq 2Q(\Delta^c)$$

Demostración : Se tiene: $\mu(A) - \nu(A) = Q(A \times E) - Q(E \times A) \leq 2Q(\Delta^c)$; igualmente para $\nu(A) - \mu(A)$. Se obtiene entonces la desigualdad anunciada. \square

Se dice que el acoplamiento es maximal si se verifica la igualdad. Existen condiciones que aseguran la existencia y permiten la construcción del acoplamiento maximal.

Un acoplamiento de cadenas de Markov.

Vamos a proponer aquí un primer acoplamiento natural de cadenas de Markov. Otros acoplamientos más eficientes serán estudiados en el capítulo 5. Sea P una matriz de transición sobre E , $X = (X_0, X_1, X_2, \dots)$ y $Y = (Y_0, Y_1, Y_2, \dots)$ dos cadenas independientes con la misma matriz de transición P , $X_0 \sim \mu$ y $Y_0 \sim \nu$. Sea $T = \inf\{n \geq 0 \text{ t.q. } X_n = Y_n\}$ el tiempo en que las dos trayectorias se cruzan por la primera vez. Definamos el acoplamiento siguiente $Q = (X', Y')$:

$$\begin{cases} X' \equiv X \text{ si } t < T, \\ Y'_t = Y_t \text{ y } Y'_t = X_t \text{ en otro caso} \end{cases}$$

A partir del instante T , la trayectoria de Y' no es otra que la de X : T es el *instante de acoplamiento* de las dos cadenas X y Y . Y' es una cadena de Markov de matriz de transición P , $Q = (X', Y')$ es un acoplamiento de (X, Y) . A partir de esta construcción, se obtiene el resultado siguiente:

Proposición 1.4 Supongamos que $\varepsilon = \inf_{ij} P_{ij} > 0$ y denotemos por π la ley invariante de P . Entonces:

- (1) Desigualdad de acoplamiento: $\|\mu P^n - \nu P^n\|_1 \leq 2\mathbb{P}(T > n)$.
- (2) El acoplamiento tiene xito: $\mathbb{P}(T > n) \xrightarrow{n \rightarrow \infty} 0$; se verifica también $\mathbb{P}(T > n) \leq 2(1 - \varepsilon)^n$.
- (3) En particular, para la selección $\nu = \pi$, se tiene: $\|\mu P^n - \pi\|_1 \leq 2(1 - \varepsilon)^n$.

Demostración :

- (1) Se deduce de la desigualdad de acoplamiento porque $(T > n) = (X'_n \neq Y'_n) = \Delta^c$.
- (2) Resta evaluar $\mathbb{P}(T > n)$. Se tiene la descripción:

$$(T > n) = \bigcup_{a \in E^{n+1}} \{X_j = a_j, Y_j \neq a_j, j = 0, n\}$$

Como las variables X_j y Y_j independientes, se verifica:

$$\mathbb{P}(T > n) = \sum_{a \in E^{n+1}} P(X_0 = a_0, \dots, X_n = a_n) \times A(a)$$

con $A(a) = \mathbb{P}(Y_0 \neq a_0, \dots, Y_n \neq a_n) = [1 - \nu_{a_0}] \prod_{j=1, n} \mathbb{P}(Y_j \neq a_j \mid Y_{j-1} \neq a_{j-1})$. Pero, para todo $B \subset E$ no vacío, se cumple:

$$\begin{aligned} \mathbb{P}(Y_j \neq a_j \mid Y_{j-1} \in B) &= \frac{\mathbb{P}(Y_j \neq a_j, Y_{j-1} \in B)}{\mathbb{P}(Y_{j-1} \in B)} \\ &= \frac{\sum_{b \in B} \mathbb{P}(Y_j \neq a_j \mid Y_{j-1} = b) \mathbb{P}(Y_{j-1} = b)}{\mathbb{P}(Y_{j-1} \in B)} \leq (1 - \varepsilon) \end{aligned}$$

porque, en efecto, $\mathbb{P}(Y_j \neq a_j \mid Y_{j-1} = b) \leq (1 - \varepsilon)$. Esto conduce al resultado enunciado.

- (3) Sólo basta recordar que si $Y_0 \sim \pi$, entonces $Y_n \sim \pi$ para todo n . □

Veremos en el capítulo 5 que otros acoplamientos más finos son posibles, y permiten mejorar el control de la velocidad de convergencia.

1.3. Reversibilidad y medida invariante

Las matrices de transición π -reversibles van a jugar un papel crucial en la construcción de algoritmos de simulación.

Decimos que la matriz de transición P sobre E es π -reversible si verifica la *ecuación de balance detallado*:

$$\forall i, j \in E: \pi_i p_{ij} = \pi_j p_{ji} \tag{1.2}$$

La primera consecuencia es que π es una medida invariante para P . En efecto, para todo j : $\sum_i \pi_i p_{ij} = \sum_i \pi_j p_{ji} = \pi_j \sum_i p_{ji} = \pi_j$, es decir que $\pi P = \pi$. También, cada vez que se construya una matriz de transición que verifique (2.3), se sabrá que π es invariante para esta matriz de transición.

El segundo comentario concierne a la distribución de (X_0, X_1) : si la distribución inicial es $X_0 \sim \pi$ y si la transición de la cadena es P , entonces la ley de la pareja (X_0, X_1) es *simétrica*; en efecto:

$$P(X_0 = j, X_1 = i) = \pi_j p_{ji} = \pi_i p_{ij} = P(X_0 = i, X_1 = j)$$

En particular, la transición de X_1 a X_0 es también P

$$P(X_0 = i \mid X_1 = j) = \frac{P(X_0 = i, X_1 = j)}{P(X_1 = j)} = p_{ji}$$

La cadena X , de ley inicial π , es reversible.

1.4. Ergodicidad para un espacio de estados general

Presentaremos aquí los resultados de ergodicidad que parecen ser a la vez los más útiles y los más fáciles de manipular. En el contexto de este curso, se supondrá siempre conocida la medida invariante de la transición (se quiere simular π dada, y se tendrá P construida de tal manera que π sea invariante para P). También el objetivo buscado es muy distinto al del modelaje aleatorio, donde la pregunta inicial es reconocer cuando una cadena con matriz de transición P admite una medida invariante y una solución estacionaria (cf. Meyn-Tweedie [75] o Duflo [29]).

Los resultados relativos al caso de un espacio discreto están expuestos en Feller ([30], tomo 1) e Isaacson y Madsen [53]; los relativos a un espacio de estados general (del tipo $E = \mathbb{R}^p$) han sido estudiados por Tierney ([98], [99]).

La noción de irreducibilidad no cambia para un espacio discreto infinito. Por el contrario, debe ser redefinida para un espacio de estados general. Exceptuando estos cambios, diremos que el núcleo de transición P , π -irreducible, es *periódico* si existe un entero $d \geq 2$ y una sucesión $\{E_1, E_2, \dots, E_d\}$ de d conjuntos no vacíos t.q., para $i = 0, 1, \dots, d-1$ y $x \in E_i$,

$$P(x, E_j) = 1 \text{ para } j = i + 1 \pmod{d}$$

De lo contrario, diremos que P es *aperiódica*.

1.4.1. Espacio de estados discreto

Este es el caso en el cual E es finito o numerable, provisto de la σ -álgebra de todos sus subconjuntos. Consideremos una cadena con matriz de transición P .

Proposición 1.5 ([30], tomo 1, p. 394; [53], Teorema III.2.2)

Sea P una matriz de transición irreducible y aperiódica sobre un espacio discreto E . Entonces las dos proposiciones siguientes son equivalentes:

- (i) Existe una ley invariante π para P .
- (ii) P es ergódica: $\forall \nu, \nu P^n \rightarrow \pi$

Adems, la ley invariante es nica y $\pi > 0$.

La existencia de una ley invariante está garantizada si E es finito, se tienen las equivalencias:

Caso E finito : P ergódica $\Leftrightarrow P$ aperiódica e irreducible $\Leftrightarrow P$ regular

1.4.2. Espacio de estados general

Sea (E, \mathcal{E}) un espacio medible, donde \mathcal{E} admite una familia generatriz numerable (por ejemplo $(\mathbb{R}^p, B(\mathbb{R}^p))$). Un núcleo de transición sobre (E, \mathcal{E}) es una función $P : E \times \mathcal{E} \rightarrow [0, 1]$ tal que:

- (i) Para todo $x \in E$, $P(x, \cdot)$ es una probabilidad en \mathcal{E}
- (ii) Para todo $A \in \mathcal{E}$, $x \mapsto P(x, A)$ es medible.

La transición de la cadena está definida por: $P(X_{n+1} \in A \mid X_n = x) = P(x, A)$.

El núcleo P , de medida invariante π , se dice π -*irreducible* si para todo $A \in \mathcal{E}$ t.q. $\pi(A) > 0$, y todo $x \in E$, existe un entero $n = n(x, A)$ t.q. $P^n(x, A) > 0$. Escribiremos P_x la ley de la cadena que parte de x . Se dice que la cadena es *recurrente* si para todo A t.q. $\pi(A) > 0$, se verifica:

- $P_x(X_n \in A \text{ i.v.}) > 0$ para todo x
- $P_x(X_n \in A \text{ i.v.}) = 1$ π -c.s. en x

La cadena es *Harris-recurrente* si la segunda igualdad se verifica para todo x .

Se tiene entonces el resultado siguiente ([98], Teorema 1), $\|\cdot\|_{VT}$ es la norma de la variación total de una medida ($\|\mu\|_{VT} = 2 \sup\{|\mu(A)|, A \in \mathcal{E}\}$),

Proposición 1.6 ([98],[99]) *Sea P una transición de medida invariante π , π -irreducible. Entonces, P es recurrente y π es la única medida invariante. Si además P es aperiódica, entonces π -c.s. en x :*

$$\|P^n(x, \cdot) - \pi(\cdot)\|_{VT} \xrightarrow{n \rightarrow \infty} 0$$

Tierney establece también un resultado que será útil para el estudio del muestreador de un modelo de Gibbs sobre \mathbb{R}^d ,

Corolario 1.1 ([98]). *Sea P una transición π -irreducible tal que $\pi P = P$. Si para todo x $P(x, \cdot)$ es absolutamente continua con respecto a $\pi(\cdot)$, entonces P es Harris-recurrente.*

1.5. Ejercicios

Ejercicio 1.1 *Cálculo de A^n para una matriz diagonalizable*

Suponga que A es una matriz $r \times r$ diagonalizable, de autovectores por la izquierda x_1, x_2, \dots, x_r , asociados a los autovalores $\lambda_1, \lambda_2, \dots, \lambda_r$. Denotemos por C a la matriz cuyas filas están en el orden x_1, x_2, \dots, x_r , y sea D la matriz diagonal, $D_{ii} = \lambda_i$, $i = 1, r$.

- (i) Verificar que $A = C^{-1}DC$. Deducir la expresión de A^n .
- (ii) Demostrar que si se escribe $C^{-1} = ({}^t y_1, {}^t y_2, \dots, {}^t y_r)$, ${}^t y_i$ es autovector por la derecha de A para λ_i . Deducir que $A^n = \sum_{i=1, r} \lambda_i^n {}^t y_i x_i$.
- (iii) Aplicación: Verificar que la transición P ,

$$P = \begin{pmatrix} 1/2 & 7/24 & 1/6 & 1/24 \\ 1/6 & 11/24 & 1/6 & 5/24 \\ 1/2 & 7/24 & 1/6 & 1/24 \\ 1/6 & 5/24 & 1/6 & 11/24 \end{pmatrix}$$

admite como autovalores/autovectores

Autovalor	Autovector por la izquierda	Autovector por la derecha
1	$\pi = x_1 = (\frac{1}{3}, \frac{1}{3}, \frac{1}{6}, \frac{1}{6})$	$y_1 = (1, 1, 1, 1)$
1/3	$x_2 = (\frac{1}{2}, \frac{1}{2}, 0, -1)$	$y_2 = (1, -1, 1, -1)$
1/4	$x_3 = (0, \frac{1}{2}, 0, -\frac{1}{2})$	$y_3 = (-\frac{5}{3}, \frac{7}{3}, -\frac{5}{3}, \frac{1}{3})$
0	$x_4 = (\frac{1}{2}, 0, -\frac{1}{2}, 0)$	$y_4 = (\frac{1}{3}, \frac{1}{3}, -\frac{5}{3}, \frac{1}{3})$

Dar el valor de P^5 con cuatro decimales. Para la distribución inicial $\mu_0 = (1, 0, 0, 0)$, ¿cuánto difiere $\mu_0 P^{10}$ de π ?

Ejercicio 1.2 *Irreducible y aperiódica \Rightarrow regular*

Demostrar directamente que si P es una matriz de transición finita, irreducible y aperiódica, P es regular.

Ejercicio 1.3 *Transición particular sobre un espacio producto*

Sea una cadena con matriz de transición P sobre $E = \{0, 1, 2, 3, \dots\}$. Se define la matriz de transición R sobre $E \times E$:

$$r^{(i,j),(k,l)} = p_{ik}p_{jl}$$

- (i) Demostrar que $r^{(i,j),(k,l)} = p_{ik}^{(n)} p_{jl}^{(n)}$.
(ii) Demostrar que si P es irreducible y aperiódica, lo mismo vale para R .

Ejercicio 1.4 *Un modelo de meteorología markoviana*

Un régimen meteorológico se describe por los tres estados: s =sol, n =nuboso y p =lluvia, y, por las probabilidades de transición, dadas en este orden, por la matriz:

$$P = \begin{pmatrix} 0,5 & 0,5 & 0 \\ 0,5 & 0,25 & 0,25 \\ 0 & 0,5 & 0,5 \end{pmatrix}$$

¿Es irreducible la cadena? ¿aperiódica? ¿regular? Determinar su medida invariante π . ¿Es π reversible la cadena? Sobre la base de diversos controles de la velocidad de ergodicidad (diagonalización, contracción, coeficiente de contracción (cf. capítulo 3)), evaluar $\|\nu P^n - \pi\|$. Determinar N t.q., para toda distribución inicial ν :

$$\sup_{n \geq N, \nu, j} |\nu P^n(j) - \pi(j)| < 10^{-3}$$

Ejercicio 1.5 *Irreducibilidad y aperiodicidad sobre un espacio numerable*

¿Cuáles son las propiedades de irreducibilidad, aperiodicidad y regularidad, de una matriz de transición P cuya primera fila y primera columna son > 0 ?

Ejercicio 1.6 *Transición que posee “el mal del país”*

(1) Una matriz de transición es doblemente estocástica (D.E.) si la suma de cada una de sus columnas es 1. Verificar que si P es D.E., la distribución uniforme es invariante para P . Si π es la única medida invariante de P , demostrar la equivalencia: P es D.E. y reversible $\Leftrightarrow P$ es simétrica.

(2) Sea el paseo al azar sobre $E = \{1, 2, 3, 4\}$ con barreras reflejantes, de transición

$$P = \begin{pmatrix} 0,5 & 0,5 & 0 & 0 \\ 0,5 & 0 & 0,5 & 0 \\ 0 & 0,5 & 0 & 0,5 \\ 0 & 0 & 0,5 & 0,5 \end{pmatrix}$$

¿es P reversible? ¿es P regular?

(3) Se dice que una matriz de transición tiene “el mal del país” si verifica (\mathcal{R})

$$(\mathcal{R}): \forall i, j, p_{ii} \geq p_{ij}$$

Por otra parte, se dice que una transición es invariante por permutaciones si todo par de filas se cambia por permutación de columnas, $(\mathcal{P}): \forall i \neq l, \exists \sigma$ t.q. $\forall k, p_{l,k} = p_{i,\sigma(k)}$.

(3-1) Verificar que para P definida en (2), P y P^2 verifican (\mathcal{P}) .

(3-2) Demostrar que si P es simétrica y verifica (\mathcal{P}) , P^2 verifica (\mathcal{R}) (usar la desigualdad del producto escalar).

Para P definida en (2), mostrar que P^4 verifica (\mathcal{R}) .

Ejercicio 1.7 *Ergodicidad sobre un espacio numerable*

(i) Sea una transición P sobre $E = \{0, 1, 2, 3, \dots\}$ que verifica

$$\begin{cases} p_{0j} \neq 0 \Leftrightarrow j = 0, 1 \\ \text{Si } i \geq 1: p_{ij} \neq 0 \Leftrightarrow j = i - 1 \text{ o } i + 1 \end{cases}$$

¿Es P irreducible? ¿aperiódica?

(ii) Se especifica P por: $p_{00} = p_{01} = \frac{1}{2}$ y para $i \geq 1$, $p_{i,i-1} = \frac{2^{i+1}-1}{2^{i+1}}$, $p_{i,i+1} = \frac{1}{2^{i+1}}$. Determinar la probabilidad invariante π de P si ella existe. ¿ P es ergódica?

Ejercicio 1.8 *Cadenas de Markov, inversin del tiempo y reversibilidad*

Sea $X = (X_0, X_1, \dots, X_n)$ una cadena de Markov sobre E^n (E finito y n fijo), con distribución inicial μ_0 , de transiciones (no-homogéneas) P_k , $k = 0, n-1$. Denotemos μ_k la ley de X_k , $k = 0, n$.

(i) Demostrar que, después del retorno del tiempo, y si $X_n \sim \mu_n$, el proceso retornado $X^r = (X_n, X_{n-1}, \dots, X_0)$ verifica la propiedad de Markov, y que es una cadena de Markov de transiciones $(Q_k, k = n, 1)$:

$$Q_k(i_k, i_{k-1}) = P(X_{k-1} = i_{k-1} \mid X_k = i_k) = \frac{\mu_{k-1}(i_{k-1})}{\mu_k(i_k)} P_{k-1}(i_{k-1}, i_k)$$

(ii) Precisar Q si la cadena es homogénea, de transición P , la distribución inicial es la medida invariante π . ¿Bajo que condición se verifica $Q(j, i) = P(i, j)$?

Ejercicio 1.9 *Cadenas de Markov y campos de Markov*

Nos colocamos en el contexto del ejercicio anterior.

(i) Demostrar que $X = (X_0, X_1, \dots, X_n)$ posee la propiedad de Markov bilateral siguiente (hablamos de un campo de Markov): para todo k , $0 < k < n$:

$$P(X_k = i_k \mid X_l = i_l, 0 \leq l \leq n \text{ y } l \neq k) = B_k(i_k \mid i_{k-1}, i_{k+1}) = \frac{P_{k-1}(i_{k-1}, i_k) P_k(i_k, i_{k+1})}{[P_{k-1} P_k](i_{k-1}, i_{k+1})}$$

(ii) Recíprocamente, si X posee esta propiedad de Markov bilateral, más las dos propiedades de Markov naturales “de borde”, mostrar que X es una cadena de Markov (sobre un subintervalo de \mathbb{N} , hay entonces equivalencia entre la propiedad de Markov causal (cadena) y no-causal (campo); esta equivalencia nos llevará sobre \mathbb{N}^2).

(iii) Consideramos la cadena homogénea y estacionaria sobre $E = \{-1, +1\}$ con matriz de transición $p = P(X_n = 1 \mid X_{n-1} = 1)$ y $q = P(X_n = -1 \mid X_{n-1} = -1)$, $0 < p, q < 1$. Determinar la transición bilateral $Q(y \mid x, z)$, para x, y y z en E . Demostrar que esta matriz de transición puede escribirse

$$Q(y \mid x, z) = Z^{-1}(x, y) \exp\{x(h + \beta(y + z))\}$$

donde $h = \frac{1}{2} \log \frac{p}{q}$ y $\beta = \frac{1}{4} \log \frac{pq}{(1-p)(1-q)}$ ($Z^{-1}(x, y)$ es la constante de normalización que hace de Q una probabilidad en x). Interpretar las situaciones: $h = 0$ ($p = q$); $\beta = 0$ ($p + q = 1$).

Ejercicio 1.10 *Cadena de Markov con rgimen markoviano y cadena de Markov escondida*

Sean P_1, P_2 dos transiciones regulares sobre $E = \{a, b, c\}$ con medidas invariantes π_1, π_2 . Sea $X = (X_n)$ un proceso sobre E definido por:

• $I = (I_n), I_n \in \{1, 2\}$ es una cadena de Markov homogénea cuya transición está determinada por: $P(I_n \neq I_{n-1} | I_{n-1}) = p$. I es el *régimen* de X ,

•• Si sabemos que $I_n, ((X_n | I_n), n \geq 0)$ es una cadena de Markov inhomogénea de transiciones (P_{I_n}) ($(X | I)$ es entonces una cadena de Markov inhomogénea).

(1) Demostrar que $(X, I) = (X_n, I_n)_n$ es una cadena de Markov de la que se determinará la transición P . Verificar que si $0 < p < 1$, P es regular.

(2) Verificar que $X = (X_n)$ no es una cadena de Markov.

Ejercicio 1.11 *Procesos de Markov de memoria 2*

Sea el proceso de Markov sobre $\{0, 1\}$, de memoria dos, y de transición:

$$p_{ab,c} = P(X_{n+2} = c | X_n = a, X_{n+1} = b), a, b, c \in \{0, 1\}$$

Demostrar que $Y_n = (X_n, X_{n+1})$ es una cadena de Markov. Determinar su transición P . Si cada $p_{ab,c} > 0$, verificar que $P^2 > 0$ y entonces que P es regular. ¿Como se generalizan estos resultados al caso de una memoria de orden p ?

Ejercicio 1.12 *Prdida de la propiedad de Markov por agregacin de estados*

Sea X una cadena de Markov con transición P sobre $E = \{a, b, c\}$. Sea Y el proceso asociado con dos estados $\{\{a, b\}, c\}$ (se agregan los dos estados a y b):

$$Y = \{a, b\} \text{ si } X \in \{a, b\}, Y = c \text{ si } X = c$$

Demostrar que en general, Y no es una cadena de Markov. (ver que $P(Y_2 = c | Y_0 = x_0, Y_1 = \{a, b\})$ depende en general de x_0 para $x_0 = c$).

Los ejercicios siguientes se refieren a la simulación de variables aleatorias de base. Sobre esta materia, se podrá consultar Devroye [19], Jonhson y Kotz [54] y Robert [88].

Ejercicio 1.13 *Simulacin de una variable de funcin de distribucin F conocida*

Sea X una v.a. de función de distribución F conocida, y

$$F^-(u) = \inf\{x \in R; F(x) \geq u\}$$

($F^- = F^{-1}$, la función inversa de F , si F es continua). Demostrar que si U es la distribución uniforme sobre $[0, 1]$, $F^-(U)$ tiene distribución F .

Aplicaciones: obtener la simulación de distribuciones: exponencial; exponencial simétrica; Cauchy; Pareto; Gamma $\Gamma(\kappa, \lambda)$; en el caso en que el índice κ es entero.

Ejercicio 1.14 *Simulacin de una distribucin discreta*

(1) Sean p_1, p_2, \dots, p_k , k probabilidades de suma 1, y $x_i = \sum_{j=1, i} p_j$, $i = 1, k$. Verificar que si U es uniforme sobre $[0, 1]$, $P(x_i \leq U < x_{i+1}) = p_i$.

(2) Deducir un método de simulación de la distribución discreta $P(X = a_i) = p_i, i = 1, k$.

(3) *Aplicaciones.* Simular las distribuciones siguientes: Bernoulli $B(p)$; binomial $B(n, p)$ (proponer dos métodos); geométrica $G(p)$ (proponer dos métodos); Poisson $P(\lambda)$.

Ejercicio 1.15 *Simulación de la distribución gaussiana y distribuciones conexas*

(1) Justificar el hecho de que $X = \sum_{i=1,12}(U_i - 0,5)$ es aproximadamente una ley normal estándar si las U_i son i.i.d. uniformes sobre $[0, 1]$.

(2) *Método de Box-Muller*: verificar analíticamente que si U_1 y U_2 son uniformes sobre $[0, 1]$ e independientes, entonces las dos variables X_1 y X_2 definidas abajo son (exactamente) $\mathcal{N}(0, 1)$ e independientes:

$$X_1 = \sqrt{-2 \log U_1} \cos(2\pi U_2), X_2 = \sqrt{-2 \log U_1} \sin(2\pi U_2)$$

(3) Deducir la simulación de las distribuciones $\chi_n^2, T_n, F_{n,m}$

Ejercicio 1.16 *Generación de una ley de Poisson, proceso de Poisson y distribuciones exponenciales*

(1) Demostrar que si X_1, X_2, \dots son distribuciones exponenciales i.i.d. de parámetro λ , y si N es una distribución de Poisson de parámetro λ ,

$$P(N = k) = P(X_1 + \dots + X_k \leq 1 < X_1 + \dots + X_k + X_{k+1})$$

Deducir un procedimiento de simulación de la distribución de Poisson.

Denotando $S_k = \sum_{j \leq k} X_j$, $S_0 = 0$, el *proceso de Poisson* de intensidad λ sobre \mathbb{R}^+ está dado por la sucesión $\mathcal{P} = \{S_k, k \geq 0\}$.

(2) *Proceso de Poisson borrado*: Demostrar que si borramos de manera i.i.d. con probabilidad p (e independiente de \mathcal{P}) los puntos del proceso \mathcal{P} , se obtiene un proceso de Poisson de intensidad λp cuyos intervalos de separación constituyen una sucesión i.i.d. de exponenciales de parámetro λp .

Ejercicio 1.17 *Método de simulación por aceptación-rechazo*

(1) Queremos simular X una v.a. real de densidad f concentrada en $[0, 1]$.

Sea g la densidad sobre $[0, 1]$ de una *distribución fácil de simular* (por ejemplo la uniforme), y M una constante tal que sobre $[0, 1]$: $f(x) \leq Mg(x)$. Definimos el algoritmo siguiente:

- 1-Simular z siguiendo g e, independientemente, u siguiendo la ley $U([0, 1])$
- 2-Aceptar $x=z$ si $Mg(z) \times u \leq f(z)$
- 3-Si no repetir 1.

Demostrar que x tiene una distribución con densidad f .

(2) ¿Cómo generalizar este resultado si X está concentrada sobre un intervalo (α, β) , intervalo eventualmente de longitud infinita ?

(3) *Aplicaciones*: construir un simulador de

→ la ley $\Gamma(\lambda, \kappa)$, $\kappa \geq 1$, de densidad, $f(x) = \Gamma(\kappa)^{-1} \lambda \kappa e^{-\lambda x} x^{\kappa-1} \mathbf{1}(x > 0)$

→ la ley $\beta(a, b)$, $a \geq 1$ y $b \geq 1$, de densidad $\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)^{b-1} \mathbf{1}(x \in]0, 1[)$

Capítulo 2

Simulación usando Cadenas de Markov

2.1. El problema y el método

Sea $E = \{1, 2, \dots, r\}$ un espacio de estados finito, y π una distribución de probabilidad sobre E : ¿Cómo obtener un valor de la variable X con distribución π ?

Si r , el cardinal de E , no es muy grande (del orden del millar), se simula π por el método clásico : se forma la sucesión creciente en $[0, 1]$ que define la función de distribución de $X \sim \pi$,

(i) $F_0 = 0$ y para $0 < i \leq r$, $F_i = \sum_{j \leq i} \pi_j$;

(ii) se simula una distribución uniforme U en $[0, 1]$, y

(iii) se guarda el valor $x = x(U)$ si $F_{x-1} < U \leq F_x$

A pesar de que este método de simulación es directo y exacto, él es inaplicable cuando el cardinal de E es muy grande. Por ejemplo para el espacio producto $E = \{0, 1\}^{10 \times 10}$, ($x \in E$ es el valor de una variable de presencia o ausencia en la malla cuadrada $\{1, 2, \dots, 10\}^2$ hasta 100 puntos), $r = 2^{100} \simeq 1,27 \times 10^{30}$: no hay manera de poder colocar en la memoria $1,27 \times 10^{30}$ valores F_j . Notemos que este ejemplo es de “tamaño pequeño”: los valores son binarios y el tamaño de la ventana es pequeño. Un problema real en imágenes hará intervenir un valor de $r = 256^{512 \times 512}$ (256 niveles de gris, imagen con 512×512 pixeles).

El método de simulación usando cadenas de Markov.

La idea es simular aproximadamente π como una distribución límite de una cadena de Markov ergódica X de transición P . Esto implica :

(1) proponer una transición P sobre E tal que $\pi P = \pi$ (P es π invariante).

(2) asegurar la ergodicidad de la cadena : $\nu P^n \rightarrow \pi$.

(3) saber a partir de que valor n_0 se puede decir que para $n \geq n_0$, νP^n está cercana de π .

Así, para una distribución inicial ν arbitraria, y si n_0 es bastante grande, νP^{n_0} genera aproximadamente a π . Si se quiere una muestra aproximada de π , se recommienza independientemente esta operación (se tienen entonces cadenas independientes); o más aun, sobre la misma cadena, se espacian con un valor K grande los valores sucesivos, νP^{n_0} , νP^{n_0+K} , νP^{n_0+2K} , y se sigue así de manera recursiva.

Muchas selecciones de P son posibles : desde el punto de vista puramente paramétrico, P depende de $r(r-1)$ parámetros, y la invarianza de π impone $r-1$ restricciones de igualdad : el conjunto de transiciones ergódicas que convergen a π contiene un abierto no vacío de \mathbb{R} ; se dispone entonces de $(r-1)^2$ grados de libertad en la selección de P .

En relación a los puntos (1) y (2), la selección de P debe responder a la *comodidad de la implementación práctica* y a una *buena eficiencia algorítmica*. A este nivel, el *conocimiento práctico* juega un papel importante.

El punto (3) es difícil: ¿cuando se podría aceptar que la cadena que comienza con una distribución ν *entró en su régimen estacionario*? Como se ha visto en el primer capítulo, el control teórico sobre la velocidad de ergodicidad es inoperante en la práctica en la mayoría de los ejemplos reales. Presentaremos en el capítulo 5 reglas empíricas que permiten probar si la cadena ha entrado dentro de su régimen estacionario ([86]). Aquí de nuevo, el conocimiento previo y la experiencia pueden ser más útiles que los resultados teóricos no explícitos, o cuando, a pesar de ser explícitos, son ineficaces puesto que otorgan una mayoración en ρ^n para un valor de ρ muy cercano de 1.

Este punto (3) y la dificultad de responderlo están en el origen de muchos desarrollos teóricos: (i) el estudio del fenómeno del *cutoff* para una cadena de Markov ([24], [91], [108]; cf. capítulo 5): la convergencia hacia π se hace de manera abrupta alrededor de un instante de corte (cutoff). Antes de este instante, $\|\nu P^n - \pi\|$ permanece grande, haciéndose esta norma exponencialmente pequeña después.

(ii) la *simulación exacta por acoplamiento desde el pasado* ([84], [61], [78]; cf. capítulo 5) : la simulación se hace desde un tiempo pasado $-T$ y el procedimiento permite asegurar que X_0 sigue *exactamente* la distribución π si T es bastante grande.

Antes de pasar a dos grandes familias de procedimientos de simulación (Gibbs, Métropolis), resumamos las propiedades que debe verificar la transición P :

(1) P es π -invariante (2) P es irreducible y aperiodica Entonces, $\forall \nu, \nu P^n \rightarrow \pi$

Una manera simple de verificar (1) es construir P que sea π -reversible,

(1') P es π -reversible : $\forall i, j \in E, \pi_i p_{ij} = \pi_j p_{ji}$

En cuanto a (2) ella es equivalente al hecho de que P sea regular,

(2') : $\exists k \geq 1$ tal que $P^k > 0$

Notemos desde ahora que en los procedimientos de construcción de P que vamos a desarrollar, es suficiente *conocer π salvo un factor multiplicativo*, es decir conocer para todo $i \in E$ la forma $\pi_i = c e_i$, donde los e_i son explícitos, pero no la constante de normalización $c = (\sum_i e_i)^{-1}$.

2.2. El muestreador de Gibbs

La construcción del *muestreador de Gibbs* (también llamado dinámica de Glauber) *no es posible sino para un espacio de estados producto*, $E = F_1 \times F_2 \times \dots \times F_n$ o potencia $E = F^n$. Este método de simulación fue introducido por los hermanos Geman ([37], [39]) con el fin de resolver problemas de reconstrucción en el tratamiento de imágenes. Para simplificar la presentación, supondremos que $E = F^n$ es un espacio potencia finito. Nada cambia fundamentalmente para un espacio producto general $E = \prod_{i=1, n} F_i$ de componentes F_i finitas.

Algunas notaciones : se define $S = \{1, 2, \dots, n\}$ como el espacio de *sitios*; $E = F^S = F^n$; F el espacio de estados en cada sitio; $x = (x_i, i \in E)$ una *configuración* sobre el conjunto de sitios (es un elemento de E); para un subconjunto no vacío $A \subset S$ del conjunto de sitios,

$x_A = (x_i, i \in A)$ y si $A \neq S$, $x^A = (x_i, i \notin A)$ (para simplificar, se escribirá x_i por $x_{\{i\}}$ y x^i por $x^{\{i\}}$). Finalmente, se denota $\pi_i(\cdot | x^i)$ la distribución condicional a x^i en el sitio i :

$$\pi_i(x_i | x^i) = \frac{\pi(x_i, x^i)}{\pi^i(x^i)}$$

donde $\pi^i(x^i) = \sum_{a_i \in F} \pi(a_i, x^i)$ es la distribución marginal sobre $S \setminus \{i\}$. Se supondrán conocidas explícitamente estas distribuciones condicionales. De esta manera, el contexto en el cual el muestreador de Gibbs puede ser construido es el siguiente:

Contexto requerido para el Muestreador de Gibbs :

- (1) $E = F^n$ (o esp. producto)
 (2) $\forall i \in S$, las distribuciones $\pi_i(\cdot | x^i)$ son conocidas

2.2.1. Muestreador de Gibbs

Consideremos la sucesión de visitas $1 \mapsto 2 \mapsto \dots \mapsto (n-1) \mapsto n$ correspondientes a un *barrido* del conjunto de sitios S . Sea $x = (x_i)$ una configuración inicial, $y = (y_i)$ una configuración final. La transición P del muestreador de Gibbs para este barrido que hace pasar de x a y se define por :

$$P(x, y) = \prod_{i=1, n} \pi_i(y_i | y_1, y_2, \dots, y_{i-1}, x_{i+1}, x_{i+2}, \dots, x_n)$$

En el i -ésimo paso de este barrido, se *cambia* el valor x_i en el sitio i por y_i siguiendo la distribución condicional $\pi_i(\cdot | y_1, \dots, y_{i-1}, x_{i+1}, \dots, x_n)$: en el condicionamiento, los $(i-1)$ primeros valores x han sido ya cambiados, mientras que los $(n-i-1)$ últimos no lo han sido todavía. Este algoritmo es *secuencial* o *asíncrono*. Se distingue fundamentalmente de los *algoritmos simultáneos, síncronos, o paralelos* que estudiaremos más adelante.

Proposición 2.1 *Supongamos que $\pi > 0$. Entonces π es invariante para P y $P > 0$. Ms precisamente, para $\varepsilon = \inf_{i,x} \pi_i(x_i | x^i)$, se tiene, para todo $x, y : P(x, y) \geq \delta = \varepsilon^n > 0$. En particular, para toda distribución inicial ν , $\nu P^k \xrightarrow[k \rightarrow \infty]{} \pi$.*

Demostración :

Invarianza de π : para mostrar que π es P -invariante, es suficiente constatar que $P = \prod_{i \in S} P_i$ y que π es P_i -invariante, P_i correspondiente al i -ésimo paso del barrido. En efecto, si dos transiciones P y Q admiten π como medida invariante, la transición compuesta PQ admite igualmente π como medida invariante porque $\pi(PQ) = (\pi P)Q = \pi Q = \pi$. Mostremos entonces que π es invariante para P_i . P_i hace pasar de un estado u a un estado v donde sólo la coordenada i ha sido cambiada, y esto con probabilidad $\pi_i(v_i | u^i)$,

$$P_i(u, v) = \begin{cases} \pi_i(v_i | u^i) & \text{si } u^i = v^i \\ 0 & \text{si no} \end{cases}$$

Con v fijo, solamente las configuraciones u donde $u^i = v^i$ pueden llevar a v :

$$\sum_u \pi(u) P_i(u, v) = \sum_{u_i} \pi(u_i, v^i) \pi_i(v_i | v^i) = \sum_{u_i} \pi(v_i, v^i) \pi_i(u_i | v^i) = \pi(v)$$

π es invariante para P_i .

Regularidad de P : para cada x , $\pi(x) > 0$. Así, $\varepsilon = \inf_{i \in S, x \in E} \pi_i(x_i | x^i) > 0$: para todo $x, y \in E$, $P(x, y) \geq \delta = \varepsilon^n > 0$. \square

Comentarios.

(i) Es suficiente que π sea conocida salvo un factor para así construir P : si $\pi(x) = ce(x)$, la probabilidad condicional $\pi_i(\cdot | \cdot)$ no depende de c .

(ii) P no es π -reversible a pesar de que cada P_i lo sea (una composición de transiciones π -reversibles no es necesariamente π -reversible, cf. ejercicio).

(iii) Todo otro barrido de S , $\sigma(1) \mapsto \sigma(2) \mapsto \dots \mapsto \sigma(n)$ visitando todos los sitios (σ es entonces una permutación de $\{1, 2, \dots, n\}$) da el mismo resultado. Es fácil ver que se obtiene el mismo resultado de ergodicidad para un barrido de longitud N , visitando todos los sitios, ciertos sitios pueden ser visitados varias veces.

(iv) Se demostrará en el capítulo 3 que se pueden hacer barridos sucesivos que cambian en el curso del tiempo, la transición en el k -ésimo barrido siendo $P(k)$: una condición suficiente de ergodicidad de la cadena inhomogénea ($P(k)$) es que el k -ésimo barrido satisfaga (iii), al estar el número $N(k)$ uniformemente acotado en k .

(v) En la práctica, parece razonable iterar en promedio 100 veces los barridos para simular π .

Ejemplo 2.1 Simulación del modelo de Ising

Consideremos el conjunto de sitios constituidos por la malla cuadrada $S = \{1, 2, \dots, n\}^2$, $F = \{-1, +1\}$, $E = F^S$. El modelo de Ising con interacción de los 4 vecinos más cercanos (denotado v.m.c.) se define de la manera siguiente : notemos por $x \in E$, $U(x) = h \sum_{i \in S} x_i + \beta \sum_{\langle i, j \rangle} x_i x_j$, donde h, β son dos parámetros reales, y $\langle \cdot, \cdot \rangle$ es la relación de vecindad de los 4-v.m.c. : para $i, j \in S$, $\langle i, j \rangle \Leftrightarrow \|i - j\|_1 = 1$. La distribución π de energía U se define por

$$\pi(x) = Z^{-1} \exp\left\{h \sum_{i \in S} x_i + \beta \sum_{\langle i, j \rangle} x_i x_j\right\} \quad (2.1)$$

donde $Z = Z(h, \beta) = \sum_{y \in E} \exp U(y)$ es la constante de normalización que hace de π una probabilidad. La condición de positividad de π se verifica y es fácil ver que la distribución condicional en i , depende solamente de los vecinos $\partial i = \{j \in S : \langle i, j \rangle\}$ de i por intermedio de $v_i = \sum_{j \in \partial i} x_j$ (cf. capítulo 4 sobre los campos de Markov) :

$$\pi_i(x_i | x^i) = \pi_i(x_i | x_{\partial i}) = Z^{-1}(h, \beta, v_i) \exp\{x_i(h + \beta v_i)\}$$

con $Z_i(h, \beta, v_i) = 2 \cosh(h + \beta v_i)$. Notemos que a pesar de que la constante global Z es incalculable (e inútil en la construcción del algoritmo), las constantes Z_i son explícitas (e indispensables para la construcción del algoritmo).

La situación $h = 0$ da una distribución donde todas las marginales X_i son equidistribuidas sobre $F = \{-1, +1\}$. Si además $\beta > 0$, hay correlación espacial, que crece a medida que β se hace grande. En este caso, se puede tomar $\varepsilon = (1 + e^{8\beta})^{-1}$; para $\beta = 1$, $\varepsilon \simeq 3,4 \times 10^{-4}$; se entiende que la positividad de esta constante es esencial para garantizar la ergodicidad, ella, sin embargo, es inoperante para tener un control real de la velocidad de ergodicidad.

Ejemplo 2.2 Textura de niveles de gris y el Φ -modelo

Más generalmente, el muestreador de Gibbs permite la simulación de una gran clase de densidades asociadas a un modelo de Gibbs sobre $E = F^N$ donde $F \subset \mathbb{R}$, por ejemplo $F = \{0, 1, 2, \dots, 255\}$

$$\pi(x) = Z^{-1} \exp U(x)$$

En análisis de imágenes, F es un conjunto de niveles de gris; un modelaje clásico de textura de niveles de gris es el siguiente ([40], [39]) :

$$U(x) = \theta V_d(x) \text{ donde } V_d(x) = \sum_{\langle i,j \rangle} \Phi_d(x_i - x_j), \Phi_d(u) = \frac{1}{1 + (\frac{u}{d})^2}$$

La ventaja de estos Φ -modelos con respecto a los modelos gaussianos es que permiten los contrastes $|x_i - x_j|$, importantes en sitios vecinos, allí donde un modelo gaussiano no lo permite (porque le dan un peso $\exp -(x_i - x_j)^2$ muy débil). θ y d son dos parámetros que controlan el tipo de textura: los contrastes más importantes se permiten si d aumenta, θ controla la correlación espacial. Para este modelo, la distribución condicional en el sitio i es

$$\pi_i(x_i | x_{\partial i}) = Z^{-1}(x_{\partial i}) \exp\{\theta \sum_{j:\langle i,j \rangle} \Phi_d(x_i - x_j)\}$$

con $Z(x_{\partial i}) = \sum_{x_i \in F} \exp\{\theta \sum_{j:\langle i,j \rangle} \Phi_d(x_i - x_j)\}$

2.2.2. Muestreador de Gibbs con barrido aleatorio

Sea ν una probabilidad sobre S . El muestreador de Gibbs con *barrido aleatorio* corresponde al algoritmo siguiente : sea x el estado inicial,

- (1) se selecciona un sitio i , independientemente del pasado, usando ν
- (2) se cambia el valor x_i en este sitio proponiendo y_i con la distribución $\pi_i(\cdot | x^i)$

La nueva configuración es $y = (y_i, x^i)$. La transición Q asociada se concentra en los cambios en a lo sumo un sitio:

$$Q(x, y) = \sum_{i=1, n} \nu_i \mathbf{1}(x^i = y^i) \pi_i(y_i | x^i)$$

Proposición 2.2 *El muestreador de Gibbs aleatorio es ergdico y converge a π siempre que ν y π son > 0 .*

Demostración : Como en la demostración precedente, es fácil ver que π es invariante para Q . Demostremos que $Q^n > 0$. Sea $\varepsilon = \inf_{i,x} \pi_i(x_i | x^i), \varepsilon > 0$. Sean $x, y \in E$ arbitrarios. En el cálculo de la transición $Q^n(x, y)$, existe una selección de barrido aleatorio que visita sucesivamente todos los sitios $1 \mapsto 2 \mapsto 3 \mapsto \dots \mapsto n$, esto con la probabilidad $\nu_1 \nu_2 \dots \nu_n > 0$, el i -ésimo cambio que sustituye a x_i por y_i . Notemos $\Delta = \nu_1 \nu_2 \dots \nu_n$. Se obtiene entonces la minoración :

$$Q^n(x, y) \geq \Delta \prod_{i=1, n} \pi_i(y_i | y_1, y_2, \dots, y_{i-1}, x_{i+1}, x_{i+2}, \dots, x_n) \geq \Delta \varepsilon^n > 0$$

De hecho, la minoración puede ser mejorada porque el argumento es cierto para todo barrido definido por una permutación de $\{1, 2, \dots, n\}$. Se obtiene entonces $Q^n(x, y) \geq \Delta n! \varepsilon^n$; por ejemplo, para un barrido aleatorio uniforme de S , $Q(x, y) \geq n! (\frac{\varepsilon}{n})^n$. \square

Comparación de estrategias para barridos.

Amit y Grenander [4] comparan las estrategias de barrido para el caso específico gaussiano (ver más adelante el muestreador de una distribución continua sobre \mathbb{R}^p) y para el criterio de la velocidad de ergodicidad $\nu P^n \rightarrow \pi$. Este contexto permite obtener un control de los autovalores de la transición de un barrido a partir de la covarianza de la variable y de la matrix de visita de los sitios. Su conclusión es la siguiente: (1) el barrido aleatorio parece preferible al barrido sistemático; (2) existen malos barridos sistemáticos. Esto no permite sin embargo rechazar los barridos sistemáticos periódicos: más simples desde un punto de vista algorítmico, son probablemente eficientes si los sitios sucesivos visitados están alejados (lo que hace, en promedio, un barrido aleatorio).

2.2.3. Muestreador con barridos síncronos

Una manera de acelerar la simulación es la de efectuar los n cambios simultáneamente y de manera independiente: hablamos entonces de cambios *síncronos* o *simultáneos*. La transición $R(x, y)$ está dada por:

$$R(x, y) = \prod_{i \in S} \pi_i(y_i | x^i)$$

Recalquemos la diferencia con el muestreador de Gibbs secuencial: aquí el condicionamiento en el sitio i es en $(x_1, x_2, \dots, x_{i-1}, x_{i+1}, x_{i+2}, \dots, x_n)$ mientras que en el algoritmo secuencial lo era en $(y_1, y_2, \dots, y_{i-1}, x_{i+1}, x_{i+2}, \dots, x_n)$.

Para que estos cambios sean efectuados simultáneamente, es necesario disponer de una máquina con arquitectura paralela: un microprocesador se vincula a cada sitio i , $i \in S$, estos procesadores realizan simultáneamente e independientemente los cambios $x_i \mapsto y_i$ siguiendo las distribuciones $\pi_i(\cdot | x^i)$.

Desafortunadamente, a pesar de que R es ergódica, π no es en general invariante para R y este método por paralelización total no permite simular π . Se tiene el resultado siguiente :

Proposición 2.3 Muestreador con barridos sncronos.

La transición R es regular. Existe entonces una distribución μ tal que para toda distribución inicial ν , $\nu R^n \rightarrow \mu$. En general, π no es invariante para R , y entonces $\mu \neq \pi$. μ se llama distribución virtualmente asociada a π .

Demostración : Para $\varepsilon = \inf_{i,x} \pi_i(x_i | x^i)$, $\inf_{x,y} R(x, y) \geq \varepsilon^n$: R es regular. Por otra parte, es fácil ver que π no es en general invariante para R (cf. ejercicio). \square

En general, no se sabe dar una fórmula explícita analítica de μ : las relaciones entre μ y la distribución de origen π no son conocidas. En uno de los ejercicios, se estudia un caso particular del modelo de Ising (2.1) donde μ puede ser dada de manera explícita: se constata que π y μ son muy disímiles. Por ejemplo, si los vecinos de i para π son los 4-v.m.c., aquellos de μ son los 4 vecinos diagonales. Esto lleva a pensar que distribución virtual y la distribución de origen son “bastante diferentes”.

Simulación de π por paralelización parcial.

Se pueden proponer algoritmos parcialmente paralelos para la simulación de π . Para precisar esto, retomemos el contexto del ejemplo (2.1) del modelo de Ising.

Sobre $S = \{1, 2, \dots, n\}^2$, n par, consideremos la partición en dos colores : $S = B \cup N$, $B = \{i = (i_1, i_2) \text{ t.q. } i_1 + i_2 \text{ es par}\}$, $N = S \setminus B$ (en el contexto donde S es un tablero de ajedrez $n \times n$, B es el conjunto de $\frac{n^2}{2}$ cuadrados blancos, N es el conjunto de $\frac{n^2}{2}$ los cuadrados negros). Es fácil ver que la distribución de $(X_B | x_N)$ (resp. de $(X_N | x_B)$) es $P_B(x_B | x_N) = \prod_{i \in B} \pi_i(x_i | x_{\partial i})$ (resp. $P_N(x_N | x_B) = \prod_{i \in N} \pi_i(x_i | x_{\partial i})$).

Consideremos entonces el muestreador de Gibbs para el barrido siguiente de S : se comienza por visitar los $\frac{n^2}{2}$ sitios de N , después se termina por la visita de los $\frac{n^2}{2}$ sitios de B . Notemos P la transición de este muestreador; puesto que los cambios de sitios de N (resp de B) no hacen intervenir sino a x_B (resp. y_N), se tiene:

$$P(x, y) = P_N(y_N | x_B)P_B(y_B | y_N)$$

Este muestreador logra hacer bien la simulación de π . Si se dispone de una máquina paralela, este algoritmo se realiza en dos iteraciones, cada una efectúa $\frac{n^2}{2}$ cambios síncronos. Se tiene así de un algoritmo de simulación de π parcialmente paralelo.

Es fácil generalizar esta construcción para un conjunto S provisto de un grafo de vecindad para el cual es posible obtener una *partición de S en k colores*, con las propiedades de independencia antes señaladas. Si tal es el caso, se ha construido un algoritmo parcial paralelo que trabaja en k iteraciones.

2.2.4. Muestreador de Gibbs sobre $E = \mathbb{R}^n$ (o $(\mathbb{R}^d)^n$)

Los resultados de Tierney [98] presentados en el primer capítulo se ponen fácilmente en práctica en la situación siguiente, la continuidad absoluta es definida con respecto a la medida de Lebesgue sobre $E = (\mathbb{R}^d)^n$:

- π es una distribución con densidad $\pi(x) > 0$.
- la transición $P(x, y)$ admite una densidad $p(x, y) > 0$ para todo $x, y \in E$.
- π es la medida invariante para P .

En este caso, la cadena de transición P es π -irreducible, aperiódica y Harris recurrente. Se deduce que, para la norma en variación total y para toda distribución inicial ν , $\nu P^n \rightarrow \pi$.

Ejemplo 2.3 Simulación de un vector gaussiano

Supongamos que $F = \mathbb{R}$, la distribución de $X = (X_1, X_2, \dots, X_n)$ es gaussiana de media 0 y de matrix de covarianza Σ , invertible. Si la media es m , será suficiente simular una gaussiana de media 0 y sumarle luego m .

El muestreador de Gibbs de X consiste en visitar secuencialmente los sitios $1 \mapsto 2 \mapsto \dots \mapsto n$, el i -ésimo cambio se hace siguiendo la distribución condicional $(X_i | y_1, \dots, y_{i-1}, x_{i+1}, \dots, x_n)$. Esta distribución condicional es gaussiana, pudiendo calcularla de manera explícita cuando $Q = \Sigma^{-1}$ está dada en forma explícita. La transición $P(x, y)$ admite una densidad, positiva en todas partes, así como la densidad de X . El muestreador de Gibbs provee de una buena simulación de π .

Recordemos que la distribución condicional en el sitio i está dada en función de los coeficientes (q_{ij}) de Q por :

$$\mathcal{L}(X_i | x^i) \sim \mathcal{N}_1(-q_{ii}^{-1} \sum_{j:j \neq i} q_{ij}x_j; q_{ii}^{-1})$$

En el caso de un *campo gaussiano y markoviano* (cf. capítulo 4), $Q = \Sigma^{-1}$ tiene una forma simplificada, $\pi_i(x_i | x^i)$ dependiente de x^i por medio de $x_{\partial i}$, ∂i la vecindad de i para el grafo markoviano. En particular, $q_{ij} \neq 0 \Leftrightarrow j \in \partial i$. Coloquémonos en $S = \{1, 2, \dots, n\}^2$ y consideremos la variable gaussiana X sobre S de densidad :

$$\pi(x) = Z^{-1} \exp U(x) \text{ où } U(x) = -\frac{1}{2} {}^t x Q x, \text{ con } {}^t x Q x = a \sum_{i \in S} x_i^2 + b \sum_{\|i-j\|_1=1} x_i x_j$$

Si $|\frac{b}{a}| < \frac{1}{2}$, Q es definida positiva y

$$\mathcal{L}(X_i | x^i) \sim \mathcal{N}_1\left(-\frac{b}{a} \sum_{j: \|i-j\|_1=1} x_j; \frac{1}{a}\right).$$

Este método de simulación puede ser comparado al método clásico. Sea $\Sigma = T^t T$ una descomposición de Cholesky de Σ , ε una muestra gaussiana estándar de tamaño $N = n^2$: entonces, $X = T\varepsilon$ es ahora $\mathcal{N}_N(0, \Sigma)$. Por la descomposición de Cholesky y para un campo sobre $\{1, 2, \dots, 100\}^2$, Σ es de dimensión $10^4 \times 10^4$. Para simular por el muestreador de Gibbs, es necesario efectuar 100 barridos, es decir 100×10^4 visitas sucesivas a los sitios, teniendo en cada visita la simulación de una gaussiana, pero sin la búsqueda de la forma de Cholesky de Σ .

Ejemplo 2.4 *Espacio de estados mixto* $E = (\Lambda \times \mathbb{R})^n$

El muestreador de Gibbs se adapta bien a la simulación de un modelo markoviano con espacio de estados $E = (\Lambda \times \mathbb{R})^n$ mixto: $\Lambda = \{1, 2, \dots, r\}$ es un espacio cualitativo indicando la *marca* λ_i del sitio i , mientras que \mathbb{R} identifica el nivel de gris x_i en este mismo sitio i . Estos modelos, muy útiles en análisis de imágenes o en teoría de señales, serán precisados en el capítulo 4.

2.3. La dinámica de Metropolis-Hastings

Propuesta por Metropolis en 1953 ([74]), el algoritmo tomó su forma general en el artículo de Hastings de 1970 ([51]). A diferencia del muestreador de Gibbs, el algoritmo de Metropolis-Hastings (denotado a partir de ahora M.H.) se puede usar en un espacio de estados E general (para el muestreador de Gibb, E debe ser un espacio producto).

2.3.1. Construcción de la transición de M.H.

Sea $E = \{1, 2, \dots, r\}$ un espacio de estados finito, $\pi > 0$ una distribución sobre E . Dos familias de distribuciones son la base del algoritmo,

(i) Q una transición irreducible sobre E , llamada *proposición de cambio*: $Q(x, y)$ es la probabilidad de proponer el cambio $x \mapsto y$

(ii) $a : E \times E \rightarrow]0, 1]$ la función de *aceptación del cambio*: $a(x, y)$ es la probabilidad de aceptar el cambio $x \mapsto y$. Se supondrá que para todo x , $a(x, x) = 1$: si uno no se mueve, se acepta.

El algoritmo de M.H. es entonces el siguiente: sea x el estado inicial:

(1) Se propone el cambio $x \mapsto y$ siguiendo $Q(x, \cdot)$

(2) Se acepta este cambio con probabilidad $a(x, y)$. Si no, no se modifica el valor x .

La transición P de este algoritmo será entonces:

$$P(x, y) = \begin{cases} Q(x, y)a(x, y) & \text{si } x \neq y \\ Q(x, x) + \sum_{y: y \neq x} Q(x, y)[1 - a(x, y)] & \text{en otro caso} \end{cases} \quad (2.2)$$

Hay dos sorteos aleatorios independientes a realizar: el primero siguiendo la distribución $Q(x, \cdot)$ sobre E ; el segundo bajo una distribución uniforme U sobre $[0, 1]$: si $U \leq a(x, y)$, se acepta el cambio $x \mapsto y$. De lo contrario se permanece en x .

Como se puede constatar, a este nivel, Q y a son independientes de π . Impondremos a P la condición de ser π -reversible, en cuyo caso π será automáticamente P -invariante. Bajo la condición de irreducibilidad y de aperiodicidad, P será entonces ergódica, νP^n convergerá hacia π para toda distribución inicial ν .

2.3.2. π -reversibilidad de M.H. y ergodicidad

La condición de reversibilidad se escribe:

$$\forall x, y \in E, x \neq y : \pi(x)q(x, y)a(x, y) = \pi(y)q(y, x)a(y, x) \quad (2.3)$$

Dado que $\pi > 0$ y $a > 0$, $q(x, y) \neq 0 \Leftrightarrow q(y, x) \neq 0$: bajo (2.3), Q es “*débilmente simétrica*”. En otros términos, se puede ir de x hacia y , si y solamente si se puede ir de y hacia x . Esto se verifica, por ejemplo, si Q es simétrica. En la construcción del algoritmo de M.H., se comienza seleccionando Q ; a tendrá que verificar:

$$\frac{a(x, y)}{a(y, x)} = \frac{\pi(y)q(y, x)}{\pi(x)q(x, y)} \triangleq r(x, y) \quad (2.4)$$

Notemos que para verificar (2.4), es suficiente conocer π salvo un factor multiplicativo.

Demos dos ejemplos de selecciones clásicas de a . Sea $F :]0, +\infty[\rightarrow]0, 1]$, una función que verifica la condición:

$$\forall z > 0 : F(z) = zF\left(\frac{1}{z}\right)$$

Para r definida en (2.4), si a verifica:

$$a(x, y) = F(r(x, y))$$

entonces a verifica la condición de reversibilidad (2.4).

Ejemplo 2.5 *Dinámica de Barker*, $F(z) = \frac{z}{1+z}$ [8]

$$a(x, y) = \frac{\pi(y)q(y, x)}{\pi(x)q(x, y) + \pi(y)q(y, x)}. \text{ Si } Q \text{ es simétrica, } a(x, y) = \frac{\pi(y)}{\pi(x) + \pi(y)}.$$

Ejemplo 2.6 *Dinámica de Metropolis*, $F(z) = \min\{1, z\}$ [74]

$a(x, y) = \min\{1, \frac{\pi(y)q(y, x)}{\pi(x)q(x, y)}\}$. Si Q es simétrica, $a(x, y) = \min\{1, \frac{\pi(y)}{\pi(x)}\}$. El algoritmo de Metropolis toma la forma siguiente:

Algoritmo de Metropolis, el caso Q simétrico

- (i) Sea x el estado inicial: escoger y siguiendo $Q(x, \cdot)$
- (ii) Si $\pi(y) \geq \pi(x)$, guardar y . Volver a (i).
- (iii) Si $\pi(y) < \pi(x)$, lanzar una distribución uniforme U en $[0, 1]$:
 - si $U \leq p = \frac{\pi(y)}{\pi(x)}$, guardar y .
 - si $U > p$, guardar el valor inicial x .
- (iv) Volver a (i).

Ergodicidad de P .

Dado que $a > 0$, P es ergódica si Q es regular. Igualmente, P es irreducible si Q es irreducible. Bastará asegurarse la aperiodicidad de P .

Proposición 2.4 *Ergodicidad del algoritmo de M.H.*

Supongamos que Q sea irreducible y que $a(x, y)$ verifica la condición de reversibilidad (2.4). La ergodicidad de P , y la convergencia $\nu P^n \rightarrow \pi$ para toda distribución inicial ν , se asegura bajo una de las condiciones siguientes:

- (i) Q es regular.
- (ii) $\exists x_0$ t.q. $Q(x_0, x_0) > 0$.
- (iii) $\exists (x_0, y_0)$ t.q. $r(x_0, y_0) < 1$.
- (iv) Q es simétrica y π no es la distribución uniforme.

Demostración : (i) es suficiente dado que $a > 0$. Para las tres otras condiciones, demostremos que existe x_0 t.q. $P(x_0, x_0) > 0$. Esta condición garantiza la aperiodicidad de P . Bajo (ii) o (iii), utilizando la expresión (2.2) que expresa $P(x, x)$ en función de Q y de a , es fácil ver que $P(x_0, x_0) > 0$.

Examinemos (iv). Q siendo irreducible, todos los estados se comunican. Notemos $x \sim y$ si $q(x, y) > 0$ ($\Leftrightarrow q(y, x) > 0$). Puesto que π no es uniforme y que todos los estados se comunican, existe $x_0 \sim y_0$ tales que $\pi(x_0) > \pi(y_0)$. Se tiene entonces:

$$P(x_0, x_0) \geq q(x_0, y_0) \left[1 - \frac{\pi(y_0)}{\pi(x_0)} \right] > 0.$$

□

2.3.3. Ejemplos

Ejemplo 2.7 *Distribución π obtenida a partir de una energía U : $\pi(x) = Z^{-1} \exp U(x)$*

Si Q es simétrica, la dinámica de Metropolis depende de los cocientes $p = \frac{\pi(y)}{\pi(x)} = \exp\{U(y) - U(x)\}$ y del signo de $\Delta U = \{U(y) - U(x)\}$: si $\Delta U \geq 0$, se guarda y ; si no, y se retiene con probabilidad $\exp \Delta U$. Denotando por $a^+ = \sup\{0, a\}$, la transición se escribe para $x \neq y$,

$$P(x, y) = Q(x, y) \exp -[U(x) - U(y)]^+.$$

Ejemplo 2.8 *Dinámica de cambio de spins para un modelo de Ising*

Consideremos el modelo de Ising (2.1) sobre $S = \{1, 2, \dots, n\}^2$, el valor del spin en i será denotado $x_i \in \{-1, +1\}$, la distribución conjunta sobre $E = \{-1, +1\}^S$ está dada por la energía $U(x) = h \sum_{i \in S} x_i + \beta \sum_{\|i-j\|_1=1} x_i x_j$. Sea x una configuración inicial.

Se selecciona la proposición de cambio Q siguiente: se comienza por escoger al azar con distribución uniforme dos sitios i y j de S ; se propone la configuración y idéntica a x en todas partes salvo en que los valores x_i y x_j han sido intercambiados :

$$y^{\{i,j\}} = x^{\{i,j\}}, y_i = x_j \text{ e } y_j = x_i$$

Q es simétrica. Dado que U no hace intervenir sino las interacciones $x_i x_j$ para sitios vecinos, $\Delta U(x, y)$ toma una forma local simple. Si los dos sitios seleccionados por la permutación de spins son i y j , es fácil verificar que, poniendo $v_k = \sum_{l: \langle l, k \rangle} x_l$, se tiene :

$$\Delta U(x, y) = U(y) - U(x) = \begin{cases} \beta(x_j - x_i)(v_i - v_j) & \text{si } \|i - j\|_1 > 1 \\ \beta(x_j - x_i)(v_i - v_j) - \beta(x_j - x_i)^2 & \text{si } \|i - j\|_1 = 1 \end{cases}$$

Un paso del algoritmo de Metropolis de cambio de spins requiere entonces :

- (i) Sortear dos uniformes independientes definidas en $\{1, 2, \dots, n\}$
- (ii) El cálculo de $\Delta U(x, y)$
- (iii) Si $\Delta U(x, y) < 0$, el sorteo de una variable V uniforme en $[0, 1]$, independiente de (i); se acepta la permutación de spins si $V < \exp \Delta U(x, y)$. Si no, no se intercambian los spins.

Sea $x(0)$ la configuración inicial en el paso 0. La proposición de cambio Q obliga a permanecer en el espacio de configuraciones

$$E_0 = E_{x(0)} = \left\{ x \in E \text{ t.q. } \sum_{i \in S} x_i = N(0) = \sum_{i \in S} x_i(0) \right\}$$

A cada proposición de cambio, el número de spins de un signo dado es invariante. Sobre E_0 , Q es irreducible: en efecto dos configuraciones de E_0 se corresponden por (al menos) una permutación de S y toda permutación es un producto finito de transposiciones. Por otra parte, si $x(0)$ no es una configuración constante y si $\beta \neq 0$, π no es constante sobre E_0 , la transición del algoritmo es aperiódica. El algoritmo de Metropolis de cambio de spins es entonces ergódico sobre E_0 , suministrando una simulación de $\pi|_{E_0}$, la distribución π restringida a E_0 .

Ejemplo 2.9 *El muestrador de Gibbs a barrido aleatorio*

El muestrador de Gibbs a barrido aleatorio sobre $E = F^S$ es un algoritmo de M.H. para la dinámica de Metropolis siguiente:

(1) Seleccionar un sitio $i \in S$ al azar bajo la distribución ν ($\nu > 0$); la transición Q hace pasar de $x = (x_i, x^i)$ a $y = (y_i, x^i)$, donde y_i se escoge siguiendo la distribución $\pi_i(\cdot | x^i)$.

(2) Se guarda siempre y : para x e y del tipo precedente, $a(x, y) = 1$.

Esta selección corresponde a la dinámica de Metropolis, $a(x, y) = \sup\{1, r(x, y)\}$; en efecto :

$$r(x, y) = \frac{\nu_i \pi(y_i, x^i) \pi_i(x_i | x^i)}{\nu_i \pi(x_i, x^i) \pi_i(y_i | x^i)} \equiv 1$$

Ejemplo 2.10 *Simulación sobre un espacio provisto de un grafo irreducible*

Supongamos que E (no necesariamente espacio producto) está provisto de una estructura de grafo \mathcal{G} simétrico, es decir de una relación de “vecindad” $x \sim y$ simétrica para $x \neq y$. Se define la vecindad de x como $\mathcal{V}(x) = \{y | y \sim x\}$ y se supone que para todo x , $v(x) = |\mathcal{V}(x)| > 0$, donde $|A|$ es el cardinal de A . Se dirá que el grafo \mathcal{G} es irreducible si para todo $x, y \in E$, existe un camino que va de x a y .

Consideremos ahora la proposición de cambio uniforme de x hacia uno de sus vecinos,

$$q(x, y) = \begin{cases} \frac{1}{v(x)} & \text{si } y \sim x \\ 0 & \text{en otro caso} \end{cases}$$

Q es una transición simétrica, irreducible si \mathcal{G} es irreducible.

Ejemplo 2.11 *Un problema de corte maximal de un grafo*

Especifiquemos E y π en el contexto de un problema de corte maximal de un grafo. Sea $S = \{1, 2, \dots, n\}$ un conjunto de n sitios, y $w = \{w_{ij}, i, j \in S\}$ un conjunto de pesos reales y simétricos sobre $S \times S$: para todo $i, j \in S$, $w_{ij} \in \mathbb{R}$ y $w_{ij} = w_{ji}$. Pongamos $E = \mathcal{P}(S)$ el conjunto de subconjuntos de S , y para $A \in E$,

$$U(A) = \sum_{i \in A, j \notin A} w_{ij}$$

El problema a resolver es el siguiente: buscar A minimizando $U(A)$. Usaremos para esto un algoritmo de recocido simulado (cf. capítulo 3). Describiremos aquí el algoritmo de simulación de la distribución π_β siguiente :

$$\pi_\beta(A) = Z^{-1}(\beta) \exp\{-\beta U(A)\}$$

La relación entre la simulación de π_β y la minimización de U es la siguiente, ésta es la heurística del recocido simulado : el conjunto M_β de las modas de π_β tiende, cuando $\beta \rightarrow +\infty$, hacia el

conjunto M donde U alcanza su mínimo. Así, la simulación de π_β para β grande se concentra aproximadamente alrededor de M .

Algoritmo de Metropolis para la simulación de π_β .

(1) Los únicos cambios autorizados son $A \mapsto B$ donde B difiere de A en exactamente un sitio; dos casos se presentan :

$$\begin{cases} \text{(i)} & B = A \cup \{s\} \text{ si } |A| < n \text{ y si } s \notin A \text{ o} \\ \text{(ii)} & B = A \setminus \{s\} \text{ si } |A| \geq 1 \text{ y } s \in A \end{cases}$$

Notemos que $B \mapsto A$ es posible si $A \mapsto B$ lo es: el grafo de comunicación asociado $A \sim B$ es simétrico.

(2) Se selecciona s uniformemente sobre S : si $s \in A$, se toma $B = A \setminus \{s\}$; si $s \notin A$, se toma $B = A \cup \{s\}$. En estos dos casos, $q(A, B) = \frac{1}{n}$; si no, $q(A, B) = 0$.

Q es simétrico e irreducible. Notemos que Q no es regular, puesto que allí donde ella no es nula, ella cambia la paridad de un subconjunto.

(3) Evaluar $\Delta U = U(B) - U(A)$: para (i), $\Delta U = \sum_{j \notin B} w_{sj} - \sum_{i \in A} w_{is}$; para (ii), $\Delta U = \sum_{i \in B} w_{is} - \sum_{j \notin A} w_{sj}$.

(4) π_β está asociada a la energía $-\beta U$. Así, si $\Delta U \leq 0$, se guarda B . En otro caso, se guarda B con probabilidad $p = \exp -\beta \Delta U$.

Al ser Q simétrica, el algoritmo es ergódico ya que U es no constante.

2.3.4. Algunos resultados generales sobre las transiciones π -reversibles

Sea P una transición ergódica y π -reversible sobre $E = \{1, 2, \dots, r\}$, $X = (X_0, X_1, X_2, \dots)$ una cadena homogénea de transición P . Notemos $l^2(\pi) = l^2_{\mathbb{C}}(\pi)$ el espacio de las funciones reales definidas sobre E con el producto escalar $\langle f, g \rangle = \sum f(x)\bar{g}(x)\pi(x)$. Para toda distribución inicial ν y todo $f \in \Omega$, se tiene [58] :

$$v(f, \pi, P) = \lim_{T \rightarrow \infty} \text{Var}(T^{-\frac{1}{2}} \sum_{t=0}^{T-1} f(X_t)) \text{ existe y es independiente de } \nu.$$

Proposición 2.5 *Espectro de una transición reversible y valor de $v(f, \pi, P)$*

(1) Si P es π -reversible y ergódica, P es auto-adjunta sobre $l^2(\pi)$. P es diagonalizable de autovalores reales que verifican

$$\lambda_1 = 1 < \lambda_2 \leq \dots \leq \lambda_r \leq -1$$

$\lambda_r = -1$ corresponde al caso de una cadena 2-periodica.

(2) Escojamos los autovectores asociados $\{e_1 = \mathbf{1}, e_2, \dots, e_r\}$ ortonormales en $l^2(\pi)$. Entonces :

$$v(f, \pi, P) = \sum_{k=2}^r \frac{1 + \lambda_k(P)}{1 - \lambda_k(P)} \langle f, e_k \rangle_{\pi}^2 \quad (2.5)$$

Comentarios:

(1) λ_2 se llama el *autovalor subdominante* de P : él permite identificar el factor de varianza asintótica $v(f, \pi, P)$ en el T.C.L. para $T^{-\frac{1}{2}} \sum_{t=0}^{T-1} f(X_t)$ y, para este criterio, permite comparar dos transiciones.

(2) $\rho_2 = \sup\{|\lambda_2|, |\lambda_r|\}$ controla la velocidad de ergodicidad porque $\|\nu P^n - \pi\| \leq C(\nu)\rho_2(P)^n$.

(3) Los controles explícitos de λ_2 o de ρ_2 son en general imposibles. Mayoraciones y minoraciones del hueco espectral ("spectral gap") $(1 - \lambda_2)$ de P pueden conseguirse en ([29], Ch.6.II.4, p. 250).

Demostración:

(1) Poniendo $Ph(x) = \sum P(x, y)h(y)$, se verifica fácilmente, usando la reversibilidad para la segunda desigualdad, que

$$\langle f, Pg \rangle = \sum f(x)P\bar{g}(x)\pi(x) = \sum \bar{g}(x)Pf(x)\pi(x) = \langle Pf, g \rangle$$

(2) Descompongamos f en la base de autovectores : $f = \sum_{i=1, r} a_i e_i$, donde $a_i = \langle f, e_i \rangle$. Si f es real, $a_i \in \mathbb{R}$. Se tiene: (i) $\pi(f) = a_1$; (ii) $P^n f = \sum_{i=1, r} a_i \lambda_i^n e_i$. Se calcula la covarianza entre $f(X_0)$ y $f(X_n)$, en el régimen estacionario, y para f real:

$$\begin{aligned} E(f(X_0)f(X_n)) &= \sum_{i=1, r} \bar{a}_i \lambda_i^n \langle f, e_i \rangle = \sum_{i=1, r} \lambda_i^n |a_i|^2 = \sum_{i=1, r} \lambda_i^n a_i^2 \\ Var(f(X_0)) &= \sum_{i=2, r} a_i^2, \text{ et } Cov(f(X_0), f(X_n)) = \sum_{i=2, r} a_i^2 \lambda_i^n \end{aligned}$$

Se deduce entonces fácilmente (2). □

2.3.5. Comparación de diferentes dinámicas de M.H.

Definiciones y resultados preliminares.

Fijando la transición de la proposición de cambio, vamos a comparar las dinámicas de M.H. correspondientes a selecciones diferentes de la función de aceptación a .

Si P y Q son dos transiciones sobre $E = \{1, 2, \dots, r\}$, se dice que P domina Q fuera de la diagonal (lo que notaremos $P \succ Q$) si para todo $i \neq j$, $p_{ij} \geq q_{ij}$. Sea $\pi > 0$ una probabilidad en E . Un operador M sobre $\Omega = \mathbb{R}^E$ (es decir una matrix $r \times r$) es positiva si para todo $f \in \Omega$,

$$\langle Mf, f \rangle_\pi = \sum_{i, j} M_{ij} f_i f_j \pi_i \geq 0$$

con $(Mf)_i = \sum_j M_{ij} f_j$.

El resultado siguiente se debe a Peskun [82]. Él permite establecer la optimalidad de la dinámica de Metropolis en la familia de las diferentes selecciones posibles de a (2.4) conduciendo a una transición P reversible (para Q fijo), óptima para el criterio de la varianza más pequeña $v(f, \pi, P)$. Estos resultados han sido generalizados por Tierney [100] en el caso de un espacio de estados general.

Proposición 2.6 Sean P y Q dos transiciones π -reversibles tales que $P \succ Q$.

(1) $(Q - P)$ es un operador positivo. En particular :

$$Cov_P(f(X_0), f(X_1)) \leq Cov_Q(f(X_0), f(X_1))$$

(2) Para toda $f \in \Omega$, $v(f, \pi, P) \leq v(f, \pi, Q)$.

(3) En particular, $\lambda_2(P) \leq \lambda_2(Q)$.

Demostración :

(1) Pongamos $h_{ij} = \pi_i(\delta_{ij} + p_{ij} - q_{ij})$, donde δ_{ij} es la función de Dirac. $h \geq 0$ y

$$\begin{cases} \sum_j h_{ij} = \pi_i & (P \text{ y } Q \text{ son transiciones}) \\ \sum_i h_{ij} = \pi_j & (P \text{ y } Q \text{ son } \pi \text{-reversibles}) \end{cases}$$

En particular, $\sum_{ij} h_{ij} = 1$. Se tiene la sucesión de igualdades :

$$\begin{aligned} \langle (Q - P)f, f \rangle_\pi &= \sum_{i,j} (q_{ij} - p_{ij}) f_i f_j \pi_i = \sum_{i,j} [\pi_i \delta_{ij} - h_{ij}] f_i f_j \\ &= \sum_i f_i^2 \pi_i - \sum_{i,j} f_i f_j h_{ij} \\ &= \frac{1}{2} \left\{ \sum_{i,j} f_i^2 h_{ij} + \sum_{i,j} f_j^2 h_{ij} - 2 \sum_{ij} f_i f_j h_{ij} \right\} \\ &= \frac{1}{2} \sum_{i,j} (f_i - f_j)^2 h_{ij} \geq 0 \end{aligned}$$

La relación entre las covarianzas es una consecuencia directa de esta positividad tomando f centrada.

(2) cf. [82], [58].

(3) A partir de (2) y de (2.5), seleccionando $f = e_2(P)$,

$$v(f, \pi, P) = \frac{1 + \lambda_2(P)}{1 - \lambda_2(P)} \leq \sum_{k=2}^r \frac{1 + \lambda_k(Q)}{1 - \lambda_k(Q)} \langle e_2(P), e_k(Q) \rangle^2 \leq \frac{1 + \lambda_2(Q)}{1 - \lambda_2(Q)}$$

porque $\lambda_2(Q) \geq \lambda_k(Q)$ para $k \geq 2$ y $\|e_2(P)\|_\pi^2 = 1$. □

Consecuencias.

(1) *Optimalidad de la dinámica de Metropolis.*

Fijemos la transición de la proposición de cambio Q . Las selecciones de la probabilidad de aceptación a que hacen a la transición de M.H. P reversible son aquellas para las cuales

$$a(x, y) = a(y, x)r(x, y) \text{ donde } r(x, y) = \frac{\pi(y)Q(y, x)}{\pi(x)Q(x, y)}$$

la dinámica de Metropolis corresponde a la selección $a_M(x, y) = \min\{1, r(x, y)\}$. Entonces,

$$\begin{cases} \text{si } r(x, y) < 1, a_M(x, y) = r(x, y) \geq a(x, y) \\ \text{si } r(x, y) \geq 1, a_M(x, y) = 1 \geq a(x, y) \end{cases}$$

Si $x \neq y$,

$$P_M(x, y) = Q(x, y)a_M(x, y) \geq P_{MH}(x, y) = Q(x, y)a(x, y)$$

Para el criterio de la varianza $v(f, \pi, P)$, la dinámica P_M de Metropolis es óptima.

(2) *Mejoras del muestreador de Gibbs : Gibbs Metropolizado [70].*

En el muestreador de *Gibbs con barrido aleatorio*, impongamos el cambio (si el sitio i se escoge por el intercambio) $x = (x_i, x^i) \rightarrow y = (y_i, x^i)$, para la distribución $\pi_i(\cdot | x^i)$ con $y_i \neq x_i$. La probabilidad de transición es

$$P_1(x, y) = \nu_i \frac{\pi_i(y_i | x^i)}{1 - \pi_i(x_i | x^i)}$$

Para el muestreador de *Gibbs Metropolizado* y las mismas proposiciones de cambio $x \mapsto y$

$$q(x, y) = \frac{\pi_i(y_i | x^i)}{1 - \pi_i(x_i | x^i)}$$

(si $F = \{a, b\}$ es de dos estados, se cambia sistemáticamente x_i en otro valor). Se aceptará y con probabilidad

$$a_M(x, y) = \min\{1, r(x, y)\} = \min\left\{1, \frac{1 - \pi_i(x_i | x^i)}{1 - \pi_i(y_i | x^i)}\right\}$$

Para un barrido aleatorio y para $x \neq y$, la transición es

$$P_2(x, y) = \nu_i \min\left\{\frac{\pi_i(y_i | x^i)}{1 - \pi_i(x_i | x^i)}, \frac{\pi_i(y_i | x^i)}{1 - \pi_i(y_i | x^i)}\right\} \geq \nu_i \pi_i(y_i | x^i) = P_1(x, y)$$

P_2 domina P_1 : el muestreador de Gibbs Metropolizado es preferible al muestreador de Gibbs habitual.

Comparación de los algoritmos de Gibbs y de Metropolis.

Cuando E es un espacio producto, se puede definir simultáneamente el muestreador de Gibbs (G) y el algoritmo de Metropolis (M). Más generalmente Frigessi et al [33] estudian la familia de algoritmos π -reversibles intercambiado las configuraciones sitio por sitio, y comparan los autovalores subdominantes ρ_2 . Para un barrido aleatorio, los autores establecen que $\rho_2 = \lambda_2$ para los dos algoritmos. Muestran que si el intercambio de Metropolis es iterado dos veces (o un número *par* de veces k) en cada sitio (algoritmos (Mk)), (G) es preferible a (Mk). La descripción completa de estos algoritmos de intercambio sitio por sitio para el modelo de Ising se propone como ejercicio.

2.3.6. Algoritmo de M.H. para un espacio de estados general [98]

Sea E un espacio de estados general, λ una medida de referencia en E . Típicamente, $E = \mathbb{R}^d$ y λ es la medida de Lebesgue. Se estudiará también el caso de espacios del tipo $E = \bigoplus_{d=1}^N \mathbb{R}^d$ en el contexto de la simulación de procesos puntuales de Gibbs, N corresponde a una cota superior del número de puntos de la realización del proceso. Sea π la distribución a simular : se supone que π admite una densidad con respecto a λ , de nuevo llamada π y con $\pi(x) > 0$ para toda configuración x .

Consideremos ahora Q una transición de proposición con densidad q , $Q(x, dy) = q(x, y)\lambda(dy)$. Para las selecciones y posibles (es decir tales que $q(x, y) > 0$), la razón de M.H. es

$$r(x, y) = \frac{\pi(y)q(y, x)}{\pi(x)q(x, y)}$$

Por ejemplo, la dinámica de Metropolis está asociada a la probabilidad $a(x, y) = \min\{1, r(x, y)\}$. De manera más general, para una probabilidad a cumpliendo la condición de reversibilidad (2.4) y poniendo $p(x, y) = q(x, y)a(x, y)$ si $x \neq y$, 0 en otro caso, la transición de M.H. es

$$P(x, dy) = p(x, y)\lambda(dy) + [1 - \int p(x, z)\lambda(dz)]\delta_x(dy)$$

π es invariante para P . Es necesario asegurar la π -irreducibilidad y la aperiodicidad de P para tener la convergencia

$$\forall x \in E, \|P^n(x, \cdot) - \pi(\cdot)\|_{VT} \rightarrow 0$$

El hecho de que la convergencia tenga lugar para todo x está ligado al hecho de que una transición de M.H. es Harris recurrente (cf. Corolario 2 de [98]).

Si Q es π -irreducible, P también lo es. Una condición suficiente de aperiodicidad es

$$\lambda\{x : 1 - \int p(x, z)\lambda(dz) > 0\} > 0$$

Es decir, en un conjunto de λ -medida > 0 , no se cambia la configuración con una probabilidad > 0 .

Ejemplo 2.12 *Simulación de un Proceso Puntual de Gibbs en $S = [0, 1]^2$*

Presentaremos en el capítulo 4 los procesos markovianos de objetos. Cuando los objetos se reducen a un punto de \mathbb{R}^d , hablaremos de procesos puntuales. Presentamos aquí estos procesos puntuales así como su simulación.

Un proceso puntual (denotado de ahora en adelante P.P.) en $S = [0, 1]^2$ es una variable aleatoria que toma sus estados x en el espacio exponencial de configuraciones $E = \cup_{n \geq 0} S^n$. Si $x = (x_1, x_2, \dots, x_n)$, hay $n = n(x)$ puntos de S en la realización del P.P., el orden de enumeración de los puntos se mantiene (habrá entonces cada vez $n!$ realizaciones equivalentes). Para una presentación general de modelos de P.P., se puede consultar ([14], [94]).

El P.P. (la medida) de referencia es el P.P. de Poisson (P.P.P.). El P.P.P. de intensidad 1 en S está caracterizado por :

- (i) la probabilidad de que haya n puntos en la realización es $(n! \times e)^{-1}$.
- (ii) si hay n puntos, éstos están repartidos al azar uniformemente en S .

Así, la densidad del P.P.P. es

$$p(x) = \frac{1}{e} \sum_{n \geq 0} \frac{\mathbf{1}(n(x) = n)}{n!} dx_1 dx_2 \dots dx_n$$

Una manera clásica de definir otros P.P. es definir su densidad con respecto al P.P.P. Sea $U : E \rightarrow R$ una función de energía invariante por permutación de coordenadas de x ($U(x) = U(\sigma(x))$) para toda permutación σ de coordenadas de x , que verifica la condición de admisibilidad,

$$Z = \sum_{n \geq 0} \frac{1}{n!} \int_{x \in S^n} \exp U(x) dx < \infty$$

Entonces la densidad $f(x) = Z^{-1} \exp U(x)$ define un P.P. en S . Por ejemplo, $U(x) \equiv 0$ nos da el P.P.P. de parámetro 1, $U(x) \equiv a$ da el P.P.P. de intensidad e^a . Presentemos algunos modelos clásicos.

Ejemplo : *Modelo uniforme de núcleo duro (Hardcore model)*

Sea $r > 0$ un radio fijo. Un proceso de núcleo duro es un P.P.P. condicionado al hecho de que dos puntos cualesquiera de la realización están siempre a una distancia $\geq r$. La densidad de un tal modelo es, si $x = (x_1, x_2, \dots, x_n)$:

$$f(x) = c \mathbf{1}(\forall i \neq j, \|x_i - x_j\| \geq r)$$

Por ejemplo en ecología, este condicionamiento traduce la existencia de zonas de influencia de plantas situadas en los sitios x_i . En física, se habla de modelo de esferas no penetrables : existe un radio de embotellamiento irreducible.

Ejemplo : *Modelo de Strauss*

Para una realización de $n = n(x)$ puntos y para dos constantes reales a y b , el modelo de Strauss (Strauss, [95]; [87]) es un modelo de energía

$$U(x) = an(x) + bs(x)$$

donde $s(x) = \sum_{i \neq j} \mathbf{1}(\|x_i - x_j\| < r)$, para un valor dado r . Se trata de un modelo de Gibbs de potenciales evaluados en los conjuntos de un solo elemento $\Phi_{\{i\}}(x) = a$, y de potenciales de pares $\Phi_{\{i,j\}}(x) = b\mathbf{1}(\|x_i - x_j\| < r)$. a es un parámetro de intensidad, y b un parámetro de interacción. Bajo la forma exponencial, la densidad es

$$f(x) = Z^{-1} \beta^{n(x)} \gamma^{s(x)}, \text{ con } a = \log \beta \text{ y } b = \log \gamma$$

Si se impone $n(x) \leq N < \infty$, U es admisible. Sin acotar $n(x)$, U es admisible si y sólo si $b \leq 0$ ($\gamma \leq 1$) ([63]). Para $\gamma = 0$, se encuentra de nuevo el modelo de núcleo duro, $\gamma = 1$ es el P.P.P., $\gamma > 1$ (y n fijo) favorece las configuraciones de agregados, mientras que $\gamma < 1$ favorece las configuraciones más regulares.

Ejemplo: *Modelos de interacción de áreas* (cf. Capítulo 4)

Estos modelos, con un uso más sencillo que el modelo de Strauss, fueron introducidos por Baddeley y van Lieshout [6]. Sea $x = (x_1, x_2, \dots, x_n)$ una realización. Coloquemos bolas de radios r centradas en x_i , $i = 1, n$. Notemos $A(x)$ el área de la unión de estas bolas. Los modelos de interacción de áreas tienen por densidad

$$f(x) = Z^{-1} \beta^{n(x)} \gamma^{A(x)}$$

Estos modelos son admisibles sin limitación cuando $\gamma > 0$: β regula la intensidad y γ discrimina entre las situaciones más regulares ($\gamma < 1$) y aquellas con formación de agregados ($\gamma > 1$).

Estos diferentes modelos de P.P. gibsianos se prestan bien a la simulación usando la dinámica de Metropolis ([78], [48], [34]). Se examinará separadamente la situación donde $n(x) = n$ está fijo o no.

(1) *Dinámica de Metropolis* : el caso $n(x) = n$ fijo.

$f(x)$ está concentrado en $E = S^n$. Sea $E^+ = \{x \in E, f(x) > 0\}$ el soporte de f .

Simulación de un Proceso Puntual

(i) **Proposición de cambio** : sea $x = (x_1, x_2, \dots, x_n) \in E^+$

- Escoger $\eta = x_i \in x$ con la probabilidad uniforme $\frac{1}{n}$.
- Reemplazar η por ξ escogido uniformemente en $S^+(x, \eta)$

donde $S^+(x, \eta)$ es tal que $f((x \setminus \eta) \cup \xi) > 0$ donde $(x \setminus \eta) \cup \xi = (x \setminus \{\eta\}) \cup \{\xi\}$. La proposición de cambio es

$$x \mapsto y = (x \setminus \eta) \cup \xi$$

Si $E = E^+$, la densidad de transición es, para $x \mapsto y$, $q(x, y) = \frac{1}{n|S|}$, 0 en otro caso.

(ii) Aceptar y con la probabilidad $a(x, y) = \min\{1, \frac{f(y)}{f(x)}\}$. En otro caso, permanecer en x .

La transición de Metropolis es ergódica si q es irreducible y aperiódica. Lo que se cumple si $f > 0$ en todas partes. En n pasos, se puede pasar de todo x a todo y y dado que $a > 0$, la cadena es aperiódica. Pueden existir situaciones donde q no sea irreducible (cf. ejercicio).

(2) *Dinámica de Metropolis : el caso sin restricciones.*

Para simplificar la presentación, supondremos que $E = E^+$. E es de dimensión variable. Diremos que la densidad f del P.P. es *hereditaria* si para $x \in E$ y $\xi \in S$ $f(x \cup \xi) > 0 \Rightarrow f(x) > 0$.

(i) Proposición de cambio : sea $x \in E$

- Con una probabilidad $p(x)$, se incorpora un punto ξ seleccionado en S con una densidad $b(x, \cdot) : y = x \cup \xi$.

- Con probabilidad $1 - p(x)$, se quita uno de los puntos η de x con la probabilidad $d(x \setminus \eta, \eta)$, $\eta \in X : y = x \setminus \eta$.

Esto deja un gran margen para la proposición de cambio.

(ii) La probabilidad de aceptación de y es $a(x, y) = \min\{1, r(x, y)\}$ donde

$$\begin{cases} r(x, x \cup \xi) = \frac{f(x \cup \xi)}{f(x)} \frac{1-p(x)}{p(x)} \frac{d(x, \xi)}{b(x, \xi)} & \text{si } y = x \cup \xi \\ r(x, y) = \frac{1}{r(y, x \setminus \eta)} & \text{si } y = x \setminus \eta \text{ con } \eta \in x \end{cases}$$

Por ejemplo, si $p(x) = 1 - p(x) \equiv \frac{1}{2}$, si $b(x, \xi) \equiv \frac{1}{|S|}$ y $d(x, \xi) = \frac{1}{n}$ si $n(x) = n$, se tiene

$$\begin{cases} r(x, x \cup \xi) = \frac{f(x \cup \xi)}{f(x)} \frac{|S|}{n} & \text{si } y = x \cup \xi \\ r(x, y) = \frac{f(x)}{f(x \setminus \eta)} \frac{n}{|S|} & \text{si } y = x \setminus \eta, \eta \in x \end{cases}$$

La irreducibilidad de q está asegurada puesto que : para x e y dos estados de E , existe un camino de longitud $n(x) + n(y)$ que une x con y ; es suficiente en efecto pasar de x a la configuración vacía \emptyset , y borrar los $n(x)$ puntos de x punto a punto, después, al llegar a la configuración vacía, se hacen aparecer progresivamente los $n(y)$ puntos de y .

2.4. Ejercicios

Ejercicio 2.1 Distribución conjunta y distribuciones condicionales

(1) *Una distribución conjunta positiva está caracterizada por sus distribuciones condicionales.*

Sea $\pi > 0$ una distribución en $E = F_1 \times F_2 \times \dots \times F_n$, cada F_i es finito. Sea ω_i un estado de referencia para cada F_i , $\omega = (\omega_i)$. Verificar que la distribución conjunta puede ser reconstruida a partir de sus distribuciones condicionales en base a la identidad

$$\pi(x) = \pi(\omega) \prod_{i=1}^n \frac{\pi_i(x_i | \omega_1, \dots, \omega_{i-1}, x_{i+1}, x_n)}{\pi_i(\omega_i | \omega_1, \dots, \omega_{i-1}, x_{i+1}, x_n)}$$

(2) *En general y sin restricciones, una familia de n distribuciones condicionales no se integran en una distribución conjunta.*

Sea $F_1 = \{1, 2, \dots, m_1\}$ y $F_2 = \{1, 2, \dots, m_2\}$, m_1 y $m_2 > 2$. Nos damos unas familias de distribuciones $(X | Y = y)$ y $(Y | X = x)$ respectivamente en F_1 y en F_2 , para $y \in F_2$ y $x \in F_1$. Evaluar la dimensión paramétrica de estas dos familias si no hay restricciones. ¿Cuál es la dimensión de un modelo conjunto (X, Y) general en $E \times F$? Conclusiones.

Ejercicio 2.2 Muestreador de Gibbs para un proceso bivariado con componentes binarias

Sea $S = \{1, 2, \dots, n\}^2$ el toro bidimensional, provisto de la relación de 4-v.m.c., relación prolongada por periodicidad. Se considera en S el proceso bivariado $Z_i = (X_i, Y_i) \in \{0, 1\}^2, i \in S$ con distribución de Gibbs

$$\begin{aligned}\pi(z) &= Z^{-1} \exp\left[\sum_{i \in S} \Phi_1(z_i) + \sum_{\langle i, j \rangle} \Phi_2(z_i, z_j)\right] \text{ con} \\ \Phi_1(z_i) &= \alpha x_i + \beta y_i + \gamma x_i y_i \text{ y } \Phi_2(z_i, z_j) = \delta x_i x_j + \eta y_i y_j \text{ para } \langle i, j \rangle\end{aligned}$$

(1) Determinar las distribuciones condicionales siguientes : $\pi_i(z_i | z^i)$, $\pi_i^1(x_i | x^i, y)$ y $\pi_i^2(y_i | x, y^i)$.

(2) Construir un muestreador de Gibbs usando las dos familias π_i^1 y π_i^2 , $i \in S$. Establecer la ergodicidad de este muestreador.

Ejercicio 2.3 *Simulacin de un proceso gaussiano bivariado*

En el mismo contexto que el ejercicio anterior, se quiere simular la distribución gaussiana bivariada en S de densidad

$$\pi(z) = Z^{-1} \exp\left\{-\sum_{i \in S} (x_i^2 + y_i^2) + \beta \sum_{\langle i, j \rangle} (x_i y_j + x_j y_i)\right\}, |\beta| < \frac{1}{2}$$

Determinar las distribuciones condicionales $\pi_i(z_i | z^i)$, $\pi_i^1(x_i | x^i, y)$ y $\pi_i^2(y_i | x, y^i)$. Proponer dos procedimientos de simulación.

Ejercicio 2.4 *Generar una permutacin aleatoria [20]*

Se quiere generar una permutación aleatoria de distribución π uniforme en el conjunto $E = \Sigma_n$ de las permutaciones de $\{1, 2, \dots, n\}$. Demostrar que la transición P en E que permuta dos índices i y j escogidos al azar es π -reversible. Deducir una manera de sacar al azar una permutación.

Ejercicio 2.5 *Cadena de Markov de campo de Markov o dinamica espacio-temporal*

Se considera $S = \{1, 2, \dots, n\}$, $F = \{0, 1\}$ y la cadena homogénea $X = (X(t), t \geq 0)$ en $E = F^S$ de transición

$$\begin{aligned}P(x(t-1), x(t)) &= Z(x(t-1))^{-1} \exp \sum_{i \in S} x_i(t) [\alpha + \beta v_i(t) + \gamma w_i(t-1)] \text{ con} \\ v_i(t) &= x_{i-1}(t) + x_{i+1}(t) \text{ y } w_i(t-1) = x_{i-1}(t-1) + x_i(t-1) + x_{i+1}(t-1)\end{aligned}$$

con la convención $z_i = 0$ si $i \notin S$.

(1) ¿Cómo simular una tal dinámica por el muestreador de Gibbs? Escribir las distribuciones condicionales que intervienen.

(2) Poner en práctica esta dinámica para $\alpha = -\beta = 2$ y $n = 100$. Estudiar la evolución de manchas $N_t = \{i \in S : X_i(t) = 1\}$ en función del parámetro γ .

(3) Proponer un modelo análogo en $S = \{1, 2, \dots, n\}^2$ y simularlo.

Ejercicio 2.6 *La composicin de transiciones π -reversibles no es automticamente π -reversible*

(1) Sean P y Q dos transiciones π -reversibles en E finito. Demostrar que $\forall x, y$

$$\pi(x)PQ(x, y) = \pi(y)QP(y, x)$$

(2) Determinar la forma explícita del muestreador de Gibbs $P = P_1P_2$ en el caso $F = \{0, 1\}$ y $S = \{1, 2\}$ (dos barridos), para una distribución π general con tres parámetros. Deducir que existe π tal que $P_1P_2 \neq P_2P_1$, y que el muestreador de Gibbs no es π -reversible.

Ejercicio 2.7 *Muestreadores asncrono y snrono de un modelo de Ising*

Poner en práctica los muestreadores de Gibbs asíncrono (o secuencial) y síncrono para el modelo de Ising $\pi(x) = Z^{-1} \exp \beta \sum_{\langle i, j \rangle} x_i x_j$, $\beta = 0, 2, 0, 5, 1$, sobre el toro bidimensional $S = \{1, 2, \dots, N\}^2$, $N = 64$, provisto del grafo de vecindad de 4-v.m.c.. Comparar los dos tipos de algoritmos. Calcular la correlación empírica a distancia 1 para una y otra de las simulaciones y verificar que para la simulación síncrona, ésta es nula.

Ejercicio 2.8 *Simulacin de un modelo de Ising por la dinmica de intercambio de spins*

En el mismo contexto que el ejercicio anterior, efectuar la simulación del modelo de Ising por intercambio de spins, la configuración inicial se haya equilibrada entre los dos spins $+1$ y -1 . Calcular la correlación empírica a distancia 1. Comparar esta simulación a la obtenida por el muestreador de Gibbs secuencial.

Ejercicio 2.9 *Muestreador de Gibbs gaussiano : una aproximacin directa [111]*

Retomemos las notaciones del párrafo correspondiente a la simulación de un vector gaussiano $\pi \sim \mathcal{N}_n(0, \Sigma)$, con matriz de covarianza inversa $Q = \Sigma^{-1} = (q_{ij})$. Consideremos la situación de barridos periódicos $1 \mapsto 2 \mapsto \dots \mapsto n$ de S . Si $X(k)$ es el estado al tiempo k , y $X(k+1)$ al tiempo $k+1$ después del intercambio en el sitio $i = i(k)$, se tiene

$$X_i(k+1) \sim N\left(-\frac{1}{q_{ii}} \sum_{j \neq i} q_{ij} X_j(k), \frac{1}{q_{ii}}\right)$$

Si uno parte de un estado inicial $X(0) = x(0) \in \mathbb{R}^n$, $X(k)$ es un vector gaussiano para todo k . Es suficiente mostrar que la media $M(k)$ y la varianza $\Sigma(k)$ de $X(k)$ tienden respectivamente a 0 y Σ cuando $k \rightarrow \infty$.

Sean $A_i = I - B_i$ donde B_i es la matrix $n \times n$ con todas sus filas nulas salvo la i -ésima fila que vale $q_{ii}^{-1}(q_{i1}, q_{i2}, \dots, q_{in})$. Sea S_i la matrix $n \times n$ de coeficientes todos nulos salvo el término (i, i) igual a q_{ii}^{-1} .

(1) Verificar que $(X_i(k+1) | X(k)) \sim \mathcal{N}(A_i X(k), S_i)$. Deducir que :

$$M(k+1) = A_{i(k)} M(k) \text{ y } \Sigma(k+1) = S_{i(k)} + A_{i(k)} \Sigma(k)^t A_{i(k)}$$

(2) La distribución π es invariante para un paso del muestreo, deducir entonces que

$$\Sigma(k+1) - \Sigma = A_{i(k)} (\Sigma(k) - \Sigma)^t A_{i(k)}$$

Así, la ergodicidad del muestreador de Gibbs estará asegurada por la convergencia

$$A(k) = A_{i(k)} A_{i(k-1)} \cdots A_{i(0)} \rightarrow 0 \text{ para } k \rightarrow \infty$$

(3) Sea $\langle x, y \rangle_Q$ la norma inducida por Q , $\|\cdot\|_Q$ la norma matricial correspondiente. Demostrar que para este producto escalar, A_i y B_i son proyectores ortogonales. Demostrar que $\|A(n)\|_Q \leq \varrho < 1$. Deducir la ergodicidad del muestreador de Gibbs.

Observación : este resultado se extiende a barridos no necesariamente periódicos, bajo la condición de que cada barrido cubra a S y que el número de pasos de un barrido sea acotado [111].

Ejercicio 2.10 *Muestreo sincrónico, distribución virtual μ asociada a π*

(1) La distribución virtual μ difiere en general de π . Sea $F = \{0, 1\}$ y $S = \{1, 2\}$ y π definida en E por : $\pi(0, 0) = \pi(1, 1) = 0,1$ y $\pi(0, 1) = \pi(1, 0) = 0,4$. Escribir de manera explícita el núcleo de Gibbs asíncrono P y el núcleo de Gibbs sincrónico Q . Verificar que $\pi Q \neq \pi$.

(2) Cálculo explícito de la distribución virtual para el modelo de Ising de v.m.c.. En el contexto siguiente, vamos a poder mostrar la distribución virtual μ . π es un modelo binario de estados $\{0, 1\}$ sobre $S = \{1, 2, \dots, n\}$ provisto de un grafo de vecindad $\langle \cdot, \cdot \rangle$, de densidad

$$\pi(x) = Z^{-1} \exp \beta \sum_{\langle i, j \rangle} x_i x_j$$

(2.1) Verificar que la distribución condicional en i es

$$\pi_i(x_i | x^i) = \frac{e^{x_i v_i(x)}}{1 + e^{v_i(x)}} \text{ donde } v_i(x) = \beta \sum_{j: \langle i, j \rangle} x_j$$

(2.2) Determinar el núcleo de muestreo sincrónico Q . Verificar que la distribución μ

$$\mu(x) = \Gamma^{-1} \prod_{i=1}^n [1 + e^{v_i(x)}]$$

es invariante para Q (Q es también μ -reversible).

(2.3) Verificar que la distribución μ es markoviana, con conjuntos de cliques”

$$\mathcal{C} = \{\partial i = \{j \in S : \langle i, j \rangle\}, i \in S\}$$

Ejemplo : sea $S = \{1, 2, \dots, N\}^2$ el toro bidimensional, N par, y $\langle i, j \rangle$ la relación de 4-v.m.c. prolongada por periodicidad. Comparar los grafos de Markov de π y de μ . Verificar por ejemplo que si μ^+ (resp. μ^-) es la distribución marginal de μ en $S^+ = \{i = (i_1, i_2) \in S, i_1 + i_2 \text{ par}\}$ (resp. $S^- = S \setminus S^+$), entonces $\mu = \mu^+ \otimes \mu^-$.

Ejercicio 2.11 *Muestreador de Gibbs Metropolizado*

Demostrar que la transición P del algoritmo de Gibbs Metropolizado es π reversible e irreducible. Verificar que $P^2(x, x) > 0$ y caracterizar las configuraciones x t.q. $P(x, x) = 0$. Deducir que P es aperiódica si π no es la distribución uniforme.

Ejercicio 2.12 *Dinámica general de intercambio sitio por sitio [33]*

Sea $F = \{-1, +1\}$, $S = \{1, 2, \dots, n\}^2$ el toro bidimensional, $\langle i, j \rangle$ la relación de 4-v.m.c., con condiciones de bordes periódicas y

$$\pi(x) = Z^{-1} \exp \beta \sum_{\langle i, j \rangle} x_i x_j$$

Vamos a examinar las dinámicas que surgen de las transiciones P_s donde sólo un intercambio es efectuado en s : $P_s(x, y) = 0$ si $x^s \neq y^s$. Para $s \in S$ se denota $m \triangleq m(s)$ el número de spins positivos entre los cuatro vecinos de s . Si se impone :

(i) por una parte, que P_s no depende sino de m

$$P_s(x, y) = P_s(x_s \rightarrow y_s \mid m) \text{ para } x_s = y_s$$

(ii) por la otra, la simetría siguiente con respecto a la transformación $x \mapsto -x$

$$P_s(-x_s \rightarrow -y_s \mid 4 - m) = P_s(x_s \rightarrow y_s \mid m)$$

(1) ¿Cuales son los valores posibles de m ? Demostrar que P_s está enteramente determinada por los cinco parámetros siguientes :

$$\begin{aligned} a_4 &= P_s(-1 \rightarrow +1 \mid m = 4) \text{ y } a_3 = P_s(+1 \rightarrow -1 \mid m = 4) \\ a_2 &= P_s(-1 \rightarrow +1 \mid m = 3) \text{ y } a_1 = P_s(+1 \rightarrow -1 \mid m = 3) \\ a_0 &= P_s(+1 \rightarrow -1 \mid m = 2) \end{aligned}$$

(2) Demostrar que P_s es π -reversible si y solamente si

$$a_3 = a_4 \exp(-8\beta) \text{ y } a_1 = a_2 \exp(-4\beta)$$

Se supondrá que se verifican estas dos condiciones y denotaremos $a = (a_0, a_2, a_4) \in]0, 1]^3$ los tres parámetros que definen la dinámica.

(3) Indicar los valores de $a = (a_0, a_2, a_4)$ para los algoritmos siguientes : (i) Muestreador de Gibbs; (ii) muestreador de Metropolis.

(4) Demostrar que la dinámica asociada a una tal transición P_s para un selección aleatoria uniforme de s es ergódica.

Ejercicio 2.13 *Diversas transiciones de proposición de cambio para la simulación de una $\mathcal{N}(0, 1)$ [43]*

El propósito de este ejercicio es poner en evidencia la influencia preponderante de la selección de la transición de proposición de cambio $q(x \rightarrow y)$ en la dinámica de Metropolis así como del valor inicial x_0 . Se quiere simular una distribución $\pi \sim \mathcal{N}(0, 1)$. Para esto, escogeremos como proposiciones de cambio :

- (i) $q(x, \cdot) \sim \mathcal{N}(x, 0,5)$ y $x_0 = -10$
- (ii) $q(x, \cdot) \sim \mathcal{N}(x, 0,1)$ y $x_0 = 0$
- (iii) $q(x, \cdot) \sim \mathcal{N}(x, 10)$ y $x_0 = 0$

La probabilidad de aceptación de y es la de la dinámica de Metropolis :

$$a(x, y) = \min\left\{1, \frac{\pi(y)q(x, y)}{\pi(x)q(y, x)}\right\}$$

Para cada una de las selecciones, representar gráficamente la trayectoria $\{x_t, 0 \leq t \leq 500\}$. Evaluar la probabilidad de que no haya cambio en el algoritmo de Metropolis. Probar la gaussianidad de las muestras y la rapidez para entrar en el régimen estacionario $\mathcal{N}(0, 1)$.

Ejercicio 2.14 *Comparación de las varianzas de las medias para dos dinámicas de M.H.*

Hemos visto que para dos dinámicas P y Q de M.H. t.q. $P \succ Q$,

$$\text{Var}_P\left(\frac{1}{T} \sum_{t=0}^T f(X_t)\right) \leq \text{Var}_Q\left(\frac{1}{T} \sum_{t=0}^T f(X_t)\right)$$

para $T = 1$ y cuando $T \rightarrow \infty$. Como lo muestra el ejemplo siguiente [100], esto no es cierto en general para otros valores de T .

Sean las dos transiciones simétricas en el espacio E de 4 estados :

$$P = \begin{pmatrix} 0 & ,2 & ,8 & 0 \\ ,2 & 0 & 0 & ,8 \\ ,8 & 0 & ,2 & 0 \\ 0 & ,8 & 0 & ,2 \end{pmatrix} \text{ y } Q = \begin{pmatrix} ,1 & ,1 & ,8 & 0 \\ ,1 & ,1 & 0 & ,8 \\ ,8 & 0 & ,2 & 0 \\ 0 & ,8 & 0 & ,2 \end{pmatrix}$$

P y Q son π -reversibles para $\pi = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ y $P \succ Q$. Para $f = {}^t(1, -1, -3, 3)$, verificar que

$$\text{Var}_R(f(X_0) + f(X_1) + f(X_2)) = \begin{cases} 15,4 & \text{para } R = P \\ 14,8 & \text{para } R = Q \end{cases}$$

Ejercicio 2.15 *Proceso Puntual : un caso donde Q no es irreducible*

En $[0, 1]^2$, y para $n(x) \equiv 2$, determinar para un proceso de núcleo duro el valor maximal r_0 de r . Demostrar que para $\varepsilon > 0$ pequeño y $r = r_0 - \varepsilon$, la transición de proposición q propuesta para la simulación de un P.P. no es irreducible (proponer dos configuraciones de dos puntos que no se puedan comunicar entre ellas).

Ejercicio 2.16 *Dinmica de Metropolis para el P.P. de Strauss*

Programar la simulación por medio de la dinámica de Metropolis de un proceso de Strauss con $n(x) = n = 50$ fijo (sólo el parámetro b (o γ) influye) y para $r = 0,05$. Examinar las configuraciones de las simulaciones para $\gamma = 0,01, 0,5, 1, 2, 10$.

Capítulo 3

Cadena inhomogénea y recocido simulado

Una *cadena de Markov inhomogénea* sobre un espacio de estados finito E es un proceso de Markov cuya probabilidad de transición en el instante k depende de k

$$P(X_{k+1} = j \mid X_k = i) = P_k(i, j)$$

Una cadena inhomogénea está, por lo tanto, caracterizada por su distribución inicial ν y la sucesión de sus transiciones $(P_k)_{k \geq 0}$. La distribución de (X_0, X_1, \dots, X_n) es

$$P(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) = \nu(x_0) \prod_{k=0}^{n-1} P_k(x_k, x_{k+1})$$

la distribución de X_n está dada por el producto matricial $\nu P_0 P_1 \dots P_{n-1}$.

Utilizando la noción de *coeficiente de contracción* de una transición P (Dobrushin [26]; [92], [47], [106]), comenzaremos por establecer criterios de ergodicidad para una cadena de Markov inhomogénea.

A continuación presentaremos el método y los resultados de *Recocido Simulado*. Mostraremos que a un problema de optimización general:

$$\text{Optimizar } U : E \longrightarrow \mathbb{R}$$

podemos asociar canónicamente, para un esquema de temperatura (T_k) que converge a 0, y una dinámica dada, una cadena de Markov inhomogénea que se concentra en los estados que realizan el máximo de U . La búsqueda del máximo de U está de esta manera ligado al estudio de la ergodicidad de esta cadena inhomogénea.

3.1. Coeficiente de contracción de Dobrushin

3.1.1. Preliminares

Comencemos por establecer algunas propiedades preliminares. Sea μ una medida sobre E . La *norma de variación* de μ está definida por

$$\|\mu\| \triangleq \sum_{x \in E} |\mu(x)| = \|\mu\|_1$$

Tenemos la propiedad

Proposición 3.1 Sean μ y ν dos probabilidades sobre E . Entonces:

$$(1) \|\mu - \nu\| = 2 \sum_x (\mu(x) - \nu(x))^+ = 2 \sup_{A \subset E} (\mu(A) - \nu(A)) \leq 2.$$

$$(2) \|\mu - \nu\| = 2(1 - \sum_x \inf\{\mu(x), \nu(x)\})$$

$$(3) \|\mu - \nu\| = \sup_{h: |h| \leq 1} |\sum_x h(x)(\mu(x) - \nu(x))|$$

En particular, dos probabilidades son disjuntas si y solamente si $\|\mu - \nu\| = 2$.

Demostración:

(1) Sea $A = \{x : \mu(x) \geq \nu(x)\}$, $B = E \setminus A$. μ y ν de igual masa,

$$\sum_{x \in A} (\mu(x) - \nu(x)) = - \sum_{x \in B} (\mu(x) - \nu(x)) = \sum_{x \in E} (\mu(x) - \nu(x))^+ = \frac{1}{2} \|\mu - \nu\|$$

(2)

$$\frac{1}{2} \|\mu - \nu\| = \sum_{x \in A} (\mu(x) - \nu(x)) = \sum_{x \in E} \mu(x) - \sum_{x \in B} \mu(x) - \sum_{x \in A} \nu(x) = 1 - \sum_{x \in E} \inf\{\mu(x), \nu(x)\}$$

(3)

$$\|\mu - \nu\| = \sum_{x \in E} |\mu(x) - \nu(x)| \geq \max_{|h| \leq 1} \left| \sum_{x \in E} h(x)(\mu(x) - \nu(x)) \right|$$

Para la igualdad, escogemos $h(x) = \text{sign}(\mu(x) - \nu(x))$. □

La oscilación de una función $f : E \rightarrow \mathbb{R}$ se define como:

$$\delta(f) = \sup_{x, y \in E} |f(x) - f(y)| = \sup_E f - \inf_E f$$

Sea $\mu(f) = \sum_x f(x)\mu(x)$.

Proposición 3.2 Sean μ y ν dos probabilidades sobre E , $f : E \rightarrow \mathbb{R}$. Entonces,

$$|\mu(f) - \nu(f)| \leq \frac{1}{2} \delta(f) \|\mu - \nu\|$$

Demostración: Sea m un real fijo. Como μ y ν tienen igual masa total, tenemos:

$$|\mu(f) - \nu(f)| = |\mu(f - m) - \nu(f - m)| \leq \|\mu - \nu\| \times \max_{x \in E} |f(x) - m|$$

Basta entonces verificar que en el término de la acotación, $\max_{x \in E} |f(x) - m|$ se alcanza para $m = \frac{1}{2}(\sup_E f + \inf_E f) = \frac{1}{2}\delta(f)$. □

3.1.2. Coeficiente de contracción y ergodicidad

Coeficiente de contracción.

Sea P una probabilidad de transición sobre E . El coeficiente de contracción de P se define por

$$c(P) = \frac{1}{2} \max_{x, y \in E} \|P(x, \cdot) - P(y, \cdot)\| \quad (3.1)$$

Se tiene que $0 \leq c(P) \leq 1$. $c(P) = 1$ si existe $x, y \in E$ tales que $P(x, \cdot)$ y $P(y, \cdot)$ son disjuntas (es decir, tienen soportes disjuntos). Finalmente, $c(P) = 0$ significa que todas las proyecciones

$P(x, \cdot)$ de P son iguales entre si, e iguales a una distribución $\pi : P$ es la transición independiente de distribución π .

La proposición siguiente da dos expresiones más de $c(P)$. Denotemos $Ph(x) = P(x, \cdot)h = \sum_{z \in E} P(x, z)h(z)$,

Proposición 3.3 *Dos expresiones de $c(P)$.*

$$(1) \quad c(P) = 1 - \min_{x,y} \sum_z \inf\{P(x, z), P(y, z)\}.$$

$$(2) \quad c(P) = \frac{1}{2} \max_{x,y} \max_{h:|h| \leq 1} |Ph(x) - Ph(y)|$$

Demostración: Para (1) (resp. (2)) utilizamos la descripción (2) (resp. (3)) de la proposición 3.1 que describe la norma de variación. \square

En particular, si $\varepsilon(P) = \inf_{x,z} P(x, z)$ y $r = \text{card}(E)$, tenemos la acotación:

$$c(P) \leq 1 - r\varepsilon(P) \quad (3.2)$$

Esta acotación es sencilla e importante desde el punto de vista teórico. Por otro lado tenemos las propiedades de contracción siguientes,

Proposición 3.4 *Sean μ, ν dos distribuciones sobre E y P, Q dos transiciones sobre E .*

$$(1) \quad \|\nu P - \mu P\| \leq c(P) \|\nu - \mu\|$$

$$(2) \quad c(PQ) \leq c(P)c(Q)$$

Demostración:

(1) $\|\nu P - \mu P\| = \max_{|h| \leq 1} |(\nu P)h - (\mu P)h|$. Pero:

$$(\mu P)h = \sum_z (\mu P)(z)h(z) = \sum_{z,x} \mu(x)P(x, z)h(z) = \sum_x \mu(x) \sum_z P(x, z)h(z) = \mu(Ph)$$

Así,

$$\begin{aligned} \|\nu P - \mu P\| &= \max_{|h| \leq 1} |\nu(Ph) - \mu(Ph)| \\ &\leq \|\nu - \mu\| \times \max_{|h| \leq 1} \left\{ \frac{1}{2} \max_{x,y} |Ph(x) - Ph(y)| \right\} = c(P) \|\nu - \mu\| \end{aligned}$$

(2)

$$\begin{aligned} c(PQ) &= \frac{1}{2} \max_{x,y} \|PQ(x, \cdot) - PQ(y, \cdot)\| = \frac{1}{2} \max_{x,y} \|P(x, \cdot)Q - P(y, \cdot)Q\| \\ &\leq \frac{1}{2} \max_{x,y} \|P(x, \cdot) - P(y, \cdot)\| c(Q) = c(P)c(Q) \quad (\text{punto (1)}) \end{aligned}$$

\square

El siguiente resultado controla la oscilación de Pf y compara los valores propios de P en $c(P)$.

Proposición 3.5 *Oscilación de Pf , valores propios de P y $c(P)$.*

$$(1) \quad \delta(Pf) \leq c(P)\delta(f)$$

$$(2) \quad \text{Sea } \lambda \neq 1 \text{ un valor propio de } P. \text{ Entonces, } |\lambda| \leq c(P) \leq 1 - r\varepsilon(P)$$

Demostración:

(1) Para dos distribuciones μ y ν , $|\mu(f) - \nu(f)| \leq \frac{1}{2}\delta(f)\|\mu - \nu\|$. Escogiendo $\mu = P(x, \cdot)$ y $\nu = P(y, \cdot)$, obtenemos:

$$|(Pf)(x) - (Pf)(y)| \leq \frac{1}{2}\delta(f)\|P(x, \cdot) - P(y, \cdot)\|$$

Basta tomar el máximo en x, y para obtener (1).

(2) Sea λ un valor propio de P , $f \neq 0$ un valor propio asociado, $Pf = \lambda f$. f no tiene coordenadas constantes si $\lambda \neq 1$: en efecto, si f tuviese coordenadas constantes, como P es una matriz de transición, $Pf = f = \lambda f$, es decir $\lambda = 1$. Para este f , utilizamos la acotación de (1):

$$\max_{x,y} |Pf(x) - Pf(y)| = |\lambda| \max_{x,y} |f(x) - f(y)| \leq c(P)\delta(f)$$

Como $\delta(f) \neq 0$, obtenemos el resultado anunciado. \square

Sea \mathbb{P} una probabilidad sobre un espacio medible (E, \mathcal{E}) , \mathcal{F} y \mathcal{G} dos sub σ -álgebras de \mathcal{E} . El coeficiente de α -mezcla entre \mathcal{F} y \mathcal{G} está definido por ([28]),

$$\alpha(\mathcal{F}, \mathcal{G}) = \sup\{|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|, A \in \mathcal{F}, B \in \mathcal{G}\}$$

La proposición siguiente permite controlar la mezcla y la covarianza de la cadena de transiciones (P_n) :

Proposición 3.6 *Estimaciones de mezcla y de la covarianza de la cadena. Sea X una cadena de distribución inicial μ y de transición P , $f : E \rightarrow \mathbb{R}$.*

(1) Para todo $0 \leq k < m$, y todo μ , tenemos: $\alpha(X_k, X_m) \leq 2c(P^{k,m})$.

(2) Para todo $0 \leq k < m$, y todo μ , tenemos: $|Cov(f(X_k), f(X_m))| \leq 2\delta^2(f)c(P^{k,m})$.

(3) En particular, si la cadena es homognea, de transición P , tenemos:

$$Var\left[\frac{1}{N} \sum_{t=0}^{N-1} f(X_t)\right] \leq \frac{\delta^2(f)}{N} \frac{1+c(P)}{1-c(P)}$$

Demostración:

(1) Llamemos P a la transición $P^{k,m}$ y ν a la distribución de X_k . Sean A y B dos subconjuntos de E : $\mathbb{P}(A \cap B) = \sum_{i \in A, j \in B} \nu_i P_{i,j}$, $\mathbb{P}(A) = \sum_{i \in A} \nu_i$ y $\mathbb{P}(B) = \sum_{l \in E, j \in B} \nu_l P_{l,j}$. Obtenemos:

$$\Delta = \mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B) = \sum_{i \in A} \nu_i \left\{ \sum_{j \in B} (P_{i,j} - \sum_{l \in E} \nu_l P_{l,j}) \right\}$$

Escribiendo $P_{i,j} = (\delta_{\{i\}}P)_j$, se obtiene fácilmente la acotación

$$|\Delta| \leq \sum_{i \in A} \nu_i \sum_{j \in B} |(\delta_{\{i\}}P - \nu P)_j| \leq \sum_{i \in E} \nu_i \sum_{j \in E} \|\delta_{\{i\}}P - \nu P\| \leq 2c(P)$$

(2) Llamemos νf y νPf las esperanzas de $f(X_k)$ y de $f(X_m)$. Tenemos:

$$\begin{aligned} Cov(f(X_k), f(X_m)) &= \sum_{i,j \in E} \nu_i P_{i,j} (f_i - \nu f)(f_j - \nu Pf) \\ &= \sum_{i \in E} \nu_i (f_i - \nu f) \{P(i, \cdot)(f - \nu Pf)\} \\ &= \sum_{i \in E} \nu_i (f_i - \nu f) \{P(i, \cdot)(f - \nu Pf) - P(l, \cdot)(f - \nu Pf)\}, \text{ de donde} \end{aligned}$$

$$|Cov(f(X_k), f(X_m))| \leq 2 \sum_{i \in E} \nu_i |f_i - \nu f| \delta(f)c(P) \leq 2\delta^2(f)c(P)$$

(3) Basta desarrollar la varianza:

$$\begin{aligned} \text{Var}\left[\frac{1}{N} \sum_{t=0}^{N-1} f(X_t)\right] &= \frac{1}{N^2} \left[\sum_{i=0}^{N-1} \text{Var}(f(X_t)) + 2 \sum_{0 \leq t < t' \leq N-1} \text{cov}(f(X_t), f(X_{t'})) \right] \\ &\leq \frac{1}{N^2} [N\delta^2(f) + 2\delta^2(f) \sum_{k=1}^{\infty} (N-k)c(P)^k] \leq \frac{\delta^2(f)(1+c(P))}{N(1-c(P))} \end{aligned}$$

□

Ergodicidad débil y ergodicidad fuerte

Llamemos $P^{m,k} = P_m P_{m+1} \cdots P_{k-1}$ la transición de m a k .

• La cadena de Markov de transiciones $(P_k)_{k \geq 0}$ es *débilmente ergódica* si

$$\forall m \geq 0, \forall \nu_1, \nu_2, \lim_{k \rightarrow \infty} \left\| \nu_1 P^{m,k} - \nu_2 P^{m,k} \right\| = 0 \quad (3.3)$$

•• La cadena de Markov de transiciones $(P_k)_{k \geq 0}$ es *fuertemente ergódica* si

$$\exists \pi_{\infty} \text{ t.q.: } \forall m \geq 0, \lim_{k \rightarrow \infty} \left(\sup_{\nu} \left\| \nu P^{m,k} - \pi_{\infty} \right\| \right) = 0 \quad (3.4)$$

La propiedad de ergodicidad débil dice que para todo m hay *pérdida de memoria de la distribución de X_m* . La propiedad de ergodicidad fuerte dice además que la distribución de la cadena se estabiliza alrededor de una distribución π_{∞} para k grande.

Ergodicidad de cadenas homogéneas

Veremos en el párrafo siguiente que si la cadena (P_n) admite una sucesión de distribuciones invariantes (π_n) ($\pi_n P_n = \pi_n$) que verifiquen $\sum_{n>0} \|\pi_n - \pi_{n+1}\| < \infty$, entonces la ergodicidad débil implica la ergodicidad fuerte. En particular, *si la cadena es homogénea* de transición P y si P admite una distribución invariante π , las nociones de *ergodicidad débil y fuerte coinciden*. A modo de ilustración, mostremos como la ergodicidad de una cadena homogénea puede obtenerse a partir del coeficiente de contracción. Llamemos $[x]$ a la parte entera de un real x , r es el cardinal de E , $\varepsilon(P) = \inf_{x,y} P(x,y)$.

Proposición 3.7 *Ergodicidad de una cadena homogénea de transición P .*

(1) $c(P^n)$ es decreciente.

(2) Si P es regular, c.a.d. y existe m t.q. $P^m > 0$, entonces

$$c(P^n) \leq \{1 - r\varepsilon(P^m)\}^{\lfloor \frac{n}{m} \rfloor} \rightarrow 0 \text{ si } n \rightarrow \infty$$

Este resultado es una consecuencia de la propiedad de sub-multiplicatividad $c(PQ) \leq c(P)c(Q)$, de la identidad $P^{n+k} = P^n P^k$ y de la división euclidiana $n = mq + s$, $0 \leq s < m$, del entero n por m . Si P es regular, existe $\pi = \pi_{\infty}$ invariante para P . En consecuencia,

$$\|\nu P^n - \pi_{\infty}\| = \|\nu P^n - \pi_{\infty} P^n\| \leq \|\nu - \pi_{\infty}\| c(P^n) \leq 2c(P^n) \rightarrow 0$$

La cadena es ergódica.

Ergodicidad de cadenas inhomogéneas ([26];[53];[106])

Sea la condición:

$$\forall m, \lim_{k \rightarrow \infty} c(P_m P_{m+1} \cdots P_k) = 0 \quad (3.5)$$

Proposición 3.8 *Criterio de ergodicidad dbil.* Una cadena no-homognea es dbilmente erg-dica si y solamente si ella verifica (3.5).

Demostración:

- (\Leftarrow) Para todo m y para cualesquiera distribuciones μ, ν , $\|\mu P^{m,k} - \nu P^{m,k}\| \leq \|\mu - \nu\| c(P^{m,k})$. La ergodicidad débil es consecuencia de la condición $\lim_{k \rightarrow \infty} c(P^{m,k}) = 0$.
- (\Rightarrow) Recíprocamente, escojamos como distribuciones iniciales $\mu = \delta_{\{i\}}, \nu = \delta_{\{j\}}, i \neq j$. $\delta_{\{i\}} P^{m,k}$ no es más que la i -ésima fila de $P^{m,k}$. Así,

$$\left\| \delta_{\{i\}} P^{m,k} - \delta_{\{j\}} P^{m,k} \right\| = \sum_{l \in E} \left| P^{m,k}(i, l) - P^{m,k}(j, l) \right|,$$

y

$$c(P^{m,k}) = \frac{1}{2} \sup_{i,j} \left\| \delta_{\{i\}} P^{m,k} - \delta_{\{j\}} P^{m,k} \right\| \rightarrow 0 \text{ si } k \rightarrow \infty$$

□

Daremos ahora una condición *suficiente* de ergodicidad fuerte. Sea (π_n) una sucesión de distribuciones sobre E que verifica

$$\sum_{n>0} \|\pi_n - \pi_{n+1}\| < \infty \quad (3.6)$$

(π_n) es una sucesión de Cauchy sobre el espacio completo $(\mathbb{R}^E, \|\cdot\|_1)$: existe, por lo tanto, una distribución π_∞ tal que $\pi_n \rightarrow \pi_\infty$.

Sea $X = (X_n)_{n \geq 0}$ una cadena de transiciones (P_n) . Tenemos el siguiente resultado de ergodicidad,

Proposición 3.9 *Criterio de ergodicidad fuerte.* Supongamos que existe una sucesin de distribuciones (π_n) invariantes para (P_n) , tales que (π_n) verifica (3.6) y que la sucesin (P_n) verifica (3.5). Entonces existe una distribucin π_∞ tal que

$$\sup_{\nu} \|\nu P_0 P_1 \cdots P_n - \pi_\infty\| \rightarrow 0$$

Demostración: Ya hemos establecido la existencia de π_∞ .

(1) Comencemos por controlar $A_{i,k} = \|\pi_\infty P_i P_{i+1} \cdots P_{i+k} - \pi_\infty\|$. Tenemos:

$$\begin{aligned} \pi_\infty P_i P_{i+1} \cdots P_{i+k} - \pi_\infty &= (\pi_\infty - \pi_i) P_i P_{i+1} \cdots P_{i+k} + \pi_i P_i P_{i+1} \cdots P_{i+k} - \pi_\infty \\ &= (\pi_\infty - \pi_i) P_i P_{i+1} \cdots P_{i+k} \\ &\quad + \sum_{j=1}^k (\pi_{i-1+j} - \pi_{i+j}) P_{i+j} \cdots P_{i+k} + (\pi_{i+k} - \pi_\infty) \end{aligned}$$

de donde deducimos la acotación

$$A_{i,k} \leq 2c(P_i P_{i+1} \cdots P_{i+k}) + \sum_{n=i}^{\infty} \|\pi_n - \pi_{n+1}\| + \|\pi_{i+k} - \pi_\infty\|$$

En consecuencia, para todo $\varepsilon > 0$, podemos escoger i y k tales que $A_{i,k} \leq \varepsilon$.

(2) De la descomposición, para todo $i < n$:

$$\nu P_0 P_1 \cdots P_n - \pi_\infty = (\nu P_0 P_1 \cdots P_{i-1} - \pi_\infty) P_i P_{i+1} \cdots P_n + \pi_\infty P_i P_{i+1} \cdots P_n - \pi_\infty$$

deducimos,

$$\|\nu P_0 P_1 \cdots P_n - \pi_\infty\| \leq 2c(P_i P_{i+1} \cdots P_n) + A_{i,k} \text{ para } k = n - i$$

La ergodicidad resulta entonces de (3.5) y de (3.6). \square

Nota 3.1 .

(1) El control en n de la velocidad de ergodicidad, es decir de $\sup_\nu \|\nu P_0 P_1 \cdots P_n - \pi_\infty\|$ en n se obtiene a partir del de $c(P_n P_{n+1} \cdots)$ y del resto de la serie $\sum_{k \geq n} \|\pi_k - \pi_{k+1}\|$ [106].

(2) Una condición suficiente, muy utilizada en la práctica, que garantiza que $\sum_n \|\pi_n - \pi_{n+1}\| < \infty$ es:

$$(M) : \forall x \in E, n \rightarrow \pi_n(x) \text{ es monótona a partir de un cierto rango} \quad (3.7)$$

En efecto, $\sum_n \|\pi_n - \pi_{n+1}\| = 2 \sum_x \sum_n |\pi_n(x) - \pi_{n+1}(x)|^+$. Como solo hay un número finito de x , basta garantizar la convergencia de cada serie numérica con x fijo. La condición de monótona (3.7) asegura esta convergencia: en efecto, para n suficientemente grande, $|\pi_n(x) - \pi_{n+1}(x)|^+ = \pi_n(x) - \pi_{n+1}(x)$ si $(\pi_n(x))$ es decreciente, = 0 si no.

(3) Tres condiciones suficientes que garantizan la ergodicidad débil (3.5) son:

$$(3.5-1): \forall m \geq 0, \prod_{n \geq m} c(P_n) = 0$$

$$(3.5-2): \forall n, c(P_n) > 0 \text{ y } \prod_{n \geq 0} c(P_n) = 0$$

$$(3.5-3): \forall n, \exists k \geq n \text{ tal que } c(P_k) = 0.$$

(4) Una técnica clásica para controlar $c(P_i P_{i+1} \cdots P_{i+k})$ es descomponer el intervalo $[i, i+k+1[$ en l intervalos sucesivos y adyacentes $I_s = [\tau_{s-1}, \tau_s[$, $s = 1, l$. Llamando $P^{(s)} = P_{\tau_{s-1}} P_{\tau_{s-1}+1} \cdots P_{\tau_s-1}$, obtenemos:

$$c(P_i P_{i+1} \cdots P_{i+k}) \leq \prod_{s=1}^l c(P^{(s)})$$

Si además escogemos τ_s de modo que $P^{(s)} > 0$, tenemos $c(P^{(s)}) \leq (1 - r\varepsilon(P^{(s)}))$. Esta técnica es fácil de usar pero está lejos de ser óptima.

(5) La condición (3.5): $\forall m \geq 0, c(P_m P_{m+1} \cdots P_n) \rightarrow 0$, para todos los m , es necesaria (cf. ejercicio).

Cadenas inhomogéneas: Ley de grandes números y T.L.C.

Sea $X = (X_n)$ una cadena inhomogénea, de transiciones (P_n) . Si la distribución inicial es ν , llamaremos \mathbb{P}_ν a la distribución de la cadena. Llamemos además $c_n = \sup_{1 \leq i \leq n} c(P_i)$. Tenemos los dos resultados siguientes:

Proposición 3.10 Ley de Grandes Números (Gantert [35]; [52])

Suponemos que la cadena es fuertemente ergódica y converge a π_∞ . Sea $f : E \rightarrow \mathbb{R}$. Entonces,

$$\frac{1}{n} \sum_{i=1, n} f(X_i) \rightarrow \pi_\infty(f)$$

(i) en $L^2(\mathbb{P}_\nu)$ si $\lim_n n(1 - c_n) = \infty$.

(ii) \mathbb{P}_ν -c.s. si $\sum_{n \geq 1} \frac{1}{2^n(1-c_{2n})^2} < \infty$.

Proposición 3.11 *Teorema Central del Limite (Dobrushin [26]).*

Sean $f_i : E \rightarrow \mathbb{R}$, y $S_n = \sum_{i=1,n} f_i(X_i)$. Suponemos que:

- (i) Existe $c > 0$ tal que, para todo i , $\text{Var}(f_i(X_i)) \geq c$
- (ii) $\sup_i \|f_i\|_\infty < \infty$ y
- (iii) $\lim_n n^{\frac{1}{3}}(1 - c_n) = +\infty$. Entonces:

$$\frac{S_n - E(S_n)}{\text{Var}(S_n)^{\frac{1}{2}}} \rightarrow \mathcal{N}(0, 1)$$

La constante $\frac{1}{3}$ de (iii) es óptima.

Para estos dos resultados, el caso difícil es aquel en el cual $c_n \rightarrow 1_-$. Supongamos por ejemplo que $c_n = 1 - \frac{a_n}{n}$, $0 < a_n < n$ y $a_n \rightarrow \infty$. Para la ley de grandes números, la convergencia tiene lugar en $L^2(\mathbb{P}_\nu)$ y \mathbb{P}_ν -c.s. si $\sum_n a_n^{-2} < \infty$. El TCL vale si $a_n n^{-\frac{2}{3}} \rightarrow +\infty$.

3.2. Optimización por recocido simulado

Sea $U : E \rightarrow \mathbb{R}$ una función a minimizar sobre E finito, $U_* = \min\{U(x) : x \in E\}$ y $E_* = \{x \in E : U(x) = U_*\}$. Sea $\beta = T^{-1}$ el parámetro inverso de una temperatura $T > 0$ ($T \rightarrow 0$ equivale a $\beta \rightarrow \infty$), y π_β la distribución sobre E asociada a U y a β :

$$\pi_\beta(x) = Z^{-1}(\beta) \exp -\beta U(x).$$

La heurística del algoritmo de *recocido simulado* (que denotaremos de ahora en adelante por RS) es la siguiente:

$$\text{Si } U(a) < U(b), \quad \lim_{\beta \rightarrow +\infty} \frac{\pi_\beta(b)}{\pi_\beta(a)} = 0.$$

A baja temperatura (β grande), la distribución π_β se concentra en E_* , el conjunto de los mínimos de U . Más precisamente, si llamamos π_∞ a la distribución uniforme sobre E_* ,

$$\pi_\infty(x) = |E_*|^{-1} \mathbf{1}(x \in E_*), \text{ entonces: } \|\pi_\beta - \pi_\infty\| \rightarrow 0 \text{ si } \beta \rightarrow \infty$$

($|A|$ es el cardinal de A). Así, a primera vista, optimizar U equivale a simular π_β para β grande. Se presentan dos problemas:

- (i) ¿Cuándo consideramos que β es grande ?
- (ii) la convergencia de la simulación de π_β para β grande es muy lenta.

El algoritmo de RS es una respuesta a estos problemas (Kirkpatrick, Gelatt y Vecchi [65]): vamos a utilizar una dinámica (P_k) inhomogénea en el tiempo, relativa a una sucesión (β_k) que tiende al infinito y buscaremos condiciones suficientes que aseguren que la cadena (P_k) es fuertemente ergódica y converge a π_∞ . Examinaremos dos casos:

- El caso de un espacio producto E , con la dinámica del muestreador de Gibbs ([37]; [39]);
- El caso de un espacio de estados E general y de la dinámica de Metropolis-Hastings (Hajek [50]; [103]; [1]; [42]; [5]).

3.2.1. Recocido simulado para la dinámica de Gibbs

$E = F^S$ donde $S = \{1, 2, \dots, n\}$ y F es finito. Sea $U : E \rightarrow \mathbb{R}$. Sin pérdida de generalidad, supondremos que el mínimo de U es $U_* = 0$ (por lo tanto U toma valores en \mathbb{R}^+ , y para al menos un $x \in E$, $U(x) = 0$). La dinámica del muestreador de Gibbs está asociada a la distribución

$$\pi_\beta(x) = Z^{-1}(\beta) \exp -\beta U(x)$$

y a sus distribuciones condicionales $\pi_{i,\beta}(x_i | x^i)$, $i \in S$. El algoritmo es el siguiente:

Algoritmo de RS para la dinámica de Gibbs

- (i) Escogemos un esquema $(\beta_k) \uparrow +\infty$.
- (ii) Visitas a S periódicas, $\beta = \beta_k$ en el k -ésimo barrido.
- (iii) La transición en el k -ésimo barrido es:

$$P_k(x, y) = \prod_{i=1}^n \pi_{i,\beta_k}(y_i | y_1, \dots, y_{i-1}, x_{i+1}, \dots, x_n)$$

Sea $\Delta = \max\{U(x) - U(y) : x \equiv y \text{ excepto en un lugar}\}$, r el cardinal de E .

Proposición 3.12 *Convergencia del RS para la dinámica de Gibbs (Geman D. y Geman S. [37];[39]).*

Si $(\beta_k) \uparrow +\infty$ y verifica

$$\sum_{k \geq 0} \exp\{-n\Delta\beta_k\} = +\infty$$

entonces: $\forall x_0, \lim_{k \rightarrow \infty} P(X_k \in E_* | X_0 = x_0) = 1$. En otras palabras, para toda distribución inicial ν , $\nu P_0 P_1 \cdots P_k \rightarrow \pi_\infty$. Si $\beta_k = \gamma \log k$, esta convergencia tiene lugar si $\gamma \leq (n\Delta)^{-1}$.

Demostración: $\pi_k = \pi_{\beta_k}$ es invariante para P_k y $\pi_k \rightarrow \pi_\infty$. Basta verificar las dos condiciones (3.6) y (3.5). Observamos que la convergencia $\pi_k \rightarrow \pi_\infty$ no permite obtener (3.6).

(i) Verifiquemos que para todo x , la sucesión $(\pi_k(x))$ es monótona para k grande. Fijemos x . Tenemos:

$$\pi_\beta(x) = \left\{ \sum_{y \in E} \exp\{-\beta[U(y) - U(x)]\} \right\}^{-1}$$

Pongamos $a(y) = U(y) - U(x)$. Si $x \in E_*$, $a(y) \geq 0$, cada $\exp\{-\beta_k a(y)\}$ es decreciente y en consecuencia $(\pi_k(x))$ es creciente. Si no,

$$d(\beta) = \pi_\beta(x)^{-1} = |\{y \in E : U(y) = U(x)\}| + \sum_{a(y) < 0} e^{-\beta a(y)} + \sum_{a(y) > 0} e^{-\beta a(y)}$$

La derivada de d es $d'(\beta) = -\sum_{a(y) < 0} a(y)e^{-\beta a(y)} - \sum_{a(y) > 0} a(y)e^{-\beta a(y)}$. Si U no es constante, existe y t.q. $a(y) < 0$ y el primer término tiende a $-\infty$ si $\beta \rightarrow \infty$. En cuanto al segundo término, tiende a 0. Así si $x \notin E_*$, $(\pi_k(x))$ es decreciente a partir de un cierto valor.

(ii) $\pi_{i,\beta}(x_i | x^i) = \left\{ \sum_{u \in F} \exp\{-\beta[U(u, x^i) - U(x_i, x^i)]\} \right\}^{-1} \geq (re^{\beta\Delta})^{-1}$ ya que $[U(u, x^i) - U(x_i, x^i)] \geq -\Delta$. Deducimos que:

$$P_\beta(x, y) \geq (r^n e^{n\beta\Delta})^{-1} \text{ y en consecuencia } c(P_\beta) \leq 1 - e^{-n\beta\Delta}$$

La condición suficiente (ii) vale si $\forall m, \prod_{k \geq m} c(P_k) = 0$, condición equivalente a

$$\sum_k (1 - c(P_k)) = +\infty$$

Para el esquema de temperaturas propuesto, $1 - c(P_k) \geq e^{-n\beta_k\Delta} = (\frac{1}{k})^{\gamma n\Delta}$. La divergencia tiene lugar si $\gamma n\Delta \leq 1$. □

Ejemplo 3.1 *Suavizamiento de una curva por minimización de la rugosidad*

Queremos reconstruir una curva $x : [0, 1] \rightarrow F \subset \mathbb{R}$, donde F es una malla finita de \mathbb{R} . Tenemos observaciones $\{y_i, i = 0, n\}$ que asociamos a x por el modelo:

$$y_i = x_i + \varepsilon_i, \quad i = 0, n \text{ donde } x_i = x\left(\frac{i}{n}\right)$$

La rugosidad de una curva x de clase \mathcal{C}^2 se define por $F(x) = \int_0^1 x''(t)^2 dt$. Su versión discretizada es $F_D(x) = \sum_{i=1}^{n-1} (x_{i-1} - 2x_i + x_{i+1})^2$: la rugosidad es débil si x es regular. Una reconstrucción discreta de x se puede hacer sobre la base de la energía

$$U : E = F^{n+1} \rightarrow \mathbb{R}, \quad U(x | y) = \gamma F_D(x) + \|y - x\|_2^2$$

γ es un parámetro de regularización: mientras más grande sea γ , más regular será la reconstrucción; mientras más pequeño sea γ , mayor será la fidelidad de x a y . Verificaremos que U se escribe de manera aditiva y local:

$$U(x | y) = \sum_i \Phi_i(x_i | y) + \sum_{i,j} \Phi_{i,j}(x_i, x_j | y)$$

$$\text{con } \begin{cases} \Phi_i(x_i | y) = x_i^2(6\gamma + 1) - 2x_i y_i \\ \Phi_{i,j}(x_i, x_j | y) = \Phi_{i,j}(x_i, x_j) = \begin{cases} -8\gamma x_i x_j & \text{si } |i - j| = 1 \\ 2\gamma x_i x_j & \text{si } |i - j| = 2 \\ 0 & \text{si no} \end{cases} \end{cases}$$

Las distribuciones condicionales se expresan a partir de una energía condicional:

$$U_{i,\beta}(x_i | x_{\partial i}, y_i) = -\beta x_i (x_i(6\gamma + 1) - 2y_i - 8\gamma(x_{i-1} + x_{i+1}) + 2(x_{i-2} + x_{i+2}))$$

$$\pi_{i,\beta}(x_i | x_{\partial i}, y) = \exp U_{i,\beta}(x_i | x_{\partial i}, y_i) \left\{ \sum_{u \in F} \exp U_{i,\beta}(u | x_{\partial i}, y_i) \right\}^{-1}$$

3.2.2. Recocido simulado para la dinámica de Metropolis

Estudiamos ahora RS para la dinámica de Metropolis, donde suponemos que la transición de proposición Q es simétrica. El caso general de las dinámicas de Metropolis-Hastings puede ser tratado de manera similar. A temperatura $T = \beta^{-1}$, la transición es, para $x \neq y$

$$P_\beta(x, y) = Q(x, y) \exp -\beta[U(y) - U(x)]^+$$

El algoritmo es el siguiente:

R.S. para la dinámica de Metropolis

- (i) escoger un esquema $(\beta_k) \uparrow +\infty$. Estamos en $X_k = x$.
- (ii) escoger una proposición de cambio y según $Q(x, \cdot)$.
- (iii) Si $\Delta U = U(y) - U(x) \leq 0$, tomar $X_{k+1} = y$.
- (iv) Si no, generar V con distribución uniforme sobre $[0, 1]$:
 - (iv-1) Si $V \leq \exp -\Delta U$, tomar $X_{k+1} = y$.
 - (iv-2) Si no tomar $X_{k+1} = x$.

Supondremos que Q es regular. Sean $M = \inf\{n \text{ t.q. } Q^n > 0\}$, $q = \inf\{Q(x, y) : x, y \text{ t.q. } Q(x, y) > 0\}$ y $\Delta = \max\{U(y) - U(x) : x, y \in \text{t.q. } Q(x, y) > 0\}$. Adoptaremos un esquema de enfriamiento por niveles, cada uno de longitud M , la temperatura permanece constante con valor β_k^{-1} sobre el k -ésimo nivel. Tenemos el siguiente resultado ([50], [42],[107])

Proposición 3.13 *Convergencia del RS para la dinámica de Metropolis.*

Si $(\beta_k) \uparrow +\infty$ verificando

$$\sum_{k \geq 0} \exp -\{M\Delta\beta_k\} = +\infty$$

entonces la cadena de RS para la dinámica de Metropolis realiza la búsqueda del mínimo de U , es decir $\lim_{k \rightarrow \infty} P(X_k \in E_* \mid X_0 = x_0) = 1$, siempre que

Si $\beta_k = \gamma \log k$, la convergencia del RS vale si $\gamma \leq (M\Delta)^{-1}$.

Demostración: Seguimos el mismo procedimiento que para establecer la convergencia del RS respecto al muestreador de Gibbs. Llamemos $\pi_\beta = Z^{-1}(\beta) \exp -\beta U$, $\pi_k = \pi_{\beta_k}$ y $R_k = P_{\beta_k}^M$. π_k es invariante para R_k y $\pi_k \rightarrow \pi_\infty$. La ergodicidad fuerte vale bajo las condiciones (3.6) y (3.5).

El punto (i) se establece como para el muestreador de Gibbs: el hecho de que E sea un espacio producto no interviene para establecer la monotonía de las $(\pi_n(x))$.

Falta controlar el coeficiente de contracción de R_β . $Q(x, y) \geq q > 0$ para los cambios posibles $x \mapsto y$. Q es regular con $Q^M > 0$, para $x, y \in E$, existe un camino $x = x(0) \mapsto x(1) \mapsto \dots \mapsto x(M) = y$ de longitud M que une x a y :

$$R_\beta(x, y) \geq q^M \exp -\beta \sum_{l=1, M} [U(x_l) - U(x_{l-1})]^+ \geq q^M \exp -M\beta\Delta$$

Tenemos entonces: $c(R_\beta) \leq 1 - rq^M \exp -M\beta\Delta$. De donde se obtiene la primera parte del resultado. Para el esquema $\beta_k = \gamma \log k$, $1 - c(P_k) \geq rq^M (\frac{1}{k})^{\gamma M\Delta}$. La serie de término general $1 - c(P_k)$ es divergente si $\gamma \leq (M\Delta)^{-1}$. \square

Los dos resultados precedentes de convergencia del RS en los dos contextos de la dinámica de Gibbs y de la dinámica de Metropolis dan *condiciones suficientes no-óptimas* de convergencia del recocido simulado. Obtener una condición necesaria y suficiente de ergodicidad fuerte sobre la sucesión (β_k) es mucho más difícil.

Describamos una condición de este tipo obtenida por Hajek [50] en el contexto de la dinámica de Metropolis. Suponemos que Q es simétrica y define la relación de Q -vecindad: $x \sim y \Leftrightarrow Q(x, y) > 0$. El conjunto E_*^{loc} de los mínimos locales es el conjunto de los x t.q. $U(x) \leq U(y)$ si $x \sim y$. La *profundidad* $d(x)$ de un mínimo local $x \in E_*^{loc}$ es

$$d(x) = \inf\{d > 0 : \exists y \text{ t.q. } U(y) < U(x) \text{ y } \exists x \rightsquigarrow y \text{ de nivel } \leq U(x) + d\}$$

donde $x \rightsquigarrow y$ es un camino que va de x a y : para salir del pozo local de U en x a fin de ir hacia y , mas bajo que x ($U(y) < U(x)$), podemos hallar un camino que no suba más que $U(x) + d$. Definimos entonces la profundidad máxima del paisaje de la energía U por:

$$D = \text{máx}\{d(x), x \in E_*^{loc} \setminus E_*\}$$

La condición necesaria y suficiente de convergencia del RS es:

Proposición 3.14 *Condición necesaria y suficiente de convergencia del RS (Hajek, [50])*

$$\lim_{k \rightarrow \infty} P(X_k \in E_* \mid X_0 = x_0) = 1 \iff \sum_{k \geq 0} \exp -\{D\beta_k\} = +\infty$$

Observamos que $D \leq M\Delta$. Para la sucesión $\beta_k = \gamma \log k$, el resultado de Hajek implica el resultado precedente porque $\gamma \leq (M\Delta)^{-1} \leq D^{-1}$.

Ejemplo 3.2 *Alizamiento de una curva (continuación)*

Retomemos el ejemplo 3.1. Supongamos que el espacio de estados F es el intervalo $[-5, +5]$ discretizado a pasos de ancho 0,1 (F tiene 101 elementos). Sea ε pequeño ($\varepsilon = 0,1$). La proporción de cambio $x \mapsto z$ es la siguiente:

- (i) decidimos no cambiar nada con probabilidad ε
- (ii) escogemos al azar $i \in S = \{0, 1, 2, \dots, n\}$ y cambiamos el valor $x_i \mapsto z_i$ en ese sólo lugar i en $\pm 0,1$ si $x_i \notin \{-5, +5\}$, o $z_i \in \{x_i \text{ o } x_i \pm 0,1 \text{ si } x_i = \mp 5\}$, esto con probabilidad $Q(x, z) = \frac{1-\varepsilon}{2n}$;
- (iii) $Q(x, z) = 0$ si no.

Q es simétrica, regular, $M \leq 101 \times (n + 1)$. Un cálculo directo de $\Delta U = U(z) - U(x)$, condicional a y , da:

$$\Delta U = \gamma(z_i - x_i)[(z_i + x_i)(6 + \gamma^{-1}) - 2y_i - 8(x_{i-1} + x_{i+1}) + 2(x_{i-2} + x_{i+2})]$$

Conservamos z_i con probabilidad 1 si $\Delta U \leq 0$, y si no con probabilidad $e^{-\beta_k \Delta U}$.

Ejemplo 3.3 *Un problema de recubrimiento*

Tenemos N segmentos de longitudes $l_i > 0, i = 1, N$ con los cuales queremos recubrir, sin que se superpongan ni sobrepasen, un intervalo de longitud $L < l_1 + l_2 + \dots + l_N$. Llamemos $S = \{1, 2, \dots, N\}$, $E = \{A \subset S : A \neq \emptyset \text{ y } \sum_{i \in A} l_i \leq L\}$ y $U(A) = \sum_{i \in A} l_i$. Recubrir de la mejor manera el intervalo $[0, L]$ corresponde a optimizar U sobre E . Vamos a describir el algoritmo de Metropolis para la transición de proposición $Q(A, B)$ siguiente:

- (i) Escoger s uniformemente en S .
- (ii) Los únicos cambios $A \mapsto B$ son: $\begin{cases} \text{Si } s \in A, \text{ pasar a } B = A \setminus \{s\} \\ \text{Si } s \notin A, \text{ pasar a } B = A \cup \{s\} \end{cases}$

Q es una transición irreducible (pasamos de A a B en a lo sumo $|A| + |B|$ pasos) y simétrica (si $s \notin A$ y $B = A \cup \{s\}$, $Q(A, B) = \frac{1}{N} = Q(B, A)$). Tenemos

$$\Delta U = U(A) - U(B) = \begin{cases} +l_s & \text{si } B = A \setminus \{s\} \\ -l_s & \text{si } B = A \cup \{s\} \\ 0 & \text{si no} \end{cases}$$

Así, a la k -ésima iteración del algoritmo, conservamos con seguridad a B si $B = A \cup \{s\}$, y con probabilidad $\exp -\beta_k l_s$ si $B = A \setminus \{s\}$.

Nota 3.2 *Algunos comentarios sobre el recocido simulado*

(1) La analogía con el recocido simulado de la metalurgia es justificado. Para el esquema de temperatura $T(k) = \frac{T_0}{\log k}$, la condición suficiente de ergodicidad es $T_0 \geq N\Delta$. La temperatura inicial debe ser suficientemente grande, todas las partículas están libres y desordenadas. Luego, el enfriamiento debe ser suficientemente lento (esquema en $(\log k)^{-1}$), y la cristalización tiene lugar, realizando el mínimo de energía.

(2) El RS permite siempre escaparse de un mínimo local contrariamente a los métodos numéricos clásicos de optimización, por ejemplo el método del gradiente: es una condición necesaria a la convergencia hacia un mínimo global. En contrapartida, uno escapa también del mínimo global y no puede, por lo tanto, haber convergencia casi segura a E_* : la convergencia del RS tiene lugar en probabilidad y no casi seguramente.

(3) El recocido simulado en horizonte finito K ha sido estudiado por Catoni [15]. El problema es el siguiente: disponemos de un presupuesto K fijo (el número de iteraciones autorizadas en

el algoritmo de RS), y deseamos retener el mejor esquema de enfriamiento triangular $\mathcal{T}^K = \{T_1^K > T_2^K > \dots > T_K^K\}$ que minimice:

$$P(\mathcal{T}^K) = \max_{x \in E} P(X_K \notin E_* \mid X_0 = x)$$

El resultado es el siguiente: el esquema óptimo \mathcal{T}^K es un esquema con decrecimiento exponencial y $P(\mathcal{T}^K) = O(K^{-d})$ donde d es una característica del paisaje de energía. Este resultado no es asintótico. No está en contradicción con el resultado de decrecimiento lento de la temperatura para la convergencia del RS. En cierto sentido, este resultado valida el hecho de que, en la práctica, los esquemas de enfriamiento utilizados son exponenciales, de tipo $T_k^K = T_0 \rho^k$ con T_0 suficientemente grande y ρ cercano a 1 ($\rho = 0,99$): se deja T_k^K constante por niveles de longitud $l_k, k = 1, K$ (l_k a determinar); un esquema de enfriamiento estará caracterizado por $(T_0, \rho, (l_k))$.

(4) Se puede obtener un control de $\|\nu P_1 P_2 \dots P_k - \pi_\infty\|$ a partir de la acotación que permite obtener la ergodicidad fuerte de una cadena inhomogénea

$$\begin{aligned} \|\nu P_1 P_2 \dots P_{i+k} - \pi_\infty\| &\leq A_{i,k} + 2c(P_i P_{i+1} \dots P_{i+k}) \text{ con} \\ A_{i,k} &= \|\pi_\infty P_1 P_2 \dots P_{i+k} - \pi_\infty\| \leq \|\pi_\infty - \pi_i\| + \sum_{n \geq i} \|\pi_{i+1} - \pi_i\| \end{aligned}$$

En el cuadro del RS para la dinámica de Gibbs sobre $E = F^S, S = \{1, 2, \dots, N\}$, para una función U de mínimo global 0 ($U_* = 0$) y para el esquema $\beta_k = \frac{\log k}{N\Delta}$, Winkler [106] obtiene el control siguiente, en función de $m = \inf\{U(x) : x \notin E_*\}$,

$$\|\nu P_1 P_2 \dots P_k - \pi_\infty\| = 0(k^{-\frac{m}{m+N\Delta}})$$

La velocidad mejora con la profundidad del mínimo global de U respecto al resto del paisaje y con U localmente regular fuera de E_* .

3.3. Simulación y optimización con restricciones

Sea $C : E \rightarrow \mathbb{R}$ una aplicación que define la *restricción* $E_C = \{x \in E : C(x) = C_*\}$ donde $C_* = \inf\{C(x), x \in E\}$. Supondremos que C no es constante.

Sea $U : E \rightarrow \mathbb{R}$ una energía, $\pi(x) = Z^{-1} \exp\{-U(x)\}$ la distribución sobre E asociada a U , $U_{*,C} = \inf\{U(x) : x \in E_C\}$, $E_{*,C} = \{x \in E_C : U(x) = U_{*,C}\}$. La distribución de energía U restringida a E_C es

$$\pi_C(x) = Z_C^{-1} \mathbf{1}\{x \in E_C\} \exp -U(x)$$

Sea $\pi_{\infty,C}$ la distribución uniforme sobre $E_{*,C}$. Vamos a responder las dos preguntas siguientes:

- (S): ¿Simular π_C ?
- (O): ¿Minimizar U sobre E_C ?

El enfoque intuitivo es el siguiente. Sean β y λ dos parámetros reales positivos y $\pi_{\beta,\lambda}$ la distribución sobre E de energía $U_{\beta,\lambda} = \beta\{U(x) + \lambda C(x)\}$:

$$\pi_{\beta,\lambda}(x) = Z^{-1}(\beta, \lambda) \exp -\beta\{U(x) + \lambda C(x)\}$$

Los siguientes dos resultados son fáciles de verificar:

- (S): Si $\beta = 1$ y si $\lambda \uparrow \infty$, entonces $\pi_{1,\lambda} \rightarrow \pi_C$.
- (O): Si $\beta \uparrow \infty$ y si $\lambda \uparrow \infty$, entonces $\pi_{\beta,\lambda} \rightarrow \pi_{C,\infty}$.

Llamemos $\pi_k = \pi_{\beta_k, \lambda_k}$. Este comportamiento asintótico sugiere que construyamos cadenas de Markov de transiciones (P_k) relacionadas π_k . Si (π_k) es la distribución invariante de (P_k) , la ergodicidad de las cadenas será cierta bajo (3.6) y (3.5). En el caso en el cual E es un conjunto producto, escogemos la dinámica del muestreador de Gibbs: los resultados se deben a los hermanos Geman [38] (cf. también [39]). En el caso de un espacio de estados finito general y de la dinámica de Metropolis, los resultados se deben a Yao [107].

Si restamos C_* a C y $U_{*,C}$ a U , la distribución $\pi_{\beta, \lambda}$ no cambia: así, sin pérdida de generalidad, supondremos que $C_* = U_{*,C} = 0$.

3.3.1. Simulación y optimización bajo restricciones para la dinámica del muestreador de Gibbs

$E = F^S$ donde F es finito y $S = \{1, 2, \dots, n\}$. Sean (β_k) y (λ_k) dos sucesiones que crecen hacia $+\infty$, $\pi_k = \pi(\beta_k, \lambda_k)$ y P_k la transición del muestreador de Gibbs de π_k para el barrido $1 \mapsto 2 \mapsto \dots \mapsto n$ de S . Llamemos $\Delta = \max\{U(x) - U(y) : x \equiv y \text{ salvo en un lugar}\}$, $\Gamma = \max\{C(x) - C(y) : x \equiv y \text{ salvo en un lugar}\}$. Δ y Γ son > 0 . Tenemos el resultado siguiente:

Proposición 3.15 *Simulación y optimización bajo restricciones (dinámica de Gibbs, Geman D. y Geman S., [38]; [39]).*

(1) *Simulación:* fijemos para todo k , $\beta_k \equiv 1$. La cadena inhomogénea (P_k) de parámetros $((1, \lambda_k))$ converge hacia π_C siempre que la sucesión (λ_k) verifique:

$$\sum_{k>0} \exp(-n\Gamma\lambda_k) = +\infty$$

Esta condición se satisface para $\lambda_k = \lambda \log k$ siempre que $\lambda n\Gamma \leq 1$.

(2) *Optimización:* La cadena inhomogénea (P_k) de parámetros $((\beta_k, \lambda_k))$ converge a $\pi_{C, \infty}$ si las sucesiones (β_k) y (λ_k) verifican:

$$\sum_{k>0} \exp\{-n\beta_k(\Delta + \lambda_k\Gamma)\} = +\infty$$

Esta condición se satisface para $\beta_k \lambda_k = \gamma \log k$ siempre que $\gamma n\Gamma < 1$.

Demostración: Para establecer cada uno de los resultados, basta comprobar (3.6) y (3.5).

(1-(i)) *Simulación y (3.6)* : basta verificar que $(\pi_k(x))_k$ es monótona. El método es estándar: escribimos

$$\pi_k(x)^{-1} = \sum_{y \in E} \exp\{(U(x) - U(y)) + \lambda_k(C(x) - C(y))\}$$

Dividimos la suma en tres según $C(y)$ sea $<$, $=$ o $>$ que $C(x)$ y se concluye fácilmente.

(1-(ii)) *Simulación y (3.5)* : la distribución condicional en i vale

$$\pi_{i,k}(x_i | x^i) = \left\{ \sum_{a \in F} \exp[(U(x_i, x^i) - U(a, x^i)) + \lambda_k(C(x_i, x^i) - C(a, x^i))] \right\}^{-1}$$

Estas probabilidades están acotadas inferiormente por $L^{-1} \exp\{-\Delta + \lambda_k\Gamma\}$, donde $L = |F|$, y en consecuencia:

$$1 - c(P_k) \geq \exp\{-n\{\Delta + \lambda_k\Gamma\}\}$$

de donde se obtiene el resultado (1).

(2-(i)) *Optimización y (3.6)* : Vamos a utilizar el lema siguiente. Sea $(\pi_k = Z_k^{-1} \exp\{-U_k\})$ una sucesión de distribuciones sobre E .

Lema 3.1 ([39]) *Supongamos que la sucesin (U_k) es tal que se cumplen:*

(a) $\forall x \in E, \exists k_0$ tal que $U_k(x) \leq U_{k+1}(x)$ si $k \geq k_0$.

(b) $\exists x_0 \in E$ tal que $\sup_{k \geq 0} U_k(x_0) < +\infty$.

Entonces, bajo (a) y (b), la sucesin (π_k) verifica (3.6)

La demostración de este lema está a continuación de la del teorema. Verifiquemos que las condiciones (a) y (b) se satisfacen. Aquí, $U_k = \beta_k(U(x) + \lambda_k C(x))$, (β_k) y $(\lambda_k) \uparrow +\infty$.

Condición (a): si $x \in E_C$, $C(x) = 0$ y $U(x) \geq U_{*,C} = 0 : (U_k(x))$ es creciente. Si $x \notin E_C$, y si $U(x) \geq 0$, tenemos el mismo resultado. Falta examinar el caso en el cual $x \notin E_C$ y $U(x) < 0$. Tenemos:

$$U_{k+1}(x) - U_k(x) = (\beta_{k+1} - \beta_k)[U(x) + \lambda_{k+1}C(x)] + \beta_k(\lambda_{k+1} - \lambda_k)C(x)$$

Este crecimiento es positivo a partir de un cierto valor porque $C(x) > 0$ y $\lambda_k \uparrow +\infty$.

Condición (b): basta tomar $x_0 \in E_{*,C}$ porque entonces, $U(x_0) = C(x_0) = 0$, y en consecuencia, para todo $k \geq 0$, $U_k(x_0) = \beta_k\{U(x_0) + \lambda_k C(x_0)\} = 0$.

(2-(ii)) *Optimización y (3.5):* introduciendo el parámetro β_k , obtenemos

$$\pi_{i,k}(x_i | x^i) = \left\{ \sum_{a \in F} \exp \beta_k [(U(x_i, x^i) - U(a, x^i)) + \lambda_k (C(x_i, x^i) - C(a, x^i))] \right\}^{-1}$$

Estas probabilidades están acotadas inferiormente por $L^{-1} \exp -\beta_k \{\Delta + \lambda_k \Gamma\}$, y obtenemos la acotación inferior:

$$1 - c(P_k) \geq \exp -n\beta_k \{\Delta + \lambda_k \Gamma\}$$

Deducimos el resultado (2). Para el esquema logarítmico, la desigualdad $\gamma n \Gamma < 1$ estricta es necesaria: en efecto $\exp -n\beta_k \lambda_k \{\Gamma + \lambda_k^{-1} \Delta\} = (\frac{1}{k})^{n\gamma \{\Gamma + \lambda_k^{-1} \Delta\}} \geq \frac{1}{k}$ siempre que k sea suficientemente grande. \square

Demostración del Lema:

Basta demostrar que para cada x , $\sum_{k \geq 0} |\pi_{k+1}(x) - \pi_k(x)| < \infty$. Fijemos un x y llamemos $u_k = \exp -U_k(x)$. Tenemos la acotación

$$\begin{aligned} |\pi_{k+1}(x) - \pi_k(x)| &= (Z_k Z_{k+1})^{-1} |u_{k+1} Z_k - u_k Z_{k+1}| \\ &\leq (Z_k Z_{k+1})^{-1} \{u_{k+1} |Z_{k+1} - Z_k| + Z_{k+1} |u_{k+1} - u_k|\} \\ &\leq (\inf_k Z_k)^{-2} \{[\sup_k u_k] |Z_{k+1} - Z_k| + [\sup_k Z_k] |u_{k+1} - u_k|\} \end{aligned}$$

Ahora bien, $\sup_k u_k < \infty$ ya que $(U_k(x))$ es creciente a partir de un cierto valor. Igualmente, como E es finito, $\sup_k Z_k < \infty$. Además, $\inf_k Z_k \geq \exp\{-\sup_k U_k(x_0)\} > 0$. Concluimos observando que las sucesiones (Z_k) y (u_k) son positivas y decrecientes a partir de un cierto rango. \square

En [39], D. Geman da una forma más general de este resultado. En primer lugar, los relajamientos estocásticos pueden hacerse sobre subconjuntos A_i (hasta ahora hemos considerado conjuntos de un sólo elemento $A_i = \{i\}$). Los subconjuntos asociados a cada barrido deben recubrir S , pero los barridos no son necesariamente periódicos. En fin, si el barrido número k es de longitud k y tiene (β_k, λ_k) fijos, hace falta una condición mixta sobre (β_k, λ_k) y sobre τ_k para asegurar la convergencia de la cadena. Si los τ_k son acotados, los esquemas logarítmicos $\gamma \log k$ aseguran las convergencias deseadas si γ es suficientemente pequeña.

3.3.2. Simulación y optimización bajo restricción para la dinámica de Metropolis

Sea Q un núcleo de proposición simétrica, irreducible y regular (esta hipótesis puede relajarse, cf. [107]). Sea M el menor entero tal que $Q^M > 0$. La transición de Metropolis asociada a $U_k = \beta_k(U(x) + \lambda_k C(x))$ es, para $x \neq y$:

$$P_k(x) = Q(x, y) \exp -[U_k(y) - U_k(x)]^+$$

Sean (β_k) y (λ_k) dos sucesiones que crecen hacia $+\infty$. Tenemos entonces:

Proposición 3.16 *Simulación y optimización bajo restricción (dinámica de Metropolis, Yao [107]).*

(1) Para $\beta_k \equiv 1$, la cadena (P_k) realiza la simulación de π_C siempre que

$$\sum_{k \geq 0} \exp -\{M\Gamma\lambda_{kM}\} = +\infty$$

(2) La cadena (P_k) realiza la optimización de U sobre $E(C)$ siempre que

$$\sum_{k \geq 0} \exp -[M\beta_{kM}\{\Delta + \lambda_{kM}\Gamma\}] = +\infty$$

Un esquema natural consiste en dejar constante, para niveles de longitud l , los parámetros β y λ : si $l = M$ y si β_k y λ_k son los valores sobre el k -ésimo nivel, las condiciones de ergodicidad se obtienen reemplazando en los criterios precedentes el índice kM por k .

3.4. Ejercicios

Ejercicio 3.1 *Contracción sobre un espacio a 2 estados.*

Sean las transiciones $P = \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix}$ y $Q = \begin{pmatrix} 1-a' & a' \\ b' & 1-b' \end{pmatrix}$.

- (1) Mostrar que $c(P) = |1 - (a + b)|$.
- (2) Verificar que $c(PQ) = c(P)c(Q)$.
- (3) Para dos distribuciones μ y ν , verificar que $\|\mu P - \nu P\| = \|\mu - \nu\| c(P)$

Ejercicio 3.2 *Algunos cálculos de coeficientes de contracción.*

(1) Sea la transición $P = \begin{pmatrix} 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{4} & 0 & \frac{1}{4} \end{pmatrix}$.

Calcular $c(P)$, $c(P^2)$. Comparar $c(P^2)$, $c(P)^2$ y $(1 - 4 \inf_{x,y} P^2(x, y))$.

(2) Sea la distribución $\pi(x_1, x_2) = Z^{-1} \exp \beta x_1 x_2$ sobre $\{-1, +1\}^2$. Consideramos el muestreador de Gibbs de transición $P((x_1, x_2), (y_1, y_2)) = \pi_1(y_1 | x_2) \pi_2(y_2 | y_1)$. Calcular $c(P)$ y comparar con la acotación $1 - 4 \inf_{x,y} P(x, y)$.

(3) Sea la distribución π sobre $E = \{1, 2, \dots, r\}$ definida por una familia (a_i) , $\pi_i = Z^{-1} \exp a_i$. Describir la dinámica de Metropolis para la proposición de cambio: $\forall i, j \in E$, $q(i, j) = \frac{1}{r}$. Dar una acotación de $c(P)$ a partir de la oscilación $\delta(a)$ de a .

Ejercicio 3.3 *Estudio de la correlación de una cadena binaria.*

Sobre $E = \{-1, +1\}$, consideramos la cadena estacionaria $X = (X_0, X_1, \dots)$ de transición P , $P(x, x) = a, x = \pm 1, 0 < a < 1$, y de distribución inicial $\pi = (\frac{1}{2}, \frac{1}{2})$. Sea $f : E \rightarrow \mathbb{R}$, $f(+1) = -f(-1) = \alpha$.

- (1) Diagonalizar P y evaluar $c(P^n)$. Calcular $c_n(a) = \text{cov}(f(X_0), f(X_{n-1}))$.
- (2) Verificar el resultado:

$$\lim_{n \rightarrow \infty} [\text{Var} \frac{1}{\sqrt{n}} \sum_{i=1, n} f(X_i)] = v(a) = v(f, \pi, P) = \frac{1 + \lambda}{1 - \lambda} \langle f, e \rangle_{\pi}^2$$

donde λ es el valor propio de P diferente de 1, y e el vector propio asociado de norma 1 en $l^2(\pi)$. Estudiar la función $a \mapsto v(a)$.

Ejercicio 3.4 *Sobre la acotación $c(PQ) \leq c(P)c(Q)$.*

Sea la cadena de transiciones $P_{2n-1} = P, P_{2n} = Q, n \geq 1$, donde

$$P = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \quad \text{y} \quad Q = \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{pmatrix}$$

- (1) Calcular $c(P), c(Q)$ y $c(PQ)$.
- (2) Calcular $c(P_1 P_2 \dots P_{2n})$ y comparar a $c(P_1)c(P_2) \dots c(P_{2n})$.

Ejercicio 3.5 *Sobre la importancia de la condición: $\forall m$ en (3.5).*

Sean las transiciones: $P_1 = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}, P_2 = \begin{pmatrix} \frac{1}{4} & \frac{3}{4} \\ \frac{1}{4} & \frac{3}{4} \end{pmatrix}, P_n = \begin{pmatrix} \frac{1}{n^2} & 1 - \frac{1}{n^2} \\ 1 - \frac{1}{n^2} & \frac{1}{n^2} \end{pmatrix}$ si $n \geq 3$.

- (1) Calcular $P_1 P_2$ y $P_1 P_2 P_3$. Mostrar que $Q_n = P_1 P_2 \dots P_n = \begin{pmatrix} a_n & 1 - a_n \\ a_n & 1 - a_n \end{pmatrix}$.
- (2) Deducir en consecuencia que para $n \geq 2, X_n$ es independiente de X_1 . ¿Es este resultado cierto entre X_m y $X_{m+n}, m \geq 2, n \geq 1$?

Ejercicio 3.6 *Cadena débilmente ergódica y no fuertemente ergódica.*

Mostrar que la cadena siguiente es débilmente, pero no fuertemente ergódica:

$$P_{2n-1} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad P_{2n} = \begin{pmatrix} 0 & 1 \\ 1 - \frac{1}{2n} & \frac{1}{2n} \end{pmatrix}, \quad n \geq 1$$

Ejercicio 3.7 *Cadena fuertemente ergódica.*

Sea, para $n \geq 2, P_n = \begin{pmatrix} \frac{1}{3} + \frac{1}{n} & \frac{2}{3} - \frac{1}{n} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$. Demostrar que (P_n) es fuertemente ergódica.

Ejercicio 3.8 *Paseo aleatorio con barrera reflectora*

Sean a, c y $b = (1 - (a + c))$ tres reales > 0 y la transición P sobre $E = \{1, 2, \dots, n\}$ caracterizada por: $P_{i,i+1} = c$ para $1 \leq i \leq n - 1$, $P_{i,i-1} = a$ para $2 \leq i \leq n$, $P_{1,1} = a + b$ y $P_{n,n} = b + c$.

(1) Identificar la menor potencia r tal que $P^r > 0$. Si $b \geq c \geq a > 0$, verificar que, para este r , $\inf_{i,j} P^r(i, j) = P^r(n, 1)$. Deducir de allí una acotación de $c(P^r)$.

En todo lo que sigue, tomamos $n = 3$ y $b \geq c \geq a > 0$ ($b_k \geq c_k \geq a_k > 0$ si estos parámetros están indexados en k); llamamos $P_{a,c}$ a la transición de parámetros a y c .

(2) Determinar la distribución invariante de $P_{a,c}$ y calcular $c(P_{a,c})$.

(3) Sea (P_k) la cadena inhomogénea de parámetros $(a_k, c_k)_{k \geq 1}$. Para $a_k = c_k \rightarrow 0$, dar una condición suficiente de ergodicidad fuerte de (P_k) . Para $a_k = \frac{1}{k}$ y $c_k = \frac{1}{\sqrt{k}}$, $k \geq 3$, (P_k) ¿es ella fuertemente ergódica?

Ejercicio 3.9 *El problema del agente viajero.*

Un agente viajero debe visitar $(N + 1)$ ciudades V_0, V_1, \dots, V_N : parte de V_0 , y regresa a ella al final de su ciclo, y pasa una vez, y sólo una, por todas las otras ciudades. El conjunto E de tales ciclos $x = (x_0, x_1, x_2, \dots, x_N, x_0)$ ($x_0 \equiv V_0$) se identifica con el conjunto de las permutaciones de $\{1, 2, \dots, N\}$: si σ es una permutación, el ciclo es

$$0 \mapsto \sigma(1) \mapsto \sigma(2) \mapsto \dots \mapsto \sigma(N) \mapsto 0$$

Llamemos $d(i, j)$ la distancia entre las ciudades V_i y V_j . El agente viajero quiere organizar lo mejor posible su ciclo, es decir, minimizar la longitud $U(x)$ de su ciclo x :

$$U(x) = \sum_{i=0}^N d(x_i, x_{i+1})$$

Consideramos las 4 núcleos de proposiciones Q siguientes:

(1) Q_1 , *intercambio de dos ciudades*: escogemos al azar dos ciudades x_i y x_j diferentes (distintas de V_0) e intercambiamos estas dos ciudades en el recorrido.

(2) Q_2 , *intercambio de dos ciudades consecutivas*: lo mismo pero sobre x_i, x_{i+1} , $1 \leq i \leq N - 1$.

(3) Q_3 , *inversión de un segmento*: escogemos al azar un par (i, j) con $1 \leq i < j \leq N$. Este par define un segmento $[x_i \mapsto x_{i+1} \mapsto \dots \mapsto x_j]$ que invertimos $[x_j \mapsto x_{j+1} \mapsto \dots \mapsto x_i]$. Así $x_{i-1} \mapsto x_i$ (resp. $x_j \mapsto x_{j+1}$) se transforma en $x_{i-1} \mapsto x_j$ (resp. $x_i \mapsto x_{j+1}$), y todo el resto permanece inalterado.

(4) Q_4 , *inserción de una ciudad*: escoger i al azar en $\{1, 2, \dots, N\}$ y j al azar en $\{0, 1, 2, \dots, N\}$. Desplazar x_i e insertarlo entre j y $j + 1$.

Ilustrar gráficamente estas cuatro proposiciones de cambio. Estudiar la simetría y la irreducibilidad de estas transiciones. Describir la dinámica de Metropolis. Como ejemplo, simular la distribución $\pi_T = Z^{-1} \exp -\frac{U}{T}$ para $T = 1, 0,1$ y $0,01$. Representar los recorridos obtenidos.

Ejercicio 3.10 *El problema del agente viajero con restricción.*

Estamos en el contexto del ejercicio precedente. Las $(N + 1)$ ciudades pertenecen a p ($p \geq 2$) países B_0, B_1, \dots, B_{p-1} , y $V_0 \in B_0$, el país de origen del viajero. La restricción es la siguiente: una vez que uno entra en un nuevo país B_j ($j \neq 0$), debe visitar todas las ciudades de ese país antes de salir (derecho de entrada importante).

(1) Proponer restricción(es) $C : E \rightarrow \mathbb{R}^+$ que reflejen estas reglas.

(2) Programar la optimización bajo restricción sobre un ejemplo simple.

Ejercicio 3.11 *Matriz con longitud de banda mnima.*

$A_0 = (A_0(i, j))_{i,j=1,n}$ es una matriz para la cual la mayoría de los coeficientes son nulos (matriz rala). Buscamos reagrupar los términos no nulos alrededor de la diagonal por permutaciones sucesivas de filas y columnas. Sea E la clase de todas las matrices que pueden obtenerse de esta manera a partir de A_0 . El problema es minimizar la función U sobre E :

$$U(A) = \sum_{i,j=1}^n d(i, j) \mathbf{1}\{A(i, j) \neq 0\}$$

para d una pseudo-distancia sobre $\{1, 2, \dots, n\}$. Sea Q la transición de proposición de cambio siguiente: permutamos dos filas, luego dos columnas, las filas permutadas así como las columnas permutadas se escogen al azar.

- (1) Mostrar que Q es simétrica e irreducible.
- (2) Describir el algoritmo de Metropolis a temperatura constante precisando el cálculo de ΔU .
- (3) Tomar $n = 100$, $(A_0(i, j))_{i,j=1,100}$ i.i.d. Bernoulli de parámetro 0.05, $d(i, j) = |i - j|^p$, $p = 1, 2, 4$. Efectuar el RS para la minimización de U .
- (4) ¿Es regular Q ? Que pensar del algoritmo para la proposición de cambio Q' siguiente: con probabilidad $\{\frac{1}{2}, \frac{1}{2}\}$, intercambiamos bien dos filas, bien dos columnas, las filas permutadas (resp. las columnas permutadas) se escogen al azar.

Ejercicio 3.12 *Divisin maximal de un grafo.*

$S = \{1, 2, \dots, n\}$ es el conjunto de vértices de un grafo valuado no-orientado, la arista $\langle i, j \rangle$ tiene peso $b_{ij} > 0$. Sea $E = \mathcal{P}(S)$ el conjunto de subconjuntos de S y $U : E \rightarrow \mathbb{R}$ una función definida por:

$$\text{Para } A \subseteq S: U(A) = \sum_{\langle i,j \rangle: i \in A, j \notin A} b_{ij}$$

Observamos que $a_i = \sum_{j \in \partial i} b_{ij}$ donde $\partial i = \{j \in S, \langle i, j \rangle\}$. $\mathcal{P}(S)$, que no es un conjunto producto, puede ser puesto en biyección con $\Omega = \{0, 1\}^S$:

$$A \leftrightarrow x = (x_1, x_2, \dots, x_n) \text{ con } x_i = \begin{cases} 1 & \text{si } i \in A \\ 0 & \text{si } i \notin A \end{cases}$$

Verificamos que $U(A)$ se escribe en función de x :

$$U(x) = \sum_{i \in S} a_i x_i - 2 \sum_{\langle i,j \rangle} b_{ij} x_i x_j$$

Esta correspondencia hace de E un conjunto producto y permite utilizar la dinámica de Gibbs para la optimización de U por el RS. Para $\beta > 0$, sea $\pi_\beta(x) = Z^{-1}(\beta) \exp\{\beta U(x)\}$.

(I) *RS para la dinámica de Gibbs sobre $\Omega = \{0, 1\}^n$.*

(I-1) Describir la distribución condicional $\pi_{i,\beta}(x_i | x^i)$. Utilizando la positividad de b_{ij} , dar una minoración uniforme en x^i de $\pi_{i,\beta}(0 | x^i)$ y de $\pi_{i,\beta}(1 | x^i)$. Deducir en consecuencia que, uniformemente en x : $\pi_{i,\beta}(x_i | x^i) \geq \frac{1}{2} \exp\{-\beta a_i\}$.

(I-2) Sea P_β la transición del muestreador de Gibbs para el barrido sistemático $1 \rightarrow 2 \rightarrow 3 \dots \rightarrow n$. Dar una acotación del coeficiente de contracción $c(P_\beta)$ de P_β en función de β y de $a = \sum_{i \in S} a_i$.

(I-3) Si $(\beta_k)_{k \geq 1}$ es una sucesión que tiende a $+\infty$, y para $P_k = P_{\beta_k}$, dar una condición suficiente sobre la sucesión β que asegure $c(P_1 P_2 \cdots P_k) \rightarrow 0$ si $k \rightarrow +\infty$.

(II) *RS para la dinámica de Metropolis sobre $\mathcal{P}(S)$* . Escogemos la transición de proposición Q siguiente que hace pasar de $A \rightarrow B$:

- con una probabilidad ε , $0 < \varepsilon < 1$, nos quedamos en A ;

- con una probabilidad $1 - \varepsilon$, escogemos al azar uniformemente un lugar i de S . Se presentan dos situaciones: (i) si $i \in A$, pasamos a $B = A \setminus \{i\}$; (ii) si $i \notin A$, pasamos a $B = A \cup \{i\}$. No está permitido ningún otro cambio.

(II-1) Describir la transición Q . Verificar que Q es simétrica, irreducible y aperiódica.

(II-2) Mostrar que para todo A, B , $Q^{2n}(A, B) > 0$ y dar una cota inferior de estas probabilidades. ¿Es mínimo este orden $r = 2n$ que garantiza la positividad de Q^r ?

(II-3) Escribir la transición de Metropolis $P_\beta(A, B)$ asociada a la energía βU .

(II-4) Si i es el lugar que define el paso de A a B , verificar que $U(A) - U(B) \leq a_i$. Deducir en consecuencia una acotación de $c(P_\beta^{2n})$. Dar una condición sobre (β_k) que asegure la convergencia del RS al máximo de U .

Ejercicio 3.13 *División maximal de un grafo bajo restricción.*

Estamos en el contexto del ejercicio precedente, el número n de lugares de S es par, $n = 2p$. Proponer un método de optimización de U bajo la restricción que la división debe hacerse en dos partes iguales:

(1) Sea readaptando las transiciones de modo que uno permanezca en el espacio restringido $E(C) = \{A \subset S : |A| = p\}$.

(2) Sea definiendo una restricción $C : E \rightarrow \mathbb{R}$ y utilizando el procedimiento de simulación bajo restricción.

Ejercicio 3.14 *Coloración de un mapa.*

Un mapa es una colección de n países, $S = \{1, 2, \dots, n\}$, y un grafo de vecindad sobre $S : \langle i, j \rangle$ si i y j son dos países vecinos. Queremos colorear el mapa con K colores $F = \{c_1, c_2, \dots, c_K\}$. Una coloración está representada por $x = (x_1, x_2, \dots, x_n) \in E = F^S$, y queremos hacer una coloración que minimice:

$$U(x) = \sum_{\langle i, j \rangle} \mathbf{1}(x_i = x_j)$$

Llamamos ∂i al conjunto de vecinos de i , y $n_i(x) = \sum_{j \in \partial i} \mathbf{1}(x_j = x_i)$ el número de vecinos de i de color x_i . Llamamos $N = \frac{1}{2} \sum_{i \in S} |\partial i|$ el número de aristas del grafo de vecindad.

(1) *Optimización vía el muestreador de Gibbs*. Escogemos el barrido $1 \rightarrow 2 \rightarrow \dots \rightarrow n$, y la longitud de un barrido, fijamos la temperatura a β^{-1} .

(1-1) Describir $\pi_{i,\beta}(x_i | x^i)$ y demostrar que $\pi_{i,\beta}(x_i | x^i) \geq k^{-1} \exp -\beta |\partial i|$.

(1-2) Describir la transición P_β asociada a un barrido. Dar una acotación de $c(P_\beta)$. Dar una condición suficiente sobre (β_k) que garantice que para todo m , $\lim_{k \rightarrow \infty} c(P_m P_{m+1} \cdots P_k) = 0$. Deducir en consecuencia que mientras más grande es N , más lenta debe ser la convergencia de (β_k) hacia $+\infty$.

(2) *Optimización para la dinámica de Metropolis*. Sea Q la transición de proposición de cambio siguiente: se escoge un lugar i en S , uniformemente; en este lugar, se propone el cambio $x_i \rightarrow y_i$, $y_i \neq x_i$, y_i uniforme sobre $F \setminus \{x_i\}$; no hay cambio en ningún otro lugar.

(2-1) Describir Q . ¿Es Q irreducible? Describir el algoritmo de Metropolis a temperatura constante. ¿Es ergódico este algoritmo?

(2-2) ¿Es Q ergódica? (distinguir los casos $K \geq 3$ o $K = 2$).

(3) Experimentar ambos algoritmos sobre un ejemplo de mapa hecho por usted con $K = 3$ y función a minimizar

$$U^F(x) = \sum_{\langle i,j \rangle} F_{i,j}$$

donde $F_{i,j}$ es la longitud de la frontera común entre los países i y j . Para $K = 4$ y mapas más complejos, constatar que el teorema de los 4 colores es cierto.

Ejercicio 3.15 *Reconstrucción de un sistema de fallas.*

No situamos sobre el dominio $D = [0, 1]^2$. Una falla está identificada por una recta d que divide a D . Por otra parte, una valuación está asociada a esta falla d , $V_d : D \rightarrow \{-1, +1\}$, donde V_d es constante sobre los dos semi-planos definidos por d , $+1$ sobre uno, -1 sobre el otro. Sabemos que exactamente n fallas, $R = \{d_1, d_2, \dots, d_n\}$ intersectan D (conocemos n , pero no R). En consecuencia, para este sistema R , existe una función que mide la valuación total $V_R : D \rightarrow \mathbb{R}$, $V_R(x, y) = \sum_{i=1, n} V_{d_i}(x, y)$.

La información disponible es la siguiente. Se dispone de m pozos \mathcal{X} localizados en m lugares conocidos de D , $\mathcal{X} = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$. En cada uno de ellos, conocemos la valuación exacta $V_i, i = 1, m$ de la red real.

El problema de reconstrucción de R puede ser visto como un problema de simulación bajo restricción: simular n rectas R , que representan los límites de fallas y dividen a D bajo la restricción

$$C(R) = \sum_{i=1}^m (V_R(x_i, y_i) - V_i)^2 \text{ mnimo}$$

(1) Realizar la simulación bajo la restricción C de la distribución π siguiente: las n rectas valuadas son *i.i.d.* (Indicaciones: una recta d está parametrizada por un par $(r, \theta) \in \mathbb{R}^+ \times [0, 2\pi[$, los parámetros polares de d . La ecuación cartesiana de una recta de este tipo es

$$d(x, y) = x \cos \theta + y \sin \theta - r = 0$$

Consideraremos entonces n rectas independientes **y** que dividen a D (es necesario determinar el sub-conjunto Δ de $\mathbb{R}^+ \times [0, 2\pi[$ aceptable). Escoger al azar una recta d corresponde a escoger al azar uniformemente un punto de Δ . La selección de la valuación se puede efectuar así: sea ε una variable uniforme sobre $\{-1, +1\}$ independiente de d ; la valuación de d en un punto (x, y) para una recta d es: $V_d(x, y) = \text{sign}(\varepsilon d(x, y))$.

(2) Mismo problema, pero no conocemos a priori el número n de fallas que dividen a D .

Ejercicio 3.16 *Recocido simulado con dos estados, $E = \{0, 1\}$ ([101]).*

Consideramos la sucesión de transiciones donde $h_0 > h_1 > 0$, y $(a_n) \uparrow +\infty$:

$$P_n = \begin{pmatrix} 1 - e^{-h_0 a_n} & e^{-h_0 a_n} \\ e^{-h_1 a_n} & 1 - e^{-h_1 a_n} \end{pmatrix}$$

(1) Tomamos n grande de modo que $\alpha_n \triangleq e^{-h_0 a_n} + e^{-h_1 a_n} < 1$. Mostrar que $c(P_n) = 1 - \alpha_n$. ¿Cuál es la distribución invariante π_n de P_n ?

(2) Calcular $\sum_n \|\pi_n - \pi_{n+1}\|$. Identificar el límite π_∞ de (π_n) .

(3) Dar una condición sobre (a_n) que asegure que la cadena es fuertemente ergódica.

Comentarios Fijamos la distribución uniforme como distribución inicial y observamos que $e_k = P(X_k = 1)$, la probabilidad de que el algoritmo no de el máximo buscado. Fijemos n (el número de pasos del R.S.) y consideremos el problema

$$\begin{cases} \text{Minimizar } e_n = P(X_n = 1) \\ \text{entre los esquemas: } T_1 \geq T_2 \geq \dots \geq T_n \end{cases}$$

Tenemos el resultado asintótico siguiente ([101], observemos que ahora el esquema no es triangular): la solución está controlada por dos constantes positivas R_1 y R_2 , con $R_1 + \frac{\log n}{h_1} \leq \frac{1}{T_n^{\text{opt}}} \leq R_2 + \frac{\log n}{h_1}$. Comentar este resultado en relación con (3). Por otro lado, para este esquema, $e_n = O(n^{-\alpha})$ con $\alpha = \frac{h_0 - h_1}{h_1}$.

Ejercicio 3.17 *Grafos, rboles, distribuciones invariantes y grandes desviaciones (Ver, [101]; [29])*

(1) Consideramos una transición irreducible P sobre E finito. Asociamos a P su grafo (orientado) $\mathcal{G}: x \mapsto y$ si $P(x, y) > 0$. Un sub-grafo g de \mathcal{G} es un sub-conjunto de flechas. Cada sub-grafo g está valuado por un peso $\pi(g) = \prod_{x \mapsto y \in g} P(x, y)$.

Un árbol g que conduce a un punto $x \in E$ es un sub-grafo de \mathcal{G} que reagrupa todos los puntos de E y para el cual: (i) cada $y \in E \setminus \{x\}$ es el origen de una única flecha en g ; (ii) x no es el origen de ninguna flecha; (iii) no hay ciclos en g . Llamamos $\mathcal{G}(x)$ al conjunto de árboles que conducen a x . Demostrar la fórmula siguiente (lema de Bott-Mayberry): la única distribución invariante de P está dada por $\pi(x) = c \sum_{g \in \mathcal{G}(x)} \pi(g)$.

(2) Nos damos una transición irreducible q sobre E y consideramos el núcleo de Metropolis generalizado siguiente:

$$P_T(x, y) = q(x, y) \exp\left[-\frac{1}{T}V(x, y)\right] \quad \text{para } x \neq y, T > 0$$

donde V es un costo de comunicación sobre $E \setminus \{(x, x), x \in E\}$ que vale $+\infty$ si y solamente si $q(x, y) = 0$. P_T y q tienen el mismo grafo de comunicación \mathcal{G} . Definimos la energía virtual W por $W(x) = \min\{V(g), g \in \mathcal{G}(x)\}$ donde $V(g) = \sum_{x \mapsto y \in g} V(x, y)$.

Demostrar que si $T \rightarrow 0$, las distribuciones invariantes π_T de P_T satisfacen el principio de grandes desviaciones

$$\lim_{T \rightarrow 0} T \log \pi_T(x) = -[W(x) - W_*] \quad \text{para } x \in E \setminus E_*$$

donde $W_* = \min_{y \in E} W(y)$, $E_* = \{y \in E : W(y) = W_*\}$.

Capítulo 4

Diagnostico de convergencia de una cadena

En la simulación de una distribución π utilizando la dinámica de una cadena de Markov de transición P y de distribución inicial ν , las dos cuestiones centrales son:

- (i) ¿Cuándo detener las iteraciones (*burn-in time*) para garantizar que νP^n y π estén cerca?
- (ii) ¿Cuál es el sesgo de la simulación $\|\nu P^n - \pi\|$?

Existen diversos enfoques para responder a estas preguntas.

(A) El primero consiste en *acotar la velocidad* $\|\nu P^n - \pi\|$. En el caso de un espacio de estados finito, se trata de evaluar el *hueco espectral* (*spectral gap*) $s = 1 - |\lambda|$ donde λ es el valor propio de P de mayor módulo y $\neq 1$, ya que $\|\nu P^n - \pi\| \leq C(\nu) |\lambda|^n$. Las limitaciones de este control son numerosas:

- La obtención explícita de cotas apropiadas no se conoce, ni para modelos relativamente simples y a fortiori para modelos más complejos tales en la práctica. Esto se debe a la gran dimensión del espacio de estados E , es decir de la matriz de transición P .
- Los resultados teóricos existentes son frecuentemente inaplicables. Por ejemplo, si para el modelo de Ising sobre $E = \{-1, 1\}^r$, n puede escogerse polinomial en r (cf. [23], [84] y [93];⁽¹⁾), las cotas obtenidas sobre los coeficientes polinomiales son tales que el resultado es inaplicable.
- Estas acotaciones superiores pueden ser muy pesimistas, conduciendo a valores de n excesivamente grandes. Esta crítica se aplica a fortiori a las acotaciones más simples y groseras, como aquellas que se basan en el coeficiente de contracción.

(B) Un segundo enfoque consiste en proponer *diagnósticos heurísticos* basados en un control empírico de la convergencia. El lector puede consultar sobre este tema las dos recopilaciones de artículos, Gilks W.R., Richardson S. y Spiegelhalter D.J. [43] (cf. [86]), y C. Robert [90] (²) (cf. [89] y [17]). Si bien estos métodos son razonables, ellos no resuelven todos los problemas, como la metaestabilidad de ciertos regímenes en los cuales la cadena puede permanecer largo tiempo, haciendo pensar que ella entró en su régimen estacionario.

(C) Un tercer enfoque, reciente, es el de la *simulación exacta* por acoplamiento desde el pasado (*Coupling From The Past, CFTP*). En su artículo muy innovador, J.G. Propp y D.B.

¹Una dirección sobre cadenas de Markov en general es: <http://www.stat.berkeley.edu/~aldous/book.html>

²La dirección electrónica sobre el diagnóstico de métodos MCMC es: <http://www.ensae.fr/crest/statistique/robert/McDiag/>

Wilson (cf. [84], 1996) proponen un algoritmo que resuelve simultáneamente los problemas (i) y (ii), es decir que proponen un algoritmo:

- (i') que detecta automáticamente cuando debe detenerse, y
- (ii') que produce una muestra sin sesgo de π .

La idea de base es la siguiente: simulamos, regresando hacia el pasado, tantas realizaciones de las transiciones como estados haya (es decir $r = |E|$ transiciones a cada instante, tantas veces como sea necesario); si existe un instante de tiempo $-T < 0$ tal que las r cadenas que se inician en $-T$ se acoplan todas en el instante 0 (0 es el instante de *coalescencia* de las cadenas), entonces el estado común $x(0)$ en el instante 0 sigue *exactamente* la distribución π . La duración del algoritmo es aleatoria, pero casi seguramente finita: el algoritmo produce $X(0) \sim \pi$.

Tal cual, el método es inaplicable porque: (1) necesita un número muy grande de operaciones (si r , el cardinal del espacio de estados E , es muy grande, hacen falta $r \times T$ operaciones si el acoplamiento es en $-T$); (2) necesita demasiada memoria (r lugares de memoria para regresar una unidad en el tiempo). Sin embargo, en casos favorables, la puesta en práctica se beneficia de simplificaciones importantes que hacen el algoritmo realizable. Tal es el caso estudiado por Propp y Wilson, donde E tiene una relación de orden parcial que es preservado por la cadena de Markov (cadena de Markov montona). Colocándose a la temperatura crítica (una frontera buena entre la dependencia espacial débil y fuerte), Propp y Wilson obtienen la simulación exacta de un modelo de Ising atractivo sobre una red 512×512 , con un tiempo de acoplamiento del orden de $T = 30$ barridos, y un tiempo de realización de 10 min. sobre una estación Sparc.

Por su originalidad y su eficiencia, el artículo de Propp y Wilson ha redinamizado la investigación sobre la simulación por cadenas de Markov³. Uno de los objetivos es la extensión de CFTP a contextos distintos al estudiado por Propp y Wilson (E infinito, E sin elemento maximal o sin orden, dinámica no-montona, modelo no-atractivo), como los procesos puntuales (W. Kendall [61], O. Häggström, M.N.M. Van Lieshout y J. Møller [48], W. Kendall y J. Møller [62]), las variables multidimensionales sobre \mathbb{R}^d (D.J. Murdoch y P.J. Green, [80], J. Møller [79]), los campos de Markov no-atractivos (O. Häggström y K. Nelander, [49]). Por otra parte, la selección de un “buen acoplamiento” que reduzca tanto como se pueda $E(T_*)$ es crucial [56]. Para encarar este problema, J.A. Fill [31] propone otra estrategia, su nuevo algoritmo puede interrumpirse antes controlando el sesgo de simulación.

Nos limitaremos aquí a la presentación de los principales resultados de Propp y Wilson.

Sealemos que el enfoque de V.E. Johnson [55] que utiliza el acoplamiento hacia el futuro de varias cadenas se encuentra en la interface de los enfoques (B) y (C) (cf. también [32]).

(D) Un último enfoque, debido a B. Ycart [110], está asociado al estudio del fenómeno de *convergencia abrupta* (o *cutoff*) para una cadena. Si este fenómeno está presente, el estudio de un *tiempo de fusión* (*merging time*) ligado al tiempo de cutoff permite proponer un test de parada que asegura que la cadena entra en su régimen estacionario ([108],[72],[109] y (4)). El método está particularmente bien adaptado al caso de un espacio de estados parcialmente ordenado. Este enfoque permite además el estudio de este tipo de convergencia abrupta de νP^t hacia π .

Comenzaremos por describir este fenómeno.

³La dirección electrónica sobre la simulación exacta es: <http://dimacs.rutgers.edu/~dbwilson/exact.html/#surveys>

⁴Disponibles en la dirección <http://www.math-info.paris5.fr/~ycart/>

4.1. Fenmeno de Cutoff y test de parada de una cadena

4.1.1. Convergencia abrupta para una cadena

El fenmeno de *convergencia abrupta* de una cadena de Markov, o *cutoff*, ha sido observado por Diaconis y Aldous. El comportamiento de la convergencia es el siguiente:

Fenmeno de Cutoff: existe t_0 (el *tiempo de cutoff*) tal que $\forall \varepsilon > 0$;

$$\left\{ \begin{array}{l} \bullet \text{ Si } t < (1 - \varepsilon)t_0, \left\| \nu P^t - \pi \right\| \simeq 1, \\ \bullet \text{ Si } t > (1 + \varepsilon)t_0, \left\| \nu P^t - \pi \right\| \simeq o(1). \end{array} \right.$$

La convergencia es abrupta alrededor de t_0 , es peor antes de t_0 , y mejor despues, sin necesidad de buscarla ms all de $(1 + \varepsilon)t_0$.

De hecho, como lo precisaremos en los dos ejemplos que presentamos, este *fenmeno es asinttico* sobre un parmetro del modelo ligado al nmero de estados de E . Las referencias generales sobre este fenmeno son Diaconis et altri ([24], [23] y [22]) y Saloff-Coste [91], §2.4.2.

Ejemplo 1: El paseo al azar sobre $\{0, 1\}^n$

Consideremos el paseo al azar sobre $S = \{1, 2, \dots, n\}$, con estados $E = \{0, 1\}^n$. Aqu, la asinttica es en n , la dimensin del hipercubo. A cada paso, escogemos un lugar i de S al azar y cambiamos x_i por $1 - x_i$ en este lugar. La transicin es $P(x, y) = n^{-1}$ si $\|x - y\|_1 = 1$, 0 si no. Esta cadena es simtrica, irreducible (pero peridica), y admite la distribucin uniforme π como distribucin invariante. Tenemos, para todo estado inicial x y para todo $\varepsilon > 0$ ([21],[22]) :

$$\begin{aligned} \lim_{n \rightarrow \infty} \left\| P_x^t - \pi \right\| &= 1 \text{ si } t \leq \frac{1}{4}(1 - \varepsilon)n \log n \\ \lim_{n \rightarrow \infty} \left\| P_x^t - \pi \right\| &= 0 \text{ si } t \geq \frac{1}{4}(1 + \varepsilon)n \log n \end{aligned}$$

Como $t \rightarrow \left\| P_x^t - \pi \right\|$ es decreciente, esto nos da una descripcin del modo de convergencia cuando n es grande: la cadena se mantiene distante (en el mximo) de su distribucin estacionaria π antes de $(1 - \varepsilon)t_0(n)$ ($t_0(n) = \frac{1}{4}n \log n$); luego ella converge abruptamente a π desde $t = (1 + \varepsilon)t_0(n)$, sin necesidad de continuar ms las simulaciones. Una descripcin ms precisa de este modo de convergencia cerca de $t_0(n)$ se encuentra en [22].

Ejemplo 2 : n -muestra de una cadena de Markov ([72], [110])

Este segundo ejemplo examina una cadena para la simulacin de una n -muestra de una distribucin π sobre $E = \{1, 2, \dots, r\}$: aqu r est fijo y la asinttica es en n , el tamao de la muestra. Sea P una transicin sobre E , reversible, ergdica, de distribucin invariante π . Queremos simular $\tilde{\pi} = \pi^{\otimes n}$ sobre $\tilde{E} = E^n$. Para esto, consideramos la cadena formada por n cadenas paralelas e independientes de transicin P , la transicin sobre \tilde{E} es,

$$\tilde{P}((i_1, i_2, \dots, i_n), (j_1, j_2, \dots, j_n)) = \prod_{k=1, n} P(i_k, j_k)$$

Descomposicin espectral de P . Como P es reversible, sus autovalores son reales,

$$1 = \alpha_1 > \alpha_2 \geq \dots \geq \alpha_r > -1$$

La aperiodicidad implica $\alpha_r > -1$. Sea $\alpha = \max\{|\alpha_2|, |\alpha_r|\}$: α controla la convergencia de P^t a π . Sea $D = \text{Diag}(\sqrt{\pi_i})$. DPD^{-1} es simétrica, de valores propios $\{\alpha_l, l = 1, r\}$, de base propia ortonormal asociada $\{V_l, l = 1, r\}$. Seleccionaremos como primer autovector (asociado al valor propio 1) $V_1(i) = \sqrt{\pi(i)}$, $i = 1, r$. Llamemos finalmente $w(i) = \sum_{l:|\alpha_l|=\alpha} V_l^2(i)$, $\tilde{i} = (i, i, \dots, i) \in \tilde{E}$, \tilde{P}_i^t la distribución en t de la cadena inicializada en \tilde{i} .

Tenemos (Proposiciones 2.1 y 2.2 de [110]),

Proposición 4.1 *El tiempo de cutoff para \tilde{P} es $t_0(n) = \frac{\log n}{2 \log(1/\alpha)}$ y para $c > 0$:*

(i) Si $t > \frac{c + \log n}{2 \log(1/\alpha)}$:

$$\left\| \tilde{P}_i^t - \tilde{\pi} \right\| < \frac{1}{2} \left(\exp\left(\frac{e^{-c}}{\pi(i)}\right) - 1 \right)^{\frac{1}{2}}$$

(ii) Sea i tal que $w(i) > 0$. Entonces, existe $n_0(c)$ t.q. si $n > n_0(c)$ y si $t < \frac{\log n - c}{2 \log(1/\alpha)}$, tenemos:

$$\left\| \tilde{P}_i^t - \tilde{\pi} \right\| > 1 - 4 \exp\left\{ \frac{-e^c w^2(i)}{8\pi_i(1 - \pi_i)} \right\}$$

Demostracin:

(i) Utilizando la desigualdad del producto escalar, verificamos fcilmente que la distancia en variacin est acotada por la distancia chi-cuadrado

$$\|P_i^t - \pi\| \leq \frac{1}{2} \{\chi(P_i^t, \pi)\}^{\frac{1}{2}}, \text{ donde } \chi(P_i^t, \pi) = \sum_{l \in E} \frac{\{P_i^t(l) - \pi(l)\}^2}{\pi(l)}$$

Bastar mostrar la desigualdad para la distancia chi-cuadrado, que vale:

$$\chi(P_i^t, \pi) = -1 + \frac{1}{\pi(i)} \sum_{l=1, r} V_l^2(i) \alpha_l^{2t} \text{ para } t \geq 0$$

Falta adaptar esta frmula a \tilde{P} y $\tilde{\pi}$. Si \tilde{D} es la matriz diagonal de $\mathbf{i} = (i_1, i_2, \dots, i_n)$ -esimo trmino $\sqrt{\pi(i_1)\pi(i_2) \cdots \pi(i_n)}$, $\tilde{D}\tilde{P}\tilde{D}^{-1}$ es el producto de Kronecker de n copias de DPD^{-1} . El espectro de $\tilde{D}\tilde{P}\tilde{D}^{-1}$ se obtiene as: a toda aplicacin $g: \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, r\}$, asociamos

$$\mathbf{V}_g(\mathbf{i}) = \prod_{l=1, n} V_{g(l)}(i_l)$$

que es el autovector asociado al autovalor (llamando $n_l = \text{Card}(g^{-1}(\{l\}))$)

$$\alpha_g = \alpha_1^{n_1} \cdots \alpha_r^{n_r}$$

Obtenemos as los r^n autovalores/autovectores de $\tilde{D}\tilde{P}\tilde{D}^{-1}$. Aislado aquel que es igual a 1, obtenemos:

$$\chi(\tilde{P}_i^t, \tilde{\pi}) = -1 + \left(\frac{1}{\pi(i)}\right) \sum_{l=1, r} V_l^2(i) \alpha_l^{2t}$$

Por la seleccin de V_1 , $\frac{1}{\pi(i)} V_1^2(i) \alpha_1^{2t} = 1$. Como $(1 + x) \leq e^x$, obtenemos:

$$\chi(\tilde{P}_i^t, \tilde{\pi}) \leq -1 + \exp\left\{ n \sum_{l=2, r} \frac{V_l^2(i)}{\pi(i)} \alpha_l^{2t} \right\} \leq -1 + \exp\left\{ \frac{n}{\pi(i)} \alpha^{2t} \right\}$$

As, $\chi(\tilde{P}_i^t, \tilde{\pi}) \leq \eta$ siempre que $\frac{n}{\pi(i)}\alpha^{2t} < \log(1 + \eta)$, es decir, siempre que

$$t > \frac{\log n}{2 \log(1/\alpha)} - \frac{\log \pi(i)}{2 \log(1/\alpha)} - \frac{\log(\log(1 + \eta))}{2 \log(1/\alpha)}$$

Es el resultado anunciado para $c = -\log \pi(i) - \log(\log(1 + \eta))$.

(ii) Supondremos c lo suficientemente grande de tal manera que una cota inferior de la desigualdad sea ≥ 0 . Para demostrar la desigualdad, es suficiente encontrar un subconjunto $\mathbf{F}_t \subset E^n$, tal que para $t < (\log n - c)/(2 \log(1/\alpha))$ y $\varepsilon(c, i) = \frac{-e^c w^2(i)}{8\pi_i(1-\pi_i)}$, se tiene:

$$\tilde{\pi}(\mathbf{F}_t) > 1 - 2 \exp\{\varepsilon(c, i)\} \text{ y } \tilde{P}_i^t(\mathbf{F}_t) < 2 \exp\{\varepsilon(c, i)\}$$

Sea $\mathbf{i} = (i_1, i_2, \dots, i_n) \in \tilde{E}$ el valor muestreado por \tilde{P}_i^t al instante t , y, para $i \in E$, $N_i(\mathbf{i})$ el nmero de coordenadas de \mathbf{i} iguales a i . Entonces N_i es una distribucion binomial $\mathcal{B}in(n, P_i^t)$ bajo \tilde{P}_i^t , y una distribucion binomial $\mathcal{B}in(n, \pi(i))$ bajo $\tilde{\pi}$. Recordemos que $t \rightarrow P_i^{2t}$ es decreciente.

La idea es demostrar que, bajo \tilde{P}_i^t , el nmero $N_i(\mathbf{i})$ es significativamente superior a $n\pi(i)$, la media de $N_i(\mathbf{i})$ bajo $\tilde{\pi}$. Para esto, definamos:

$$\mathbf{F}_t = \{i : N_i(i) < n(\pi(i) + \frac{1}{2}[P_i^t(i) - \pi(i)])\}$$

Recordemos que $P_i^t(i) = \sum_{l=1, r} V_l^2(i)\alpha_l^t$. Guardamos $t = t(n)$ par y $c = c(n)$ de manera tal que $t = (\log n - c)/(2 \log(1/\alpha))$. Entonces,

$$P_i^t(i) = \pi(i) + n^{-\frac{1}{2}}e^{\frac{1}{2}c}w_i + o(n^{-\frac{1}{2}})$$

Para este valor de t , $E(N_i)$ excede $n\pi(i)$ en el orden \sqrt{n} . Resta utilizar la desigualdad de Chernov para distribuciones binomiales :

Lema 4.1 Sea $B \sim \mathcal{B}in(n, p)$, $b \in]0, 1[$, $h(p, b) = (\frac{1-p}{1-b})^{1-b}(\frac{p}{b})^b$. Entonces: $P(B > nb) < h^n(p, b)$ si $b > p$ y $P(B < nb) < h^n(p, b)$ si $b < p$.

Aplicando el lema a $p = \pi(i)$ (resp. a $p = P_i^t(i)$) y $b = \pi(i) + \frac{1}{2}[P_i^t(i) - \pi(i)]$, se obtiene $\tilde{\pi}(\tilde{E} \setminus \mathbf{F}_t)$ (resp. $\tilde{P}_i^t(\mathbf{F}_t)$) $< \exp\{\varepsilon(c, i) + o(1)\}$. De donde se concluye el resultado. \square

Podemos preguntarnos si existe un control uniforme en el estado inicial $\mathbf{i} = (i_1, i_2, \dots, i_n)$ de la cadena. La repuesta es si (Proposicin 2.4, [110]): para toda funcion $c(n) \rightarrow +\infty$, poniendo $t^-(n) = \max\{0, \frac{\log n}{2 \log(1/\alpha)} - c(n)\}$, $t^+(n) = \frac{\log n}{2 \log(1/\alpha)} + c(n)$, se verifica:

$$\lim_{n \rightarrow \infty} \left\| \tilde{P}_i^{t^-(n)} - \tilde{\pi} \right\| = 1 \text{ y } \lim_{n \rightarrow \infty} \left\| \tilde{P}_i^{t^+(n)} - \tilde{\pi} \right\| = 0$$

dado que $\mathbf{i} = (i_1, i_2, \dots, i_n)$, cada modalidad $i \in E$ es de frecuencia asinttica > 0 en \mathbf{i} .

4.1.2. Ejemplo 2 (continuacin) : tiempo de fusin y regla de parada [109]

Sea $f : E \rightarrow \mathbb{R}$ y $\{X(t)\}_{t \geq 0} = \{X_1(t), X_2(t), \dots, X_n(t)\}_{t \geq 0}$ la cadena que comienza en $\tilde{i} = (i, i, \dots, i)$. La media empirica de f al tiempo t es:

$$S_i^{(t)}(f) = \frac{1}{n} \sum_{m=1}^t f(X_m(t))$$

Definición 4.1 Sean i_1 y i_2 dos estados tales que $f(i_1) < E_\pi(f) < f(i_2)$. El tiempo de fusin (merging time) de la cadena asociada a (i_1, i_2, f) es la variable aleatoria :

$$T_{i_1, i_2}(f) = \inf\{t > 0 : S_{i_1}^{(t)}(f) \geq S_{i_2}^{(t)}(f)\}$$

Sea $w_i(f) = \sum_{l:|\alpha_l|=\alpha} \sum_{j \in E} f(j) \sqrt{\frac{\pi_j}{\pi_i}} V_l(i) V_l(j)$ y $\langle f, P_i^t \rangle$ la esperanza de f bajo P_i^t .

Proposición 4.2 ([109] y [110]) Si las dos condiciones siguientes se verifican:

- (a) uno de los dos valores $w_{i_1}(f)$ y $w_{i_2}(f)$ no es cero,
- (b) las dos funciones $\langle f, P_{i_1}^t \rangle$ y $\langle f, P_{i_2}^t \rangle$ son crecientes en t ;

Entonces:

$$\lim_{n \rightarrow \infty} T_{i_1, i_2}(f) \left\{ \frac{\log 2n}{2 \log(1/\alpha)} \right\}^{-1} = 1 \text{ en Probabilidad}$$

Este resultado permite dar una regla de parada que asegura la convergencia de la cadena \tilde{P} : si $\varepsilon > 0$ y si n es grande, parar la cadena en $T = (1 + \varepsilon)T_{i_1, i_2}(f)$ asegura la proximidad de \tilde{P}^T y de π . Sealemos que (b) se verifica cuando P se reemplaza por P^2 (los valores propios de P^2 son positivos, si se itera la cadena por pasos de 2 y se detiene la cadena en un instante par). Las condiciones $f(i_1) < E_\pi(f) < f(i_2)$ sobre (f, i_1, i_2) y (b) se obtienen por ejemplo si E posee un orden parcial \prec , si f es montona, si i_1 es minimal y si i_2 es maximal en E . Por ejemplo, sobre $E = \{-1, +1\}^n$ para el orden:

$$x \prec y \Leftrightarrow \forall i = 1, n : x_i \leq y_i$$

$-1 = (x_i \equiv -1)$ es minimal, $+1 = (x_i \equiv +1)$ es maximal se puede seleccionar $f(x) = \sum_i x_i$ que es creciente.

Diferentes contextos de aplicacin ⁽⁵⁾ se desarrollan en [110].

4.2. Simulacin exacta por acoplamiento desde el pasado

4.2.1. Acoplamiento desde el pasado

Expondremos aqu los resultados principales ⁽⁶⁾ de J.G. Propp y D.B. Wilson ([84], 1996). Sea P una transicin ergdica sobre $E = \{1, 2, \dots, r\}$, de distribucin invariante π . Se quiere simular π segn la dinmica P . Llamemos $P(i, \cdot)$ la distribucin de la cadena que sale de i luego de un paso.

Las simulaciones (\mathcal{S}). El acoplamiento desde el pasado se define a partir de una sucesin \mathcal{S} de simuladores

$$\mathcal{S} = \{f_{-t}(i), t \geq 1, i \in E\}$$

donde $f_{-t}(i)$ tiene como distribucin $P(i, \cdot)$. Supondremos los generadores $\mathcal{S}_t = \{f_{-t}(i), i \in E\}$ i.i.d. en el tiempo $t \geq 1$ (W. Kendall y J. Møller debilitan esta condicin, [62]). En cuanto a las variables $f_{-t}(i), i \in E$ (t est fijo), ellas no son necesariamente independientes. Veremos tambn que es posible e interesante que todas ellas provengan de un nico generador U_t uniforme sobre $[0, 1]$. El algoritmo evoluciona desde el pasado:

⁵Cutoff para una cadena con espacio numerable, cadena de nacimientos y de muertes sobre un rbol, sobre R , red de Jackson

⁶Propp y Wilson estudian aplicaciones en combinatoria, en mecnic estadstica y en cubrimientos (Ising, random-cluster, ice, y dimer models).

• *Iteración entre $-t$ y $-t+1$ ($t \geq 2$) :*

- Para cada $i \in E$, simular $i \mapsto f_{-t}(i) \in E$
- Se obtienen as r -transiciones $\{i \mapsto f_{-t}(i), i \in E\}$ de $-t$ a $-t+1$.

El flujo de t_1 a t_2 , $t_1 < t_2$: $F_{t_1}^{t_2} : E \rightarrow E$.

Es la transformación aleatoria:

$$F_{t_1}^{t_2} = f_{t_2-1} \circ f_{t_2-2} \cdots \circ f_{-t_1+1} \circ f_{t_1}$$

$F_{t_1}^{t_2}(i)$ es el estado al tiempo t_2 de la cadena que ha comenzado en i al tiempo t_1 . As, F_t^0 ($t < 0$), el flujo entre el instante t y 0, verifica:

$$F_t^0 = F_{t+1}^0 \circ f_t, \text{ con } F_0^0 = Id \quad (4.1)$$

Esta última relación muestra que F_t^0 se obtiene recursivamente utilizando r lugares de memoria.

Tiempo y estado de coalescencia para (\mathcal{S}) .

$T < 0$ es un tiempo de coalescencia del algoritmo si F_T^0 es constante,

$$\exists i_* \in E \text{ t.q. } : \forall i \in E, F_T^0(i) = i_*$$

Las r -cadenas, que comienzan en los r -estados al instante T , se acoplan todas al instante 0: se dice de nuevo que T es un *instante de acoplamiento desde el pasado* (coupling from the past, CFTP), i_* el estado de coalescencia. La relación (4.1) nos dice que si $T' < T$, entonces T' es todavía un tiempo de coalescencia, el estado de coalescencia al instante 0 permanece intercambiado. Llamemos $F_{-\infty}^0 = F_{-T_*}^0 = F_{-M}^0$ para todo $M \geq T_*$ el valor constante y común obtenido para el primer instante de acoplamiento, llamado T_* ,

$$T_* = \inf\{t > 0 : F_{-t}^0(\cdot) \text{ es constante}\}$$

Proposición 4.3 *Simulación exacta por CFTP(\mathcal{S})*

Casi seguramente, el acoplamiento tiene lugar en un tiempo finito : $P(T_ < \infty) = 1$. Además, $i_* = F_{-\infty}^0$ se distribuye siguiendo π .*

Demostración : Como la cadena es ergódica, existe $L < \infty$ tal que, para todo i, j , $P^L(i, j) > 0$. As, F_{t-L}^t tiene una probabilidad $\varepsilon > 0$ de ser constante. Cada una de las aplicaciones $F_{-L}^0, F_{-2L}^{-L}, \dots$ tiene probabilidad ε de ser constante, y como estos eventos son independientes, al menos uno se cumple con probabilidad 1. As c.s. F_{-M}^0 es constante si M es lo suficientemente grande. Se tiene $P(T_* < \infty) = 1$.

La sucesión $\mathcal{S}_t = \{f_{-t}(i), i \in I\}$, $t = -1, -2, \dots$ es *i.i.d.*, $F_{-\infty}^{-1}$ a $F_{-\infty}^0$ tienen la misma distribución ν . Por otra parte, se pasa de una a otra por P . Se tiene entonces $\nu P = \nu$, esto es $\nu = \pi$. \square

Este resultado ha sido extendido a un contexto mucho más general por W. Kendall y J. Møller (Teorema 3.1,[62]).

Comentarios.

(1) La distribución de la variable $\mathcal{S}_1 = \{f_{-1}(i), i \in I\}$ es el germen de la simulación. El test de parada es implícito al procedimiento.

(2) *Complejidad algorítmica.* Para obtener F_t^0 , es necesario efectuar $(-t) \times r$ operaciones, lo que es irrealista si r es grande. Igualmente, r lugares de memoria serían necesarios. Dos procedimientos son posibles: (i) las transiciones $\mathcal{S}_t = \{f_{-t}(i), i \in I\}$ pueden ser construidas a partir de un nico germen aleatorio U_{-t} , limitando así el número de llamadas a un generador aleatorio; (ii) en el caso de un espacio E con un orden parcial \prec y de una cadena montona (nocin que definiremos luego), ser suficiente probar la igualdad $F_T^0(i) = F_T^0(\bar{i})$ para *dos estados solamente*.

(3) El procedimiento MCMC es distinto: en efecto, se escoge un estado inicial i_0 y se calcula $F_0^n(i_0)$ para n grande: *una sola trayectoria es simulada*. La dificultad está en la selección de n .

(4) Podemos pensar que un buen procedimiento consiste en simular, desde el instante 0 (es decir hacia el futuro), r cadenas comenzando de r estados, hasta el primer instante T^* donde ellas se acoplan, es decir tal que $F_0^{T^*}$ es constante. Esta es la idea estudiada por Jonhson [54] (cf. también [32]). La ventaja consiste en hacer desaparecer el sesgo de la selección de un punto inicial particular i_0 , la complejidad algorítmica se sitúa entre T^* (la de MCMC) y $r \times T^*$ (la de CFTP) (en efecto a medida que las trayectorias se acoplan, se disminuye el número de operaciones a efectuar). Esta idea es tanto más natural que T_* y T^* tienen, como veremos, la misma distribución. Sin embargo, $F_0^{T^*}$ y $F_{-T^*}^0$ no tienen la misma distribución, y en general, $F_0^{T^*}$ no posee la distribución π . Para convencerse, es necesario considerar una transición P tal que existe un estado j_0 que tiene un solo predecesor: $F_0^{T^*}$ no carga a j_0 por lo tanto no puede tener la distribución π , porque $\pi(j_0) > 0$ (cf. ejercicio).

(5) Lo que distingue el acoplamiento hacia el pasado del acoplamiento hacia el futuro es que, para el acoplamiento hacia el pasado, se puede escribir el estado de coalescencia, $x(0) = F_{-\infty}^0(\cdot)$: es como si $x(0)$ fuese el estado al tiempo 0 de la cadena que comienza en $-\infty$. Este estado tiene distribución π . Una escritura e interpretación de este tipo no es posible para el estado de acoplamiento hacia el futuro, ya que la cadena comienza en el tiempo fijo 0.

4.2.2. Un nico germen para definir $\mathcal{S}_{-t} = \{f_{-t}(i), i \in I\}$

Sean $\dots, U_{-3}, U_{-2}, U_{-1}$ una sucesión i.i.d. de variables uniformes sobre $[0, 1]$, y $\Phi : E \times [0, 1] \rightarrow [0, 1]$ una aplicación medible tal que

$$\forall i, j : P\{\Phi(i, U_{-1}) = j\} = P(i, j)$$

Se puede entonces seleccionar, para todo $i \in E : f_{-t}(i) = \Phi(i, U_t)$. En cada instante, una sola simulación es necesaria. *Es necesario no obstante, evaluar todos los valores $\Phi(i, U_t)$ para $i \in E$.*

El caso “favorable” de una *cadena montona* permite eliminar esta dificultad: en este caso, veremos que es suficiente seguir solo dos trayectorias para identificar el acoplamiento. Esto reduce considerablemente el número de operaciones a efectuar así como la dimensión de la memoria til.

4.2.3. Algoritmo de Monte-Carlo montono

Presentación general

Supongamos que E está provisto una relación de orden parcial, llamada $x \prec y$, para la cual existe un elemento minimal $\mathbf{0}$ y un elemento maximal $\mathbf{1} : \forall x \in E, \mathbf{0} \prec x \prec \mathbf{1}$. Supongamos además que la regla de actualización preserva este orden, es decir:

$$\forall x, y \in E \text{ t.q. } x \prec y \text{ entonces: } \forall u \in [0, 1] : \Phi(x, u) \leq \Phi(y, u)$$

Diremos que este *algoritmo de Monte-Carlo* es *montono*. Definamos entonces:

$$\Phi_{t_1}^{t_2}(x, \mathbf{u}) = \Phi_{t_2-1}(\Phi_{t_2-2}(\dots \Phi_{t_1}(x, u_{t_1}), \dots, u_{t_2-2}), u_{t_2-1})$$

donde $\mathbf{u} = (\dots, u_{-1}, u_0)$. La monotona del algoritmo asegura que :

$$\forall x \prec y \text{ y } t_1 < t_2, \Phi_{t_1}^{t_2}(x, \mathbf{u}) \leq \Phi_{t_1}^{t_2}(y, \mathbf{u})$$

En particular, si $u_{-T}, u_{-T+1}, \dots, u_{-1}$ son tales que $\Phi_{-T}^0(\mathbf{0}, \mathbf{u}) = \Phi_{-T}^0(\mathbf{1}, \mathbf{u})$, entonces $-T$ es un tiempo de acoplamiento desde el pasado. Estas dos configuraciones $\mathbf{0}$ y $\mathbf{1}$ caracterizan el tiempo de acoplamiento. En efecto, para todo x , las trayectorias $\{F_{-T}^t(x), -T \leq t \leq 0\}$ son "ensanguchadas" ([61]) entre $\{F_{-T}^t(\mathbf{0}), -T \leq t \leq 0\}$ y $\{F_{-T}^t(\mathbf{1}), -T \leq t \leq 0\}$.

Complejidad del algoritmo.

Propp y Wilson proponen el algoritmo siguiente. Se prueba sucesivamente $T = 1, 2, 4, 8, \dots$ hasta conseguir un valor 2^k tal que $\Phi_{-2^k}^0(\mathbf{0}, \mathbf{u}) = \Phi_{-2^k}^0(\mathbf{1}, \mathbf{u})$. Los valores de u_t utilizados deben ser los mismos para dos ensayos que hacen intervenir un mismo instante t . Estos valores de u_t son entonces progresivamente colocados en la memoria.

Para este algoritmo, el nmero de operaciones necesarias para probar la coalescencia es $2 \times (1 + 2 + 4 + \dots + 2^k) \approx 2^{k+2}$ (es necesario seguir dos trayectorias, tantas veces como ensayos). Como $T_* > 2^{k-1}$, este nmero de operaciones es a lo sumo 4 veces el nmero ptimo $2T_*$: el procedimiento es razonable.

Ejemplo de un modelo de Ising

Los sistemas de spins atractivos son ejemplos-tipos de modelos que permiten construir una dinmica de Monte-Carlo montona. El conjunto de sitios es $S = \{1, 2, \dots, n\}$ y el de estados, $E = \{-1, +1\}^S$ ($|E| = r = 2^n$). E posee el orden:

$$x \prec y \Leftrightarrow \{\forall i \in S, x_i \leq y_i\}$$

Para este orden, $\mathbf{0} = (x_i = -1, \forall i)$ es minimal, $\mathbf{1} = (x_i = +1, \forall i)$ es maximal.

Una distribucin π sobre E se llama *atractiva* si, para todo i :

$$x \prec y \Rightarrow \pi_i(+1 | x) \leq \pi_i(+1 | y) \quad (4.2)$$

donde $\pi_i(\cdot | z)$ es la distribucin condicional en z (realmente en z^i) en i . Como no figura en el sitio i donde el intercambio de spin se ha efectuado, y escribiendo $x \uparrow$ (resp. $x \downarrow$) la configuracin $(+1, x^i)$ (resp. $(-1, x^i)$), la relacin (4.2) es equivalente a

$$x \prec y \Rightarrow \frac{\pi(x \downarrow)}{\pi(x \downarrow) + \pi(x \uparrow)} \geq \frac{\pi(y \downarrow)}{\pi(y \downarrow) + \pi(y \uparrow)}$$

Consideremos entonces el algoritmo de Gibbs (secuencial determinista o de barrido aleatorio), y definamos:

$$f_t(x, u_t) = \begin{cases} x \downarrow & \text{si } u_t < \frac{\pi(x \downarrow)}{\pi(x \downarrow) + \pi(x \uparrow)} \\ x \uparrow & \text{si } u_t \geq \frac{\pi(x \downarrow)}{\pi(x \downarrow) + \pi(x \uparrow)} \end{cases}$$

Si π es atractiva, los f_t definen un algoritmo de Monte-Carlo montono porque $f_t(x, u_t)(i) = +1$ y $f_t(y, u_t)(i) = -1$ son incompatibles para $x \prec y$.

Si se especifica π como la distribucin de un modelo de Ising de energia $U(x)$

$$\pi(x) = Z^{-1} \exp\{U(x)\} \text{ con } U(x) = \sum_i \alpha_i x_i + \sum_{i < j} \beta_{i,j} x_i x_j$$

la propiedad de ser atractiva es equivalente a la condicin : $\forall i, j, \beta_{i,j} \geq 0$. Si nos colocamos en la temperatura critica ($\beta \simeq 0,441$) para un modelo de Ising isotrpico y de 4-v.m.c. sobre una malla 512×512 , Propp y Wilson observan que aproximadamente $T = 30$ barridos hacia el pasado son suficientes para la simulacin CFTP del modelo (esto es 10 min. sobre una estacin Sparc).

Otro inters de un algoritmo de Monte-Carlo montono es que permite simular *simultneamente* muestras de una distribucin π_T dependiente de un parmetro de T , para diversos valores de T (.°mnithermal algorithm”,[84]).

Evaluacin del tiempo de acoplamiento

Instantes de acoplamiento hacia el pasado o hacia el futuro. El tiempo de acoplamiento T_* hacia el pasado es el primer instante tal que $F_{-t}^0(\mathbf{0}) = F_{-t}^0(\mathbf{1})$. El tiempo de acoplamiento T^* hacia el futuro es el primer $t > 0$ tal que $F_0^t(\mathbf{0}) = F_0^t(\mathbf{1})$. *Estos dos tiempos tienen la misma distribucin* : en efecto, para cada $t > 0$, $P(T_* > t)$, la probabilidad que F_{-t}^0 no sea constante, es igual a la probabilidad de que F_0^t no sea constante porque se pasa de F_{-t}^0 a F_0^t por la traslacin de $+t$ en el tiempo y que $\mathcal{S} = \{\mathcal{S}_t\}$ es estacionaria (cuidado, $F_{-T_*}^0$ y $F_0^{T_*}$ no tienen las mismas distribuciones, cf. ejercicio).

Eficiencia de un algoritmo de acoplamiento hacia el pasado Como la duracin del algoritmo es lineal en T_* , el algoritmo ser tanto ms eficiente cuando ms pequeo sea T_* . Otra manera de comprender este resultado es la siguiente: el tiempo de acoplamiento T^* controla la convergencia MCMC clsica, porque la desigualdad de acoplamiento da

$$\left\| \mu P^k - \pi \right\| \leq P(T^* > k)$$

Ser necesario entonces, dentro de la multiplicidad de las selecciones $\mathcal{S} = \{\mathcal{S}_t\}$ posibles, guardar aquellas que generan trayectorias con una fuerte probabilidad de coalescencia. Se habla de acoplamientos ptimos (cf. [46], [83], [56]).

Algunas evaluaciones sobre el tiempo de acoplamiento T_* . Permaneceremos dentro del contexto de un algoritmo de Monte-Carlo montono. Trabajaremos con el tiempo de acoplamiento hacia el futuro T^* , conceptualmente ms simple.

Algunas notaciones :

- Para $k > 0$, $\bar{d}(k) = \sup\{\|\mu P^k - \nu P^k\|, \mu, \nu \text{ distribuciones iniciales}\}$
- el tiempo de mezcla T_{mix} de P es el primer ndice k tal que $\bar{d}(k) < e^{-1}$
- l es la longitud mxima de una cadena que respeta el orden parcial: $l = \text{m}ax\{k : \text{existe un camino } x_1 \rightarrow x_2 \cdots \rightarrow x_k \text{ t.q. } x_l \prec x_{l+1} \text{ para } l = 1, k-1\}$

Proposicin 4.4 *Evaluacin de la distribucin de T^* (o de la distribucin de T_*).*

- (1) T_* y T^* tienen la misma distribucin.
- (2) $l^{-1}P(T^* > k) \leq \bar{d}(k) \leq P(T^* > k)$
- (3) $P(T^* > k)$ es submultiplicativa :

$$P(T^* > k_1 + k_2) \leq P(T^* > k_1)P(T^* > k_2)$$

- (4) $\forall k : k \times P(T^* > k) \leq E(T^*) \leq k/P(T^* > k)$

Demostracin:

- (1) Es una consecuencia de la estacionaridad de $\mathcal{S} = (\mathcal{S}_t)$.

(2) Sea $x \in E$, y $h(x)$ la longitud de la cadena ms larga de vrtice x . Sean $X_0(k)$ (resp. $X_1(k)$) el estado de la cadena inicializada en $\mathbf{0}$ (resp. $\mathbf{1}$) despues de k pasos. Si $X_0(k) \neq X_1(k)$, entonces $h(X_0(k)) + 1 \leq h(X_1(k))$. Lo que conduce a:

$$\begin{aligned} P(T^* > k) &= P(X_0(k) \neq X_1(k)) \\ &\leq E[h(X_0(k)) - h(X_1(k))] \\ &= E_{P_0^k}[h(X)] - E_{P_1^k}[h(X)] \\ &\leq \left\| P_0^k - P_1^k \right\| \{ \text{m}ax h(x) - \text{m}in h(x) \} \\ &\leq \bar{d}(k) \times l \end{aligned}$$

Para establecer la otra desigualdad, consideremos el acoplamiento de dos copias de la cadena, una con distribucin inicial μ , la otra con distribucin ν , las dos cadenas se acoplan en el primer instante de encuentro. La monotona implica que las dos copias se encuentran antes de k con una probabilidad al menos igual a $P(T^* \leq k)$. Deducimos que μP^k y νP^k son diferentes sobre un evento de probabilidad $\leq P(T^* > k)$, es decir:

$$\left\| \mu P^k - \nu P^k \right\| \leq P(T^* > k)$$

(3) El evento $\{F_0^{k_1}$ es constante $\}$ y el evento $\{F_{k_1}^{k_1+k_2}$ es constante $\}$ son independientes. Y si uno ocurre, $F_0^{k_1+k_2}$ es constante. De lo que se concluye la submultiplicatividad.

(4) La minoracin es consecuencia de la positividad de T^* . Pongamos $\varepsilon = P(T^* > k)$. Entonces, dada la submultiplicatividad, $P(T^* > ik) \leq \varepsilon^i$, y

$$E(T^*) \leq k + k\varepsilon + k\varepsilon^2 + \dots = k/P(T^* \leq k)$$

□

Estos resultados permiten concluir que: “si la cadena es rpidamente mezclante, las trayectorias se acoplan rpidamente”. En efecto, $\bar{d}(k)$ es, ella tambn, submultiplicativa (cf. captulo 3), y entonces despues de $k = T_{mix}(1 + \log l)$ pasos, $\bar{d}(k) \leq 1/el$, es decir $P(T^* > k) \leq 1/e$, ms an :

$$E(T^*) \leq 2T_{mix}(1 + \log l).$$

Enfoque emprico para la evaluacin de T^ .*

V.E. Johnson [55] estudia empricamente la curva $b \mapsto T^*(b)$ para un modelo de Ising isotrpico de 4-v.m.c. de parmetro $b > 0$ (atractivo) sobre una malla 64×64 :

$$\pi_b(x) = Z^{-1}(b) \exp\left\{ b \sum_{\|i-j\|_1=1} x_i x_j \right\}, x \in \{-1, +1\}^{64 \times 64}$$

El autor estima T^* a partir de 20 simulaciones independientes, obteniendo la estimacin emprica de la curva $b \mapsto E(T^* | b)$ y tambn los histogramas de la muestra $\{T^*(b, k), k = 1, 20\}$ para diversos valores del parmetro de correlacin espacial b (su parametrizacin es de estados $\{0, 1\}$, su parmetro es $\beta = 2b$; el umbral crtico es $b \simeq 0,441$ (valor escogido por Propp y Wilson en su simulacin)). Obtiene los siguientes resultados :

b	0.1	0.2	0.3	0.4	0.5
$T^*(b)$	7	15	50	500	200.000

Para valores moderados de b , como 0,3, T^* es bastante dbil: no es til iterar un gran nmero de barridos del muestreador de Gibbs para alcanzar el estado estacionario. Por el contrario, T^* presenta rpidamente un crecimiento superexponencial ms all del umbral crtico 0,441.

4.3. Ejercicios

Ejercicio 4.1 *Acoplamiento hacia el pasado o hacia el futuro : $F_0^{T^*}$ y $F_{-T^*}^0$ no tienen la misma distribución.*

Consideremos la cadena sobre $E = \{0, 1\}$ de transición $P(0, 0) = P(0, 1) = \frac{1}{2}$, y $P(1, 0) = 1$.

(1) Proponer diferentes trayectorias: (i) que se acoplan hacia el pasado; (ii) que se acoplan hacia el futuro.

(2) Determinar la distribución de T^* , el tiempo de acoplamiento hacia el futuro (esta distribución es la misma que la distribución de T_* , el tiempo de acoplamiento hacia el pasado).

(3) Verificar, suministrando ejemplos de trayectorias simuladas, que $F_{-T^*}^0$ carga los dos estados, pero no $F_0^{T^*}$. Demostrar que $F_0^{T^*} \sim \delta_{\{0\}}$, y que $F_{T^*}^0 \sim \pi = (2/3, 1/3)$.

Hacer el mismo estudio para la cadena con 3 estados de transición $P = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \\ 1/2 & 0 & 1/2 \end{pmatrix}$.

Ejercicio 4.2 *Un caso donde $F_0^{T^*}$ tiene distribución π*

Sea la cadena sobre $E = \{0, 1\}$ ($0 < 1$) de transición, para $p, q \in]0, 1[$, y $1 - q < p$:

$$P = \begin{pmatrix} p & 1 - p \\ 1 - q & q \end{pmatrix}$$

Los dos sorteos asociados a dos filas de $P(0, \cdot)$ y $P(1, \cdot)$ se obtienen a partir de una misma distribución uniforme U sobre $[0, 1]$:

$$\begin{cases} \bullet 0 \mapsto 1 \text{ si } U > p \text{ y } 0 \mapsto 0 \text{ si no} \\ \bullet 1 \mapsto 1 \text{ si } U > 1 - q \text{ y } 1 \mapsto 0 \text{ si no} \end{cases}$$

(1) Verificar que el algoritmo es monótono.

(2) Determinar la distribución de T^* , el tiempo de acoplamiento hacia el futuro.

(3) Verificar que $F_0^{T^*} \sim \pi = (2 - (p + q))^{-1}(1 - q, 1 - p)$, la distribución invariante de P .

Ejercicio 4.3 *Programación de algoritmos de acoplamiento*

(1) Escoger una matriz de transición ergódica P sobre $E = \{1, 2, \dots, r\}$ ($r = 3, 5, 10, 20$). Determinar su distribución invariante π y su espectro. Hacer selecciones para las cuales un algoritmo de Monte-Carlo monótono es posible (para el orden natural sobre E).

(2) Proceder a simular por acoplamiento hacia el pasado. Identificar T_* , el tiempo de acoplamiento hacia el pasado, y el valor retornado $x(0)$. Repetir la experiencia N veces ($N = 100, 1000, \dots$) de manera independiente y evaluar: la distribución empírica de T_* (esperanza, varianza empírica, y también su *min* y su *max*), la estimación $\hat{\pi}$ de π , como la distancia Chi2 $\chi(\pi, \hat{\pi})$.

(3) Rehacer el procedimiento si un algoritmo monótono existe y comparar los resultados.

(4) Para una precisión $\chi(\pi, \hat{\pi})$ análoga, ¿cuántas simulaciones *MCMC* serán necesarias? Proceder a efectuar esta simulación, y repetirla 1000 veces. Comparar los resultados.

(4) Programar el algoritmo de acoplamiento hacia el futuro. Verificar empíricamente que T^* , el tiempo de acoplamiento hacia el futuro, tiene la misma distribución que T_* . Dar la distribución empírica del estado de acoplamiento hacia el futuro.

Los ejercicios que siguen tratan sobre el acoplamiento de variables aleatorias (o de procesos) y fueron extrados del artículo de T.H.T. Thorisson [97]. Una referencia general sobre el acoplamiento es el libro de T. Lindvall [68]. Demos la definición general del acoplamiento de dos distribuciones.

Definición : acoplamiento de dos distribuciones. Sean μ y ν dos distribuciones sobre el mismo espacio medible (E, \mathcal{E}) . Una variable (X, Y) sobre $(E \times E, \mathcal{E} \times \mathcal{E})$ efecta un acoplamiento de μ y de ν si $X \sim \mu$ y $Y \sim \nu$.

Existen siempre acoplamientos: por ejemplo el acoplamiento independiente $\mu \otimes \nu$, el acoplamiento dependiente a partir de una misma variable uniforme U ($X = F_\mu^{-1}(U)$ y $Y = F_\nu^{-1}(U)$), pasando por muchas otras posibilidades. Este “grado de libertad” en la selección del acoplamiento permite obtener resultados ms o menos interesantes. La dificultad de un mtodo de acoplamiento es : “como seleccionar un acoplamiento bien adaptado al problema considerado”. Recordemos que la distancia en variación entre μ y ν est definida por :

$$\|\mu - \nu\| = \sup\{|\mu(A) - \nu(A)| : A \in \mathcal{E}\}$$

Ejercicio 4.4 Desigualdad de acoplamiento

Sea (X, Y) un acoplamiento de μ y de ν . Demostrar la desigualdad de acoplamiento :

$$\|\mu - \nu\| \leq P(X \neq Y).$$

Si (X, Y) verifica la igualdad, se dice que el acoplamiento es maximal. Existe siempre un acoplamiento maximal.

Ejercicio 4.5 Acoplamiento de una distribución binomial y de una distribución de Poisson con la misma media

(1) Acoplamiento de una distribución de Bernoulli y de una distribución de Poisson.

Sea $p \in]0, 1[$, $X \sim \mathcal{Ber}(p)$ y $Y \sim \mathcal{P}(p)$. Sea $I \in \{0, 1\}$, independiente de X y de Y , definida por : $P(I = 0) = e^p(1 - p)$. Pongamos :

$$\begin{cases} X' = 0 \Leftrightarrow Y = 0 \text{ y } I = 0 \\ X' = 1 \Leftrightarrow Y > 0 \text{ donde } I = 1 \end{cases}$$

Verificar que (X', Y) es un acoplamiento de X y de Y (es realmente un acoplamiento maximal). Verificar que : $P(X' \neq Y) = p(1 - e^{-p}) \leq p^2$.

(2) Sean $X_i \sim \mathcal{Ber}(p_i)$, $i = 1, n$, n distribuciones de Bernoulli independientes, $Y_i \sim \mathcal{P}(p_i)$, $i = 1, n$, n distribuciones de Poisson independientes, $X = \sum_i X_i$ y $Y = \sum_i Y_i$ (Y es una distribución de Poisson de media $\sum_i p_i$). Proponer un acoplamiento de X y de Y a partir de n -acoplamientos (X'_i, Y_i) de X_i y de Y_i , $i = 1, n$. Utilizando la inclusin :

$$(X' \neq Y) \subseteq \bigcup_{i=1}^n (X'_i \neq Y_i),$$

deducir la desigualdad de Chen-Stein :

$$\|\mathcal{L}(X) - \mathcal{L}(Y)\| \leq 2 \sum_{i=1}^n p_i^2.$$

(3) Aplicacin :

$$\|\mathcal{Bin}(n, p) - \mathcal{P}(np)\| \leq 2np^2 = 2 \frac{\lambda^2}{n} \text{ si } \lambda = np.$$

Ejercicio 4.6 *Dominación estocástica*

Para dos variables aleatorias reales, se llama $X \preceq Y$ la relación de dominación estocástica: para todo $x \in \mathbb{R}$, $F_X(x) \leq F_Y(x)$.

Sean X y Y dos distribuciones tales que $X \preceq Y$, $(X_i, i = 1, n)$ (resp. $(Y_i, i = 1, n)$) una n -muestra de X (resp. una n -muestra de Y), $S_X = \sum X_i$ y $S_Y = \sum Y_i$. Demostrar que $S_X \preceq S_Y$. *Indicación*: usaremos los acoplamientos elementales dependientes $(F_X^{-1}(U_i), F_Y^{-1}(U_i))$ de X_i y de Y_i , $i = 1, n$, donde $(U_i, i = 1, n)$ es una muestra $\mathcal{U}[0, 1]$.

Ejercicio 4.7 *Acoplamiento maximal de dos cadenas ([46], [83])*

Para dos cadenas discretas de transición P sobre (E, \mathcal{E}) , verificar que la transición \mathbf{P} siguiente es un acoplamiento de P y de P sobre $(E, \mathcal{E}) \otimes (E, \mathcal{E})$:

$$\begin{cases} \mathbf{P}[(i, i), (k, k)] = p_{ik} \\ \mathbf{P}[(i, j), (k, k)] = \inf\{p_{ik}, p_{jk}\} \\ \mathbf{P}[(i, j), (k, l)] = \delta_{ij}^{-1} \{(p_{ik} - p_{jk})^+ (p_{il} - p_{jl})^-\} \end{cases}$$

donde $\delta_{ij} = \frac{1}{2} \sum_k |p_{ik} - p_{jk}|$. Este acoplamiento es maximal.

Ejercicio 4.8 *El acoplamiento multigamma ([68],[56],[80])*

El espacio de estados es $E \subset \mathbb{R}^d$. Sea $f(\cdot | x)$ una densidad de transición difusa y ergódica con distribución invariante π . Supongamos que:

$$\forall x, y \in E : f(y | x) \geq r(y) \geq 0 \text{ con } 0 < \rho = \int r(y) dy < 1.$$

Sean $R(y) = \rho^{-1} \int r(u) du$, $Q(y | x) = (1 - \rho)^{-1} \int_{-\infty}^y [f(u | x) - r(u)] du$, U_1 y U_2 dos distribuciones uniformes independientes sobre $[0, 1]$. Se define la función de actualización Φ :

$$\Phi(x, U_1, U_2) = \begin{cases} R^{-1}(U_2) & \text{si } U_1 < \rho \\ Q^{-1}(U_2 | x) & \text{si no} \end{cases}$$

(1) Verificar que: $P(Y_{n+1} \leq y | Y_n = x) = \rho R(y) + (1 - \rho)Q(y | x)$. Deducir que $\Phi(x, U_1, U_2)$ tiene distribución $f(\cdot | x)$. Si hay acoplamiento, “se hace en un solo paso”, ¡para todas las trayectorias!

(2) Demostrar que el tiempo de acoplamiento (hacia el pasado), usando esta actualización, es c.s. finito.

(3) Dado que R^{-1} se llama con una probabilidad fija, igual a ρ , se puede proponer el algoritmo siguiente de simulación exacta de π : (i) sortear T_* el tiempo de acoplamiento con distribución geométrica de parámetro ρ ; (ii) sortear Y_{-T_*} con distribución $r(\cdot)/\rho$; (iii) simular hacia el futuro la distribución de la cadena con transición $Q^{-1}(U_2 | Y_{t-1}) \rightarrow Y_t$. Verificar que $Y_0 \sim \pi$.

Comentarios: este acoplamiento permite extender los resultados de Propp y Wilson a un espacio de estados continuo. Es necesario sin embargo notar que el campo de aplicación es limitado: (1) la función de minoración $r(\cdot)$ tiene el riesgo de ser muy pequeña en la práctica (la esperanza del tiempo de acoplamiento es ρ^{-1}), y a menudo $r(\cdot) \equiv 0$; (2) es necesario conocer explícitamente las constantes de normalización que intervienen en las densidades. Murdoch y Green [80] responden en parte a estas preguntas.

Ejercicio 4.9 *Control de la velocidad de convergencia $\|\nu P^n - \pi\|$ por acoplamiento*

Sea P una transición sobre un espacio general (E, \mathcal{E}) , verificando, para ν una probabilidad sobre (E, \mathcal{E}) :

$$\exists \varepsilon > 0 \text{ t.q.}, \forall x \in E, \forall A \in \mathcal{E} : P(x, A) \geq \varepsilon \nu(A)$$

Sea $x \in E$ un estado fijo, X^* la distribución de la cadena que comienza en x , Y^* la distribución de la cadena que comienza a partir de su distribución invariante π . Vamos a construir un acoplamiento multigamma $(X, Y) = (X_n, Y_n)_{n \geq 0}$ de X^* y de Y^* (cf. ejercicio precedente). En $n = 0$, tomamos $X_0 = x$ y $Y_0 \sim \pi$. La simulación en el paso $n \mapsto n + 1$ es la siguiente:

$$\begin{cases} (a) \text{ con probabilidad } \varepsilon : \text{ tomar } X_{n+1} = Y_{n+1} = y \text{ donde } y \sim \nu \\ (b) \text{ si no tomar } X_{n+1} \sim \frac{P(X_{n,\cdot}) - \varepsilon \nu(\cdot)}{1 - \varepsilon}, Y_{n+1} \sim \frac{P(Y_{n,\cdot}) - \varepsilon \nu(\cdot)}{1 - \varepsilon} \end{cases}$$

las dos simulaciones de (b) provienen de la misma distribución uniforme. Las cadenas van a terminar por acoplarse en el sentido siguiente: luego de encontrarse, digamos al instante de acoplamiento T (no importa que la causa sea (a) o (b)), continúan sus trayectorias unidas: $\forall t \geq T, X_t \equiv Y_t$.

- (1) Verificar que (X, Y) es un acoplamiento de X^* y de Y^* .
- (2) Usando la desigualdad de acoplamiento, demostrar que :

$$\|P_x^n - \pi\| \leq P(T > n) \leq (1 - \varepsilon)^n$$

(3) Si $E = \{1, 2, \dots, r\}$ es finito si P es regular ($\exists m$ t.q. $\delta = \inf\{P^m(i, j), i, j \in E\} > 0$), deducir de (2) que

$$\|P_x^n - \pi\| \leq (1 - r\delta)^{\lfloor \frac{n}{m} \rfloor}.$$

Capítulo 5

Campos de Markov, problemas inversos y reconstrucción bayesiana

En este capítulo veremos cómo la asociación de un modelo markoviano con un criterio de selección bayesiano, nos proporciona una herramienta matemáticamente rigurosa, así como numéricamente eficaz, para la resolución de problemas inversos.

Comenzaremos por definir la noción de campo de Gibbs y su relación con los campos de Markov. En primer lugar, daremos modelos de reticulados definidos en un conjunto finito de sitios $S = \{1, 2, \dots, n\}$. Luego introduciremos la noción de procesos markovianos de objetos, que generalizan a los procesos puntuales de Markov y que son conceptos básicos para la geometría aleatoria.

A través de algunos ejemplos de problemas inversos en la teoría de señales y de imágenes, se mostrará cómo el modelado markoviano puede capturar adecuadamente la información *a priori* que se tiene sobre el problema, la cual es fundamental para su resolución. Luego, veremos cómo los algoritmos aleatorios, de simulación o de optimización, cuya construcción está relacionada con nuestro criterio, dan métodos numéricos estructurados y eficientes para resolver un problema inverso.

5.1. Modelo de Gibbs y campo de Markov en un conjunto finito

Sea $S = \{1, 2, \dots, n\}$ un conjunto finito dotado de una medida referencial ν . Si F es discreto, se escoge como \mathcal{F} a la σ -álgebra discreta y si F es finito, entonces ν será la medida de contar. Si $F = \mathbb{R}^p$, entonces \mathcal{F} será la σ -álgebra de Borel y ν será en general, una medida absolutamente continua con respecto a la medida de Lebesgue. Sea $E = F^S$, $\mathcal{E} = \mathcal{F}^{\otimes S}$ y $\mu = \nu^{\otimes S}$. Una configuración en S se denota como $x = (x_1, x_2, \dots, x_n)$, y, z .

Referencias: Para conocer más sobre modelos de Gibbs en una red finita, se recomienda revisar los siguientes textos: Besag ([10]), Kinderman y Snell ([64]), Prum ([85]) y Guyon ([47]). El delicado tema de los campos de Gibbs en redes infinitas, escapa a los objetivos de este curso; no obstante, siempre se puede consultar el libro de Georgii [41], que es fundamental para el tema.

5.1.1. Modelo de Gibbs

Sea U una aplicación medible de E en \mathbb{R} . Un *modelo de Gibbs* en (E, \mathcal{E}) de energía U es una distribución π absolutamente continua con respecto a μ , de densidad π tal que

$$\pi(x) = Z^{-1} \exp U(x)$$

Esta notacin slo tiene sentido cuando la energia U es *admisibile*; es decir, si

$$Z = \int_E \exp\{U(y)\} \mu(dy) < \infty.$$

Una densidad de este tipo es estrictamente positiva. Al revs, toda densidad estrictamente positiva se escribe de esta manera, tomando como funcin de energia $U = \log \pi$. Especifiquemos ahora el modelo.

Sea \mathcal{C} una familia de subconjuntos de S que incluye, por convencin, al conjunto vaco y a los subconjuntos de un solo elemento de S . Un subconjunto cualquiera A de \mathcal{C} se denomina una *claque*. Diremos que U es la *derivada del potencial* $\Phi = \{\Phi_A, A \in \mathcal{C}\}$ si

$$U = \sum_{A \in \mathcal{C}} \Phi_A$$

donde $\Phi_A : F^A \rightarrow \mathbb{R}$, el *potencial* restringido a la claque $A \subset S$, slo depende de las coordenadas de x en A , $\Phi_A(x) = \Phi_A(x_A)$. Denotaremos como π_Φ el campo asociado a Φ ; mas, si no hay lugar a ambigedades, omitiremos el subndice Φ . Si se desea generar un modelo no saturado, escogeremos \mathcal{C} diferente de $\mathcal{P}(S)$. Cuando $A = \{i\}$ (resp. $A = \{i, j\}$, $A = \{i, j, k\}$...), nos referiremos a un potencial de conjunto unitario (resp. un potencial de pares, de tripletas, ...).

5.1.2. Especificacin condicional y propiedad de Markov

Distribuciones condicionales

Las *distribuciones condicionales* $\pi_B(\cdot | x^B)$, $B \subset S$, provienen de la energia

$$U_B(x_B | x^B) = \sum_{A \in \mathcal{C}: A \cap B \neq \emptyset} \Phi_A(x)$$

$$\pi_B(x_B | x^B) = Z_B^{-1}(x^B) \exp U_B(x_B | x^B)$$

donde $Z_B(x^B) = \sum_{a_B} \exp U_B(a_B | x^B)$ es la constante de normalizacin.

Grafo de Markov de un modelo de Gibbs

A una familia de clagues \mathcal{C} se le asocia el *grafo de vecindades* de Markov $\mathcal{G} = \mathcal{G}(\mathcal{C})$ simtrico siguiente, donde $\langle i, j \rangle$ representa la existencia de una arista entre i y j ,

$$\text{Si } i \neq j \text{ en } S: \langle i, j \rangle \iff \exists A \in \mathcal{C} \text{ tal que } \{i, j\} \subset A \quad (5.1)$$

Sea $\partial B = \{j \in S : \exists i \in B \text{ t.q. } \langle i, j \rangle\}$ la *frontera de la vecindad* de B . Es fcilmente verificable que la distribucin condicional $\pi_B(\cdot | x^B)$ slo depende de x^B a travs de $x_{\partial B}$. En particular, $\pi_i(\cdot | x^i)$ depende de x_j slamente cuando j es un vecino de i ,

$$\pi_i(x_i | x^i) = \pi_i(x_i | x_{\partial i}) Z_i^{-1}(x_{\partial i}) \exp U_i(x_i | x_{\partial i})$$

5.1.3. Campo de Markov y campo de Gibbs

Supongamos que F es discreto. Un campo de Markov ϖ sobre $E = F^S$ asociado a un grafo simétrico \mathcal{G} , es una distribución tal que, para todo subconjunto B de S , la distribución condicional es local en el siguiente sentido:

$$\varpi_B(\cdot | x^B) = \varpi_B(\cdot | x_{\partial B})$$

Definamos ahora al conjunto $\mathcal{C}(\mathcal{G})$ de claque asociadas a $\mathcal{G} : \emptyset$ y los conjuntos unitarios pertenecen a $\mathcal{C}(\mathcal{G})$, un subconjunto A de S con por lo menos dos puntos, ser una claque cuando

$$A \in \mathcal{C}(\mathcal{G}) \iff \forall i, j \in A, \langle i, j \rangle \quad (5.2)$$

Proposición 5.1 (Hammersley-Clifford, [10],[85],[47])

(i) Todo campo de Gibbs π con familia de claque \mathcal{C} es un campo de Markov de grafo \mathcal{G} definido por (5.1).

(ii) Recíprocamente, todo campo de Markov ϖ de grafo \mathcal{G} que verifique la condición de positividad: $\forall x \in E, \varpi(x) > 0$, es un campo de Gibbs para la familia de claque \mathcal{C} definida en (5.2).

La relación (i) acaba de ser explicada. Precisemos (ii). Los potenciales de Gibbs ($\Phi_A, A \in \mathcal{C}$) del campo de Markov se obtienen a partir de la fórmula de inversión de Möbius: sea $\mathbf{0}$ un estado de referencia de F ; entonces, si $A \in \mathcal{C}(\mathcal{G})$, y si $\mathbf{0}_B$ es la configuración de $\mathbf{0}$ sobre B ,

$$\Phi_A(x) = \sum_{B \subset A} (-1)^{|A \setminus B|} H^B(x) \quad \text{donde} \quad H^B(x) \log \left\{ \frac{\varpi_B(x_B | x_{\partial B})}{\varpi_B(\mathbf{0}_B | x_{\partial B})} \right\}$$

Identificabilidad de los potenciales. Diremos que la representación en potencial Φ es identificable si $\Phi \rightarrow \pi_\Phi$ es inyectiva. Si la energía U (resp. el potencial Φ_A) se modifica como $U + c$ (resp. como $\Phi_A + c$) donde c es una constante, la distribución π asociada no cambia. Falta pues imponer condiciones para que Φ sea identificable. Es fácil verificar que las restricciones siguientes hacen que Φ sea identificable,

$$\forall A \neq \emptyset, \Phi_A(x) = 0 \text{ para todo } x \in E \text{ siempre que } \exists i \in A \text{ con } x_i = 0 \quad (5.3)$$

Existen otras colecciones de restricciones que permiten garantizar que Φ sea identificable; por ejemplo; las del análisis de la varianza. Esas condiciones de identificabilidad muestran toda su importancia en el análisis estadístico de los modelos log-lineales. En este contexto son secundarias.

5.1.4. Ejemplos

Modelos de Ising en el cubo d -dimensional

Sea $S = \{1, 2, \dots, N\}^d$, $F = \{-1, +1\}$ y sea la relación de vecindad en los $2d$ -v.m.p.: $\langle i, j \rangle \iff \|i - j\|_1 = 1$. El modelo de Ising de campo exterior h y de interacción de pares β , es el modelo de energía

$$U(x) = h \sum_S x_i + \beta \sum_{\langle i, j \rangle} x_i x_j$$

Los auto-modelos de Besag ([10]; [47], §2.2.3)

Definición 5.1 *Un campo π con valores reales ($F \subset \mathbb{R}$) es un auto-modelo si su distribucion π es de energia U*

$$U(x) = \sum_S \Phi_i(x_i) + \sum_{\{i,j\}, i \neq j} \beta_{i,j} x_i x_j \quad (5.4)$$

Slamente son aceptables los potenciales de conjuntos unitarios y los potenciales de pares; los potenciales de pares tienen la forma particular $\Phi_{\{i,j\}}(x_i, x_j) = \beta_{i,j} x_i x_j$ (automticamente, $\beta_{i,j} = \beta_{j,i}$). Para cambiar la medida de referencia en x_i , escribamos $\Phi_i(x_i) = \alpha_i x_i + C_i(x_i)$. Las energas condicionales son

$$U_i(x_i | x^i) = x_i(\alpha_i + \sum_{j:j \neq i} \beta_{i,j} x_j) + C_i(x_i) \quad (5.5)$$

Para cada i , $\mathcal{L}(X_i | x^i)$ perteneciente a la familia exponencial de densidades

$$f_i(x_i; x^i) = Z^{-1}(x^i) g(x_i) \exp\{x_i \times \theta_i(x^i)\}, \text{ con } \theta_i(x^i) = \alpha_i + \sum_{j:j \neq i} \beta_{i,j} x_j \quad (5.6)$$

La propiedad ms importante de una familia de este tipo es que si las distribuciones condicionales son de la forma (5.6), ellas definen un modelo conjunto, el auto-modelo (5.4) ([10]; [47], propiedad 2.3.). Veamos algunos ejemplos.

Modelo auto-binomial. Las distribuciones binomiales condicionales, $\mathcal{L}(X_i | x^i) \sim \text{Bin}(m; p_i(x^i))$, $i \in S$, de parmetros

$$p_i(x^i) = [1 + \exp -\{\alpha_i + \sum_{j:j \neq i} \beta_{i,j} x_j\}]^{-1}$$

son las distribuciones condicionales de un auto-modelo de potenciales

$$\Phi_i(x_i) = \alpha_i x_i + \log \binom{m}{x_i} \text{ y } \Phi_{\{i,j\}}(x_i, x_j) = \beta_{i,j} x_i x_j$$

Es suficiente observar que una distribucion binomial $B(m, p)$ admite la representacin exponencial $P(X = x) = \exp\{x \log \frac{p}{1-p} + \log \binom{m}{x} + m \log(1-p)\}$.

Modelo auto-logstico. Este es el modelo precedente con $m = 1$. El modelo tiene energia

$$U(x) = \sum_i \alpha_i x_i + \sum_{i \neq j} \beta_{i,j} x_i x_j$$

Los modelos auto-binomiales permiten la simulacin de una gran variedad de texturas. La diversidad de modelos (nmero de niveles de gris y distribucion marginal, comportamiento de la correlacin, isotropa o no, inhibicin, formas organizadas, ...) se obtienen jugando con los parmetros m (nmero de niveles de gris), ∂_i , la forma y la talla de la vecindad, α y β los valores de los parmetros de las interacciones. La dinmica utilizada para la simulacin es o bien la dinmica de Gibbs, o bien la dinmica de Metropolis de cambio de spins (cf. Cross y Jain [18]; [47], §2.2.4.).

El modelo *auto-poissoniano* corresponde a $F = \mathbb{N}$, teniendo por distribuciones condicionales los sitios i de las distribuciones de Poisson de parmetros

$$\lambda_i(x^i) = \exp\{\alpha_i + \sum_{j:j \neq i} \beta_{i,j} x_j\}$$

Este modelo es admisible si, y solamente si, para todo i, j , $\beta_{ij} \leq 0$. La distribucion conjunta admite como potenciales $\Phi_i(x_i) = \alpha_i x_i + \log(x_i!)$, $\Phi_{\{i,j\}}(x_i, x_j) = \beta_{ij} x_i x_j$.

Los *modelos gaussianos* son los *modelos auto-normales*: efectivamente, la distribucion $\mathcal{N}_S(\mu, \Sigma)$ de energia cuadratica $U(x) = -\frac{1}{2} (x - \mu) Q (x - \mu)$, de potenciales

$$\Phi_i(x_i) = -\frac{1}{2} q_{ii} x_i^2 + \alpha_i x_i \text{ y para } i \neq j, \Phi_{\{i,j\}}(x) = \beta_{ij} x_i x_j$$

donde $Q = (q_{ij}) = \Sigma^{-1}$, $\alpha_i = \sum_{j \in S} q_{ij} \mu_j$, $\beta_{ij} = -\frac{1}{2} q_{ij}$.

Reciprocamente, la familia de distribuciones gaussianas unidimensionales $\mathcal{N}(a_i + \sum_{j:j \neq i} c_{ij} x_j, v_i)$, $i \in S$, se reduce a una gaussiana multidimensional si la matriz Q definida por $q_{ii} = v_i^{-1}$ y $q_{ij} = -\frac{c_{ij}}{v_i}$, es simtrica y definida positiva. Q es claramente la inversa de la matriz de covarianza de la distribucion conjunta, distribucion de media $\mu = (I - C)^{-1} a$ (C es la matriz de las c_{ij} completada por 0 en la diagonal).

Modelo de regiones coloreadas (Strauss, [96])

Este modelo es particularmente simple y muy til en el analisis de imagenes. Consideremos un modelo X con estados cualitativos $F = \{c_1, c_2, \dots, c_K\}$ (K colores, K modalidades, ...) y definido en S finito con una relacion de vecindad $\langle i, j \rangle$ dada. Una configuracion de X es vista por las K regiones coloreadas $S_k = \{i \in S : x_i = c_k\}$, $k = 1, K$. Estas regiones forman una particion de S . Definamos los potenciales

$$\begin{cases} \Phi_i(x_i) = \sum_k \alpha_k \mathbf{1}(x_i = c_k) \\ \Phi_{\{i,j\}}(x_i, x_j) = \sum_{k \neq l} \beta_{k,l} \mathbf{1}[(x_i, x_j) = (c_k, c_l)] \end{cases}$$

La energia asociada a una configuracion x es

$$U(x) = \sum_k \alpha_k n_k + \sum_{k < l} \beta_{k,l} n_{kl} \quad (5.7)$$

donde n_k es el nmero de sitios de color c_k y n_{kl} el nmero de sitios vecinos de colores $\{c_k, c_l\}$. El parametro α_k regula la importancia marginal del color c_k ; mientras mayor sea c_k , ms importante ser la region S_k . $\beta_{k,l}$ controla la verosimilitud de configuraciones vecinas (c_k, c_l) ; si se desea vetar a la vecindad de dos colores c_k y c_l , se escoger β_{kl} grande y negativa; al contrario, las configuraciones vecinas (c_k, c_l) sern bastante probables si $\beta_{kl} > 0$ es grande. Notemos que, como $\sum_k n_k = n$, el modelo est mal parametrizado: se puede fijar, por ejemplo $\alpha_K = 0$. Si los K colores son intercambiables (los α_k son idnticos, los $\beta_{k,l}$ son todos iguales con un mismo coeficiente β), la energia es descrita por el nico parametro β , lo cual vale cuando $n(x) = \sum_{\langle i,j \rangle} \mathbf{1}(x_i = x_j)$

$$U(x) = -\beta n(x) \quad (5.8)$$

Mientras mayor sea β positivo, mayor ser la probabilidad de observar regiones regulares.

Modelo en un espacio de estado mixto $F = \Lambda \times \mathbb{R}^d$

El espacio de estados F puede ser multidimensional. Supongamos por ejemplo que $Z = (X, Y)$ es un valor en $\Lambda \times \mathbb{R}^d$ donde, con la imagen del ejemplo precedente, $\Lambda = \{c_1, c_2, \dots, c_K\}$. Un modo clasico de modelar Z es definiendo jerrquicamente la distribucion: primero se propone un modelo para X ; luego, para cada $x \in \Lambda^S$, se propone una familia de modelos condicionales para $(Y | x)$. Por ejemplo, X es un modelo de K colores, de energia (5.7). Una realizacion x

define, entonces, una particin de S en K subconjuntos y se escoge para $(Y | x)$ un modelo tal que las realizaciones de Y en cada una de las regiones diferentes sean independientes entre ellas, siendo $\{Y_i, i \in S_k\}$ una textura del nivel de gris de \mathcal{T}_k , $k = 1, K$. Para \mathcal{T}_k se puede seleccionar un modelo gaussiano d -dimensional, correlacionado o no al interior de cada regin, o un Φ -modelo que penalice menos a los gradientes pronunciados de Y . Es este el tipo de modelo usado para la segmentacin de texturas por C. Graffigne ([45]).

Dinmica de un campo de Markov

Los modelos de dinmicas espaciales (o modelos espacio-temporales) describen numerosos fenmenos en ecologa, epidemiologa, meteorologa, economa o en geografa espacial. Son tambien modelos tiles para el anlisis de series de imgenes, la deteccin del movimiento y, ms generalmente, la robtica. De esos modelos, presentaremos las Cadenas de Markov-Campos de Markov (MCMF, [9]).

Un proceso MCMF $X = (X(t), t \in \mathbb{N})$ es antes que nada una cadena de Markov en el tiempo, siendo $X(t) = (X_i(t), i \in S)$ el estado en el tiempo t . Luego, para un grafo \mathcal{G} de vecindad instantnea $\langle i, j \rangle_{\mathcal{G}}$ sobre S (tambin denotada $\langle i, j \rangle$), $(X(t) | X(t-1) = x)$ es un *campo de Gibbs, condicional* de x , con potenciales $\Phi^x = (\Phi_A^x)$.

Una clase de modelos simples y manejables es la siguiente: sea x (resp. y) el estado en el tiempo $t-1$ (resp. t): la transicin $P(x, y)$ proviene de dos familias de potenciales

$$P(x, y) = Z^{-1}(x) \exp\left\{\sum_A \Phi_A(y) + \sum_{A,B} \Phi_{B,A}(x, y)\right\}$$

(Φ_A) son los potenciales instantneos asociados al grafo \mathcal{G} . $(\Phi_{B,A})$ son potenciales que definen la interaccin temporal entre los sitios. Estn asociados a \mathcal{G}^- , un grafo orientado $(t-1) \rightarrow t$ el cual determina la influencia de $(i, t-1)$, el sitio i en el instante $t-1$, sobre (j, t) , el sitio j en el instante t . Por ejemplo, una dinmica de estados $\{0, 1\}$ es :

$$P(x, y) = Z^{-1}(x) \exp\left\{\alpha \sum_{\langle i,j \rangle^-} x_i y_j + \beta \sum_{\langle i,j \rangle} y_i y_j\right\}$$

Estos modelos se prestan bien para la simulacin: una vez simulado $X(t) = x$, se simula el campo de Markov $(X(t) | x)$, por ejemplo con el muestreador de Gibbs, sobre la base de las distribuciones

$$\pi_i(y_i | x, y^i) = Z^{-1}(x, y^i) \exp\left\{\sum_{A:i \in A} [\Phi_A(y) + \sum_B \Phi_{B,A}(x, y)]\right\}$$

Estas distribuciones se explican fcilmente. Permiten de igual forma un estudio estadstico simple de esos modelos a travs de la pseudoverosimilitud (PV)

$$\mathcal{P}\mathcal{L}(x(1), x(2), \dots, x(T) | x(0)) = \prod_{t=1, T} \prod_{i \in S} \pi_i(x_i(t) | x(t-1), x^i(t))$$

Cuando la transicin P_θ pertenece a una familia exponencial de parmetro θ , esta pseudoverosimilitud es cncava en θ y los algoritmos determinsticos usuales convergen al nico estimador del mximo de PV.

5.2. Procesos puntuales de Gibbs

En el capítulo presentamos algunos modelos de procesos puntuales. Estos modelos son definidos por su densidad en referencia a un proceso de Poisson. De manera más general, los modelos de Gibbs pueden ser definidos por *objetos* u pertenecientes a una colección F de subconjuntos cerrados de \mathbb{R}^d (cf. Matheron [73], [94]). En general, se supone que F es un espacio métrico separable y localmente compacto. Por ejemplo $u = (x, r) \in \mathbb{R}^2 \times \mathbb{R}^+$ es un disco centrado en x y de radio $r > 0$. Se habla también de *proceso puntual marcado* cada vez que $u = (x, m)$, $x \in \mathbb{R}^p$ tomando en cuenta la ubicación del objeto y m , la *marca*, que contiene las otras informaciones sobre el objeto: talla, orientación, forma, ... Tales modelos de la *geometría estocástica* son tiles en análisis de imágenes cada vez que la observación (degradada o no) pertenece a un espacio continuo. La presentación que nosotros damos aquí sigue el desarrollo del artículo de van Lieshout ([105]).

Una configuración de objetos es un conjunto no ordenado $x = \{x_1, x_2, \dots, x_n\}$ de objetos de F . El espacio de todas las configuraciones es el espacio exponencial $E = \bigoplus_{n \geq 0} F^n$. Sea μ una medida finita y no atómica sobre F . Para el modelo de referencia sobre F , el proceso poissoniano de objeto (PPO), el número de objetos en la ventana de observación sigue una distribución de Poisson de media $\mu(F)$. Además, si n objetos están presentes, ellos están uniformemente distribuidos e independientemente los unos de los otros sobre F : $P(x_i \in B) = \frac{\mu(B)}{\mu(F)}$. En particular, no existe interacción entre los objetos.

Para construir un proceso espacial con interacción, especificaremos la densidad $p : E \rightarrow \mathbb{R}^+$ del nuevo proceso respecto al del PPO. La distribución del número de objetos es

$$q_n = P(N = n) = \frac{e^{-\mu(F)}}{n!} \int_{F^n} p(\{x_1, \dots, x_n\}) d\mu(x_1) \cdots d\mu(x_n)$$

y condicionalmente a $(N = n)$, la distribución espacial es

$$p_n(x_1, \dots, x_n) = \frac{\mu(F)^n e^{-\mu(F)}}{n! q_n} p(\{x_1, \dots, x_n\})$$

Sea \sim una relación simétrica y reflexiva sobre F : por ejemplo dos objetos son vecinos si su intersección es no vacía. Una clase particular de densidades es la de los modelos de Gibbs con interacciones de pares

$$p(x) = \alpha \beta^{n(x)} \prod_{x_i \sim x_j} g(x_i, x_j) = \alpha \beta^{n(x)} \exp \sum_{x_i \sim x_j} \Phi(x_i, x_j) \quad (5.9)$$

α, β son dos constantes > 0 , $n(x)$ es el número de objetos, y $g : F \times F \rightarrow \mathbb{R}^+$ es la función de interacción (Φ es el potencial de interacción), el producto se extiende a los objetos x_i y x_j tales que $i < j$. La función p debe ser integrable; es decir, Φ debe ser admisible.

El caso $g \equiv 1$ corresponde al proceso de Poisson de intensidad $\beta\mu(\cdot)$; $g(x, y) \equiv 0$ si $\|x - y\| < r$ en el proceso con núcleo duro de radio r . El proceso de Strauss (cf. capítulo 2 y [95]) corresponde a $g \equiv \gamma$, γ una constante. $\gamma < 1$ da configuraciones regulares (modelo con *inhibición*), $\gamma > 1$ y $n(x) \leq n_0$ configuraciones con agregaciones (modelo con *atracción*). Este último modelo no es considerado en la literatura porque: (i) no es admisible (sin la restricción $n(x) \leq n_0$) y (ii) conduce a configuraciones muy inestables oscilando entre un P.P.P. y un proceso fuertemente agregado (cf. [77]).

5.2.1. Propiedades de Markov

Si $u \in F$ y $u \notin x = \{x_1, x_2, \dots, x_n\}$ con $p(x) > 0$, y para el modelo (5.9), la relacin

$$\frac{p(x \cup \{u\})}{p(x)} = \beta \prod_{x_i: x_i \sim u} g(x_i, u)$$

depende de u y slo objetos x_i de x vecinos de u . Esta propiedad es comparable a la de los campos de Markov. La relacin $\frac{p(x \cup \{u\})}{p(x)}$ mide la densidad condicional para que un objeto sea ubicado en u siendo conocida la configuracin x alrededor. La definicin precisa de un objeto markoviano es de Ripley y Kelly [87]

Definición 5.2 *Un objeto aleatorio de densidad p es markoviano en la relacin \sim si para todo x de E*

(a) $p(x) > 0$ implica $p(y) > 0$ si $y \subset x$ (propiedad de herencia)

(b) Si $p(x) > 0$, $\frac{p(x \cup \{u\})}{p(x)}$ depende nicamente de u y de $\mathcal{N}(u, x) = \{x_i \in x : u \sim x_i\}$, los vecinos de u en x .

Ripley y Kelly establecen as la equivalencia del teorema de Hammersley-Clifford: un proceso de objetos de densidad $p : E \rightarrow \mathbb{R}^+$ es markoviano si y slo si

$$p(x) = \prod_{y \subset x, y \text{ claques}} q(y) \quad (5.10)$$

El producto se extiende a las claques $y \subset x$, siendo las claques definidas relativamente en la relacin \sim sobre $F \times F$. All tambn, como para los procesos sobre una red discreta, se impone por convencin que el conjunto vaco y todos los conjuntos unitarios sean claques. $q : E \rightarrow \mathbb{R}^+$ es una funcin de interaccin general.

Esos modelos verifican la propiedad de Markov global siguiente: sea A un subconjunto medible de F ; entonces la distribucin de $X \cap A$ condicionalmente a $X \cap A^c$ depende de X en la vecindad de $\mathcal{N}(A) \cap A^c = \{u \in A^c : u \sim a \text{ para un } a \in A\}$

$$\mathcal{L}(X \cap A \mid X \cap A^c) \mathcal{L}(X \cap A \mid X \cap \mathcal{N}(A) \cap A^c)$$

Los procesos markovianos de objetos fueron generalizados por Baddeley y Møller ([7]) al proceso de objetos a los vecinos ms cercanos: para estos modelos, la interaccin (es decir la relacin de vecindad) puede depender de la configuracin. Por ejemplo, dos objetos de x son vecinos, $x_i \sim_x x_j$ si estn en una misma componente conexa de $\cup_{x_i \in x} x_i$. Una versin especfica del teorema de Hammersley-Clifford existe para estos modelos ([7]), la generalizacin de la propiedad de Markov espacial se puede encontrar en ([77], [60]).

5.2.2. Un ejemplo: el proceso de interaccin de reas

Sea (F, d) el espacio mtrico de los objetos. Un proceso de interaccin de rea ([6]) se define por la densidad

$$p(x) = \alpha \beta^{n(x)} \gamma^{-\nu(S(x))}$$

$\beta, \gamma > 0$ son dos constantes, $S(x) = \cup_{i=1, n} B(x_i, r)$ es la reunin de d -bolas de radio r centradas en los puntos de la realizacin x , ν es una medida totalmente finita y regular sobre F . A la inversa del proceso de Strauss, esta densidad es admisible sin limitacin sobre γ . Adems, como $\nu(S(x)) \leq \nu(F) < \infty$ uniformemente en x , la densidad con respecto al PPO de densidad $\beta\mu(\cdot)$

est uniformemente acotada. $\gamma < 1$ corresponde a una distribucion regular de objetos, ms regular a medida que γ sea pequeo (proceso de inhibicin); $\gamma > 1$ hace aparecer los agregados.

Para la relacin $u \sim v \iff B(u, r) \cap B(v, r) \neq \emptyset$ (aun si $d(u, v) < r$), este proceso es markoviano en el sentido de (5.10) y para la funcin de interaccin q

$$\begin{cases} q(\emptyset) = \alpha; & q(a) = \beta\gamma^{-\nu(B(a, r))} \\ \forall k \geq 2, q(\{y_1, \dots, y_k\}) = \gamma^{(-1)^k B_k} \text{ con } B_k = \nu(\cap_{i=1, k} B(y_i, r)) \end{cases}$$

Este proceso tiene potenciales de todos los rdenes. Como veremos, estos modelos de inhibicin ($\gamma < 1$) estn bien adaptados para la reconstruccin de objetos : as como un criterio de AIC para la seleccin de un modelo penaliza una dimensin demasiado importante, un modelo de inhibicin penalizar la reconstruccin de muchos objetos en los casos en los cuales en la imagen original un solo objeto est presente.

5.3. Modelizacin de Gibbs y problemas inversos

La utilizacin de modelos de Gibbs para la resolucin bayesiana de problemas inversos ha conocido un desarrollo importante a partir del artculo fundador de los hermanos Geman ([37], IEEE-PAMI, 1984). Los dominios cubiertos son muy amplios: el anlisis de imgenes, el tratamiento de seales, la reconstruccin tridimensional, la deteccin del movimiento, la robtica y ms generalmente, el reconocimiento de formas.

Los artculos sobre estos temas son numerosos. Es indispensable que el lector motivado por estos problemas consulte esos trabajos: all encontrar elementos importantes sobre los fenmenos mismos, el proceso de formacin de la observacin y otros aspectos ad hoc especficos de la situacin estudiada, expuestos a menudo de manera no matemtica, pero cruciales para una buena comprensin del problema y para su solucin. Nosotros no tocaremos esos aspectos, limitndonos a citar algunas obras de referencia sobre el tema: D.Geman [39], Li [67], Chalmond [16], y, en el contexto de la teora de seales, O'Ruanaidh y Fitzgerald [81]. En estas obras se encontrarn: (i) una bibliografa detallada; (ii) indicaciones acerca del problema delicado de la seleccin de los parmetros que intervienen en los modelos, en particular para la seleccin del parmetro de regularizacin; (iii) el tratamiento efectivo de problemas. Se puede tambin consultar [47] (§2 y §6). Presentaremos algunos problemas inversos, su modelaje y su resolucin bayesiana, poniendo en prctica algoritmos aleatorios.

5.3.1. Formacin de la imagen y reconstruccin bayesiana

Formacin de la imagen

Sea $x = \{x_s, s \in S\}$ una imagen (una seal, una forma, ...) a reconstruir, $x_s \in F$, donde S es un conjunto finito de sitios. La imagen observada y es definida por el modelo de degradacin

$$y = f(H(x), b), y = \{y_t, t \in T\}, y_t \in F, T \text{ finito}$$

H es un filtro, generalmente local (mas no siempre, por ejemplo en la tomografa). b es el ruido. f y H resumen la deformacin determinista. Supondremos que el modelo tiene ruido b , que la deformacin (f, H) es conocida y que $(x, b) \rightarrow (x, y)$ es inversible, es decir que b se explique unvocamente en $(H(x), y)$

$$b = g(H(x), y) \tag{5.11}$$

T no es necesariamente idéntico a S : por ejemplo, en el caso de la tomografía por rayos X, S es el conjunto de píxeles de la imagen x a reconstruir, T es el conjunto de rayos disponibles (orientaciones \times longitud de barrido) y $|T| \ll |S|$ [25]. La situación en la que H es desconocido cubre la deconvolución ciega. No es esencial que x_s y y_t tomen sus valores en el mismo espacio de estados F .

El método de reconstrucción bayesiana

Restaurar x , o filtrar x , es proponer una estimación $\hat{x}(y)$ de x sobre la base de la observación y . Si el cardinal de T es muy inferior al de S , el problema es fuertemente subdeterminado. También diremos que está mal planteado. A fin de enfrentar las dificultades relacionadas con el paso $x \rightarrow y$ (ruido, deformación, subdeterminación), los métodos bayesianos proponen integrar una información a priori sobre el objeto x a reconstruir, información cuyo objetivo es llenar las indeterminaciones encontradas. Esta información a priori debe ser elegida conjuntamente por el experto (quien conoce de manera fina el problema y tiene un modelo a priori sobre x) y por el ingeniero (quien sabe qué es un modelo y lo que deben ser las cualidades numéricas de un algoritmo). La mecánica es la siguiente:

(1) Se escoge una “buena” información a priori sobre x , expresada en términos de $U(x)$. Mientras más importante sea $U(\hat{x})$ más razonable es \hat{x} para el experto y \hat{x} es más probable para la distribución a priori $P(x) = Z_1^{-1} \exp\{U(x)\}$.

(2) Suponiendo el modelo de ruido b independiente de x , y de la energía $V(b)$, la densidad conjunta (x, y) es, según (5.11)

$$P(x, y) = Z_2^{-1} \exp\{U(x) + V(g(H(x), y))\}$$

(3) Se escoge $\hat{x}(y)$ de acuerdo a la distribución a posteriori $(X | y)$

$$P(x | y) = Z^{-1}(y) \exp\{U(x) + V(g(H(x), y))\}$$

Notese que $P(x, y)$ y $P(x | y)$ tiene, salvo una constante en y , la misma energía.

(4) Se pone en práctica por un algoritmo aleatorio el criterio de selección para $\hat{x}(y)$.

El término en V mide la relación a los datos, el término en U la regularidad a priori del objeto x . El método consiste entonces en hacer una selección de \hat{x} , por un lado coherente con la información a priori y por el otro bien ajustado a las observaciones y (V grande), esos dos términos varían en sentidos inversos. $U(x)$ es un factor de penalización sobre x .

MAP, MPM, MPE y ICM

En las fases (3) y (4), la selección debe ser precisada.

El Máximo a posteriori ([37]) El MAP minimiza el costo promedio asociado a la función de costo global $C(x, \hat{x}) = \mathbf{1}(x \neq \hat{x})$,

$$\hat{x}(y) = \text{Arg máx}_x P(x | y) \equiv \text{Arg máx}_{x \in E} \{U(x) + V(g(H(x), y))\}$$

El algoritmo de búsqueda de \hat{x} es un algoritmo de optimización. El recocido simulado es uno de los algoritmos (en este caso, aleatorio) adaptados a la búsqueda del MAP.

El Modo posterior máximo (MPM, [71]): esta selección minimiza el costo promedio para la función aditiva y local $C(x, \hat{x}) = \sum_{s \in S} \mathbf{1}(x_s \neq \hat{x}_s)$,

$$\hat{x}(y) = \{\hat{x}_s = \text{Arg máx}_{x_s \in F} P(x_s | y), s \in S\}$$

$P(x_s | y)$ es la marginal de X en el sitio s , condicionalmente al conocimiento global de y . En general, no se conoce la expresin analtica de la distribucin marginal. El algoritmo se basa entonces en la simulacin de la distribucin condicional ($X | y$). Si los modelos sobre x y sobre b son markovianos, si H es un filtro de soporte acotado, la distribucin a posteriori es markoviana local. El muestreador de Gibbs es entonces un buen algoritmo para realizar estas simulaciones. Las probabilidades son obtenidas por el mtodo de Monte Carlo.

La *Esperanza posterior mxima* (EPM): cuando $x_s \in F \subset \mathbb{R}$, un criterio, anlogo al MPM pero referido a la esperanza, es la seleccin EPM,

$$\hat{x}(y) = \{\hat{x}_s = E(X_s | y), s \in S\}$$

All tambin, el clculo de la esperanza reposar en la simulacin de la distribucin ($X | y$).

El “*Modo condicional iterativo*” (ICM, Besag, [11]) es una variante determinista y muy rpida del recocido simulado para la la reconstruccin MAP. Se basa en la lectura de la identidad siguiente,

$$P(x | y) = P_s(x_s | x^s, y)P(x^s | y)$$

As, relajar el valor x_s en s proponiendo en modo condicional de ($X_s | x^s, y$) slo aumentar la la probabilidad $P(x | y)$. Se deduce el algoritmo:

- (i) Se escoge una sucesin de barrida de sitios, por ejemplo peridica.
- (ii) Si $s = s_k$ es el sitio visitado en el tiempo k , slo el valor en este sitio se cambia a

$$x_s(k+1) = \text{Arg máx}_{x_s} P_s(x_s | x^s(k), y)$$

$P_s(x_s | x^s(k), y)$ se expresa analticamente, el algoritmo es muy rpido. Por el contrario, como en todo algoritmo de optimizacin determinista, es crucial partir de una buena inicializacin. La experiencia muestra que hay que detener al algoritmo luego de un nmero pequeo de iteraciones (aprox. 5 barridas, cf. [11]). En efecto, si $k \rightarrow P(x_s(k+1) | x^s(k), y)$ no para de crecer, se observa en algunos problemas reales que no hay convergencia para la sucesin de imgenes $(x(k))_{k \geq 0}$.

5.4. Algunos ejemplos

5.4.1. Carcter markoviano de la distribucin a posteriori $P(x | y)$

Supongamos que las energas $U(x)$ y $V(b)$ provienen respectivamente de los potenciales $\Phi = \{\Phi_C(x), C \in \mathcal{C}\}$ y $\Psi = \{\Psi_D(b), D \in \mathcal{D}\}$. La energia condicional es entonces

$$U(x | y) = \sum_{\mathcal{C}} \Phi_C(x) + \sum_{\mathcal{D}} \Psi_D(g(H(x), y))$$

Supongamos adems que la funcin de convolucin H sea local ⁽¹⁾

$$H(x) = \{H_s(x(V_s)), s \in S\}$$

donde V_s es una ventana local centrada en s . Luego, la energia condicional es local y vale

¹Esto no ocurre en tomografa: la dependencia en x en el pixel s hace intervenir todos los x_t tales que t y s estn sobre un mismo rayo. Los mtodos markovianos no pueden ser puestos en prtica tan directamente (cf.[25]).

$$U(x | y) = \sum_{\mathcal{C}} \Phi_{\mathcal{C}}(x) + \sum_{\mathcal{D}} \Psi_{\mathcal{D}^*}^*(x, y)$$

donde $\Psi_{\mathcal{D}^*}^*(x, y) = \Psi_{\mathcal{D}}[(g_s(H_s(x(V_s))), y_s), s \in \mathcal{D}]$ y $\mathcal{D}^* = \cup_{s \in \mathcal{D}} V_s$

La nueva familia de claques para $P(x | y)$, que permitir construir las especificaciones en x y los algoritmos de simulacin o de optimizacin, es: $\mathcal{C}^* = \mathcal{C} \cup \mathcal{D}^*$ donde $\mathcal{D}^* = \{\mathcal{D}^*, \mathcal{D} \in \mathcal{D}\}$.

Examinemos el caso particular de un modelo de ruido b independiente. Viendo \mathcal{N}_s la vecindad de s para la distribucin a priori, la vecindad de dependencia de la distribucin condicional $P_s(\cdot | x^s, y)$ es

$$\mathcal{N}_s^* = \mathcal{N}_s \cup V_s^2 \text{ donde } V_s^2 = \{t \in S : \exists u \in S \text{ t.q. } s \text{ y } t \in V_u\}$$

5.4.2. Segmentacin de imgenes

Un bonito ejemplo de segmentacin es aquel estudiado por los hermanos Geman [37]. La segmentacin puede ser mejorada introduciendo, adems del proceso x que representa los niveles de gris, un proceso de bordes δ . Este modelo de bordes refuerza la segmentacin an si δ no es observada inicialmente. Una eleccin sensata de un modelo a priori para (x, δ) da origen a reconstrucciones ms eficientes. Este ejemplo se ha convertido en un clsico de la literatura, como podr constatar el lector ([37]; [39]; [47], §2.4, ejemplo 3).

Describamos un ejemplo importante ms sencillo que el anterior. Sea x una imagen sobre S . Suponemos que el espacio de estados corresponde a K modalidades cualitativas (K colores, K etiquetas en un problema de teledeteccin, ...). La observacin y_s en el sitio s se distribuye

$$\mathcal{L}(y_s | x_s = k) \sim F_k(y_s), y_s \in G, k = 1, K$$

siendo estas distribuciones independientes en cada sitio. El modelo a priori escogido para x es el modelo de regiones de colores de Strauss (5.7), con modelo de energia

$$U(x) = \sum_k \alpha_k n_k + \sum_{k < l} \beta_{k,l} n_{kl}.$$

Ya se ha discutido con anterioridad el efecto que tiene la eleccin de los parmetros α y β sobre la configuracin de la imagen x , en particular sobre el rol de regularizacin de β . Si escribimos $F_k(y_s) = \exp\{u_k(y_s)\}$, la energia a posteriori est dada por

$$U(x | y) = U(x) + \sum_{s \in S} u_{x_s}(y_s) \quad (5.12)$$

Para un modelo en el cual los K estados son intercambiables, un slo parmetro β caracteriza a $U(x)$. En este caso la energia condicional en s puede escribirse como,

$$U_s(x_s | x^s, y) = \beta n_s(x) + u_{x_s}(y_s) \quad (5.13)$$

Dos ejemplos clsicos de ruido son los siguientes:

1. *Ruido de transmisin de canal*: $F = G$, y $P(y_s \neq x_s) = \varepsilon$, uniformemente sobre los $(K - 1)$ valores diferentes a x_s .
2. *Respuesta y_s multidimensional*: esta es la situacin ms comn en teledeteccin, donde el satlite enva, para cada pixel, las respuestas multiespectrales de p -receptores, $F = \mathbb{R}^p$. En el caso de una respuesta Gaussiana, $F_k \sim N_p(\mu_k, \Sigma_k)$.

Les energas (5.12) y (5.13) se escriben fcilmente en este caso.

5.4.3. Detección de rupturas en una señal

La versión para señales de un problema de segmentación es el problema de detección de rupturas en una señal, donde no necesariamente conocemos el número de clases ([81],[66]).

Describamos el marco de este problema. Un proceso de régimen no observable $x_t \in F$, ($t \in S = \{1, 2, \dots, n\}$), gobierna una señal observable $y = \{y_s, s \in S\}$, $y_s \in \mathbb{R}^p$. El problema consiste en detectar los momentos de cambio de régimen o momentos de ruptura.

Si $r = \{r_t, t \in S\}$, es el proceso, con valores en $\{0, 1\}$, de los instantes de ruptura ($r_t = 1$ si hay una ruptura en t), escogeremos como modelo a priori para r el modelo Bernoulli, siendo los $r_i \text{ Ber}(p)$ i.i.d.,

$$P(x) = C \exp\{-\alpha n_r\} \text{ con } \alpha = \log \frac{1-p}{p}$$

donde n_r es el número de rupturas. Siendo que $\alpha \in \mathbb{R}$ regula la densidad de rupturas, un valor positivo grande penaliza la ocurrencia de un gran número de rupturas.

Denotemos $S = I_1 \cup I_2 \cup \dots \cup I_{n_r+1}$ a la partición asociada a cada zona de régimen homogéneo. Si la verosimilitud de un modelo sobre el l -ésimo intervalo se estima usando $q(y(I_l), \hat{\theta}_l)$, con energía u , la distribución a posteriori está dada por

$$P(r | y) = C \exp\{-\alpha n_r + \sum_{l=1}^{n_r+1} u(y(I_l), \hat{\theta}_l)\} \quad (5.14)$$

Examinemos, por ejemplo el caso en el cual el cambio de régimen afecta únicamente a la media,

$$y_t = \mu_l + \varepsilon_t \text{ si } t \in I_l, \varepsilon \sim BB(0, \sigma^2)$$

Sobre el intervalo I_l , esta media se estima con \bar{y}_l , y la variancia residual con $\widehat{\sigma}^2(r) = \frac{1}{n} \sum_{l=1, n_r+1} \sum_{t \in I_l} (y_t - \bar{y}_l)^2$.

En este caso la energía a posteriori está dada por

$$U(r | y) = \frac{n}{2} \log \widehat{\sigma}^2(r) - \alpha n_r$$

5.4.4. Reconocimiento de objetos

Consideremos un ejemplo de reconocimiento de un objeto modelado sobre el espacio continuo \mathbb{R}^2 . Seguiremos nuevamente la exposición de [105]. La imagen inicial es una foto "granulada" ("pellets") extraída del álbum de texturas de Brodatz [13]. Esta foto está escaneada en $S = 128 \times 128$ píxeles (hemos pasado a un espacio de sitios discretos). Los granos se tratan como discos con radio fijo de 4 píxeles. Supondremos que la degradación es aditiva Gaussiana

$$y_s = H_s(x) + \varepsilon_s, s \in S$$

donde el soporte de la función de ruido H_s está contenido en la ventana 3×3 centrada en s , con pesos relativos 4 en el centro, 2 en los 4 v.m.c., y 1 en los 4 v.m.c. diagonales. La variancia del ruido se estima por muestreo de la distribución de los niveles de gris de y .

Una reconstrucción de x por máxima verosimilitud conduce a la siguiente observación: los granos se identifican correctamente pero el método puede hacer aparecer agregados alrededor de los verdaderos objetos. Esto se explica fácilmente: al no haber un término de penalización, la verosimilitud aumenta con la dimensión del modelo y agregar granos en ciertas regiones puede mejorar la verosimilitud sin degradación (los granos se recubren). Una manera de hacer desaparecer este efecto es imponer un modelo de inhibición, por ejemplo el modelo de inhibición de Strauss o el

modelo de interaccin de reas de Baddeley-van Lieshout, como distribucin a priori para x . Estos modelos penalizan la agregacin injustificada de granos. Verificamos as que la reconstruccin por MAP no presenta estos defectos de las respuestas mltiples.

En [105], el autor utiliza, tanto para la simulacin, como para la optimizacin, una dinmica de procesos de nacimiento y muerte (PNM, [76]). Esta dinmica se define sobre la imagen discretizada. Para la optimizacin por recocido simulado, van Lieshout establece una condicin suficiente sobre el esquema de temperaturas para asegurar la convergencia [104]. Son preferibles las dinmicas de Metropolis a las de PNM [78].

Si se quiere trabajar directamente sobre el objeto continuo $x \subset \mathbb{R}^2$, los resultados de [98] permiten controlar la convergencia de las simulaciones. El problema de convergencia del recocido simulado en este espacio sigue abierto. Sin embargo, la experiencia muestra que una buena aplicacin del recocido simulado sobre un espacio de estados $E \subset \mathbb{R}^p$ continua “funcionando bien”.

5.5. Ejercicios

Ejercicio 5.1 Restricciones de identificabilidad (5.3) para un potencial

Si F tiene K estados y S tiene n sitios, un campo sobre $E = F^S$ depende de $(K^n - 1)$ parmetros. Mostrar que esta dimensin se consigue usando la representacin en potencial bajo las restricciones (5.3).

Ejercicio 5.2 No se pueden “pegar” distribuciones condicionales arbitrarias

(1) $A \times B$ tiene $m \times n$ elementos. Cul es la dimensin paramtrica conjunta de las dos familias de distribuciones condicionales sin restricciones $\{f(x | b), b \in B\}$ sobre A y $\{g(y | a), a \in A\}$ sobre B ? Cul es la dimensin de una distribucin conjunta sobre $A \times B$? Conclusin.

(2) Sobre \mathbb{Z} , y para un espacio con K estados, se considera un ncleo homogneo $\pi(y | x, z) = P(X_0 = y | X_{-1} = x, X_1 = z)$. Cul es la dimensin de este ncleo? Evale el nmero de restricciones que deben imponerse para que las distribuciones correspondan a las de un proceso markoviano con dos v.m.c.

(3) Considere la familia de distribuciones binomiales, homogneas, $\pi(X_0 = x | X_{-1} = x_{-1}, X_1 = x_1) = B(m, \theta(\cdot))$ donde $\theta(\cdot) = \frac{\exp A(\cdot)}{1 + \exp A(\cdot)}$ con $A(\cdot) = \alpha + \beta(x_{-1} + x_{+1})^2$. Estas distribuciones condicionales se “pegan” bien?

Ejercicio 5.3 Procesos Markovianos sobre un arreglo triangular

Sobre un arreglo triangular plano, considrese el campo de Markov binario invariante por traslacin con 6 v.m.c.. Describa las claques, los potenciales, las distribuciones condicionales. Identificar los submodelos: isotrpico; reversible ($\forall A, \Phi_A(x_A)$ es invariante por permutacin de los ndices de x en A); el modelo sin potenciales de orden > 2 .

Ejercicio 5.4 Modelo de Markov con ruido

Sea X el campo de Markov binario con 2-v.m.c. sobre $S = \{1, 2, \dots, n\}$. La transmisin $X \rightarrow Y$ se lleva a cabo con un ruido de canal, independiente sitio a sitio, con distribucin: $P(X_i = Y_i) = 1 - \varepsilon = 1 - P(Y_i \neq X_i), 0 < \varepsilon < 1$. Escribir la distribucin de (X, Y) . Es Y un campo de Markov? Es $(X | Y)$ un campo de Markov? Escriba las distribuciones condicionales.

Ejercicio 5.5 *Modelo bivariado sobre $\{0, 1\}^2$*

Considere el modelo bivariado sobre $S = \{1, 2, \dots, n\}$ de energía

$$U(x, y) = \alpha \sum_i x_i + \beta \sum_{\|i-j\|=1} x_i x_j + \gamma \sum_i y_i + \delta \sum_i x_i y_i$$

- (1) Detalle un algoritmo de simulación de (X, Y) .
- (2) Solo y es observado. Describa los algoritmos MAP y MPM de reconstrucción de x .
- (3) Se observan n gaussianas Z , independientes, t.q. $\mathcal{L}(z_i | x, y) = \mathcal{L}(z_i | y_i) = \mathcal{N}(\mu(y_i), \sigma^2)$. Determinar la distribución conjunta de (X, Y, Z) y verificar que $\pi(y, z | x) = \prod_i \pi(y_i, z_i | x)$. Calcular $\pi(y_i = 0, z_i | x)$ y $\pi(y_i = 1, z_i | x)$. Obtener $\pi(z_i | x)$, $\pi(z | x)$ y $\pi(x, z)$. Observamos z . Dar el algoritmo MPM de reconstrucción de x .

Ejercicio 5.6 *Cadena con ruido observada cada dos sitios*

Considere el campo binario sobre $S = \{0, 1, 2, 3, \dots, 2n\}$ de energía

$$U(x) = \alpha \sum_i x_i + \beta \sum_{\|i-j\|=1} x_i x_j + \gamma \sum_{\|i-j\|=2} x_i x_j$$

- (1) S está particionada en $P \cup I$, el conjunto de sus índices pares y de sus índices impares. Determinar la distribución $\pi_I(x_I | x_P)$. Verificar que $(X_I | x_P)$ es una cadena de Markov, independiente si $\gamma = 0$.
- (2) Suponga que $\gamma = 0$. Se observa Y_i , solo para $i \in P$, de distribución: $P(Y_i = x_i) = 1 - P(Y_i \neq x_i) = \varepsilon$, siendo estas variables independientes sitio a sitio. Se observa y_P . Determinar las distribuciones de $\pi_i(x_i | x^i, y_P)$ para $i \in S$ (examinar separadamente los casos $i \in I$ y $i \in P$). Proponer un algoritmo de reconstrucción de $(x | y_P)$.

Ejercicio 5.7 *Degradación exponencial*

Se quiere reconstruir un objeto $x \in \{0, 1\}^S$, $S = \{1, 2, \dots, n\}$. Escogemos el siguiente modelo para x :

$$\pi(x) = Z^{-1} \exp\left\{\beta \sum_{\langle i, j \rangle} x_i x_j\right\}, \beta > 0$$

La observación disponible $y = (y_i)_{i \in S}$ ($y \in (\mathbb{R}^+)^S$), degradación de x , sigue el siguiente modelo: (i) Para todo i , $\mathcal{L}(y_i | x) = \mathcal{L}(y_i | x_i) = \text{Exp}(\lambda(x_i))$ (distribuciones exponenciales, $\lambda(0)$ y $\lambda(1)$ conocidos); (ii) Estas distribuciones son independientes entre ellas. Determinar la distribución conjunta de (x, y) , su energía así como la distribución a posteriori $(x | y)$. Identificar las distribuciones condicionales $(x_i | x^i, y)$. Describir el algoritmo de reconstrucción MPM de x .

Ejercicio 5.8 *Distribución condicional y distribución marginal de un campo de Gibbs*

Sea $X = (X_i, i \in S)$ un campo de Markov binario sobre $\Omega = \{0, 1\}^S$, $S = \{1, 2, \dots, n\}$, de distribución π :

$$\pi(x) = Z^{-1} \exp U(x), \text{ donde } U(x) = \alpha \sum_S x_i + \beta \sum_{\langle i, j \rangle} x_i x_j$$

- (1) Sea k un sitio de S . Determinar la energía y los potenciales de la distribución de $(X^k | x_k)$.
- (2) Escribiendo $\pi^k(x^k) = \pi(0_k, x^k) + \pi(1_k, x^k)$, identifique la energía de la distribución marginal π^k de X sobre $A = S \setminus \{k\}$. Constatar la aparición de un nuevo potencial $\Phi_{\partial k}$. Describir el nuevo grafo de vecindades de π^k (analizar separadamente los casos $|\partial k| \leq 1$ y $|\partial k| > 1$).

Ejercicio 5.9 *Segmentacin y modelo (5.12)*

Considere la energia de segmentacin definida en (5.12) as como los dos modelos de ruido descritos en esa seccin (ruido de canal, ruido gaussiano).

1. Determinar la distribucin a posteriori $\pi(x | y)$ y las distribuciones condicionales $\pi_i(x_i | x^i, y)$. En caso de un modelo de ruido gaussiano, examinar las dos situaciones: (i) las poblaciones se distinguen nicamente por sus medias (μ_k); (ii) las poblaciones se distinguen nicamente por sus varianzas. Describir los procedimientos de reconstruccin MAP y MPM.

2. Aplicacin : sea $S = \{1, 2, \dots, 64\}^2$. Escoger una particin de tres colores y aadir ruido; reconstruir la particin inicial.

Ejercicio 5.10 *Deconvolucin*

Sobre el toro de una dimensin, $S = \{1, 2, \dots, 64\}$, considrese la seal Gaussiana autoregresiva de energia $U(x) = -\frac{1}{2\sigma^2}(\sum_i x_i^2 + 2c \sum_i x_i x_{i+1})$, $|c| < \frac{1}{2}$. La seal observada es una convolucin perturbada por un ruido blanco Gaussiano $N(0, \sigma_b^2)$,

$$Y_i = \sum_{k=-p}^p \theta_k X_{i-k} + \varepsilon_i$$

El filtro y la varianza del ruido son conocidos. Describir la distribucin a posteriori $(X | y)$, precisando sus potenciales y sus distribuciones condicionales $\pi_i(x_i | x^i, y)$.

Ejercicio 5.11 *Deteccin de rupturas*

Consideremos el modelo (5.14) que describe la distribucin a posteriori en un problema de deteccin de rupturas de medias. Para la reconstruccin por MAP, se utiliza una dinmica de Metropolis con un ncleo de proposicin $q(r, r')$ sobre $\{0, 1\}^n$: (i) se escoge un punto i de $\{1, 2, \dots, n\}$ al azar; (ii) $r'_i = 1 - r_i$, y $r'_j = r_j$ en los otros puntos. Es q irreducible? simtrico? Calcular la variacin de energia para $r \rightarrow r'$. Describir el algoritmo de optimizacin asociado.

Se desea mejorar el ncleo q en q' : (i) para un r t.q. $n_r = 0$ o n , $q'(r, r') = q(r, r')$; (ii) si no, se lanza una Bernoulli de parametro p . Si vale 1, tomamos $q'(r, r') = q(r, r')$; si no, escogemos al azar una de las n_r rupturas de r , y la trasladamos a la derecha o a la izquierda con la misma probabilidad si las dos plazas estn libres. En caso contrario no hacemos nada. Describir el ncleo. Mostrar que es irreducible y simtrico. Escribir un algoritmo de optimizacin.

Ejercicio 5.12 *Tomografa Gamma en imaginera mdica ([12])*

Para establecer un diagnostico sobre el estado de un rgano x , se hace ingerir una sustancia radioactiva al paciente. Este producto se concentra en el rgano y emite fotones que son captados y amplificados. Se obtiene as una imagen Gamma y de x . Las concentraciones en el rgano son $x = (x_s, s \in S)$, $x_s \in F = \{c_1, c_2, \dots, c_K\}$, K valores ordenados y > 0 . Las emisiones de fotones $y = \{y_t, t \in T\}$ son observadas sobre una red T ; ellas siguen un modelo independiente en cada sitio t , de distribuciones

$$\mathcal{L}(y_t | x) = P(\lambda_t(x)) \text{ con } \lambda_t(x) = \sum_{s \in S} h_{t,s} x_s$$

El ncleo $(h_{t,s})_{t \in T, s \in S}$ es una caracterstica conocida del instrumento. Se propone la utilizacin del siguiente modelo a priori sobre x

$$\pi(x) = \frac{C}{\prod_s x_s} \exp\{\gamma \sum_{\langle r,s \rangle} |\log x_r - \log x_s|\}$$

- (1) Determinar las distribuciones condicionales a posteriori $\pi_s(x_s | x^s, y)$.
- (2) Se quiere estimar MAP utilizando un recocido simulado con la dinámica de Metropolis para el siguiente núcleo de proposición: (i) se escoge un sitio i de S al azar; (ii) en ese sitio i , escribiendo $c_k = x_s$, x_s se reemplaza por x'_s escogido al azar sobre $\{c_{k-1}, c_k, c_{k+1}\}$ (con las correcciones naturales en los extremos de F), dejando iguales los demás valores. Verificar que q es simétrico e irreducible. Calcular la variación de energía para este esquema de actualización. Escribir un algoritmo de optimización.

Ejercicio 5.13 *Descompresión de imagen*

X es una imagen con valores sobre $S = \{1, 2, \dots, n\}^2$, $n = 3m$. Se transmite la información $Y = \{y_c, c \in C\}$ donde $y_c = \sum_{V_c} x_i$, V_c la ventana 3×3 alrededor del punto $c = (2 + 3i, 2 + 3j)$, $i, j = 0, m - 1$ de S . Proponer un modelado a priori de x que permita su reconstrucción a partir de y si: (1) x es un modelo de valores de gris con variación regular; (2) x es un modelo de 3 niveles de gris, con cambios bruscos en la frontera.

Bibliografía

- [1] Aarts E. et Korst J., 1989, *Simulated annealing and Boltzman machines : Stochastic approach to combinatorial optimization and neural computing*, Wiley
- [2] Aldous D. et Diaconis P., 1986, Shuffling cards and stopping times, *Amer. Math. Monthly*, 93, 333-348
- [3] Aldous D. et Fill J. 1993-1996, *Reversible Markov chains and random walks on graphs*, Monograph in preparation, www.stat.berkeley.edu/~aldous/book.html
- [4] Amit Y. et Grenander U., 1991, Comparing sweep strategies for stochastic relaxation, *J. Mult. Anal.* 37, 197-222
- [5] Azencott R., Ed., 1992, *Simulated annealing : Parallelization techniques*, Wiley
- [6] Baddeley A. et van Lieshout M.N.M., 1995, Area-interaction point processes, *Ann. Inst. Stats. Math.* 47, 601-619
- [7] Baddeley A.J. et Moller J., 1989, Nearest-neighbour Markov point processes, *Ann. Inst. Stat. Math.*, 57, 89-121
- [8] Barker A.A., 1965, Monte Carlo calculations of the radial distribution functions for a proton-electron plasma, *Austr. J. Phys.* 18, 119-133
- [9] Bayomog S., Guyon X. et Hardouin C., 1998, Markov Chain Markov Field Dynamic, Preprint SAMOS, n°100, 25p.
- [10] Besag J., 1974, Spatial interaction and the statistical analysis of lattice systems, *JRSS B-36*, 192-236
- [11] Besag J., 1986, On the statistical analysis of dirty pictures, *JRSS B*, 43, n°3, 302 – 318
- [12] Besag J. et Green P.J., 1993, Spatial statistics and Bayesian computation, *JRSS B* 55, 25-37
- [13] Brodatz P., 1966, *Texture : a photographic album for artists ans designers*, Dover, New York
- [14] Carter D.S. et Prenter P.M., 1972, Exponential states and counting processes, *Z. Wahrsch. Verw. Geibete* 21, 1-19
- [15] Catoni O., 1992, Rough large deviation estimates for simulated annealing : application to exponential schedules, *Ann. Proba.*, Vol 20, n°3, 1109-1146
- [16] Chalmond B., *Elments de modlisation pour l'analyse d'image*, Springer, à paraître

- [17] Chauveau D., Dibolt J. et Robert C.P., 1998, Control by the Central Limit Theorem, 99-126, in *Discretisation and MCMC convergence assesement*, Ed. Robert C., L.N.S. n°135, Springer
- [18] Cross G.R. et Jain A.K., 1983, Markov random field texture models, IEEE Trans. PAMI, 5, n°1, 25 – 39
- [19] Devroye L., 1986, *Non-Uniform random variable generation*, Springer
- [20] Diaconis P. et Sahshahani M, 1981, Generating a random permutation with random transpositions, Z. Wahrsch. Verw. Geb., 57, 159-179
- [21] Diaconis P. et Shahshahani M., 1987, Time to reach stationarity in the Bernoulli-Laplace diffusion model, SIAM J. Math. Anal., 18, 208-218
- [22] Diaconis P., Graham R. et Morrison J., 1990, Asymptotic analysis of a random walk on an hypercube with many dimensions, Rand. Struct. Algorithms, 1, 51-72
- [23] Diaconis P. et Saloff-Coste L., 1995, What do we know about the Metropolis algorithm ?, in Proc. of the Twenty-Seventh Annual ACM Symp. on the Theo. Comp., 112-129
- [24] Diaconis P., 1996, The cutoff phenomenon in finite Markov chains, Proc. Natl. Acad. Sci. USA, 93: 1659-1664
- [25] Dinten J.M., 1990, *Tomographie à partir d'un nombre limité de projections: rgularisation par champs markoviens*, Thèse Univ. Paris Sud
- [26] Dobrushin L.R., 1956, Central Limit Theorem for non-stationary Markov chain, I et II, Theo. Proba. Appl. 1, 65-80 et 329-383
- [27] Doeblin W., 1938, Expos de la thorie des chaînes simples constantes de Markov à nombre fini d'tats, Rev. Math. Union Interbalkan. 2, 77-105.
- [28] Doukhan P., 1995, *Mixing : Properties and Examples*, L.N.S. n°85, Springer
- [29] Duflo M., 1997, *Algorithmes stochastiques*, Springer
- [30] Feller W., 1968, *An introduction to probability theory and its applications*, Vol.1, 3-ème Ed., Wiley
- [31] Fill J.A., 1998, An interruptible algorithm for perfect simulation via Markov chains, Ann. Appl. Proba., 8, 131-162
- [32] Foss S.G. et Tweedie R.L., 1998, Perfect simulation and backward coupling, Stochastic models, 14, 187-203
- [33] Frigessi A., Hwang C.R., Di Stephano P. et Sheu S.J., 1993, Convergence rates of Gibbs sampler, the Metropolis algorithm and other single-site updating dynamics, J.R.S.S. B 55, 205-219.
- [34] Gaetan C., 1994, A stochastic algorithm for maximun likelihood estimation of Gibbs point processes, Preprint SAMOS 36, Univ. Paris 1
- [35] Gantert N, 1989, Laws of large numbers for the Annealing Algorithms, Stoch. Proc. Appl. 35, 309-313

- [36] Gantmacher F.R., 1959, *Applications of the theory of Matrices*, Wiley
- [37] Geman D. et Geman S., 1984, Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, IEEE-PAMI-6, 721-741
- [38] Geman D. et Geman S., 1987, Relaxation and annealing with constraints, Tech. report N°35, Div. Appl. Maths., Brown University
- [39] Geman D., 1990, *Random fields and inverse problem in imaging*, L.N.M. n°1427, Springer
- [40] Geman S. et Graffigne C., 1987, Markov random field models and their application in computer vision, in Gleason M. (Ed.), Proc. Int. Congress of Maths. (1986), A. M. S., Providence, 1496-1517
- [41] Georgii H.O., 1988, *Gibbs measure and phase transitions*, De Gruyter
- [42] Gidas B., 1991, *Metropolis-type Monte Carlo simulation algorithms and simulated annealing*; Trends in Contemporary Probability, 88 p., A.M.S., Providence
- [43] Gilks W.R., Richardson S. et Spiegelhalter D.J., 1996, *Markov Chain Monte-Carlo in Practice*, Chapman et Hall
- [44] Godement R., 1962, *Cours d'algèbre*, Hermann
- [45] Graffigne C., 1987, *Experiments in texture analysis and segmentation*, PhD thesis, Brown Univ.
- [46] Griffeath D., 1975, A maximal coupling for Markov chains, Z. fur W., 31, 95-106
- [47] Guyon X., 1995, *Random Fields on a Network : Modeling, Statistics and Applications*, Springer
- [48] Häggström O., van Lieshout M.N.M. et Møller J., 1999, Characterisation results and Markov chain Monte Carlo algorithms including exact simulation for some point processes, à paraître in Bernoulli
- [49] Häggström O. et Nelander K., 1998, On exact simulation of Markov random fields using coupling from the past, preprint Göteborg Univ.
- [50] Hajek B., 1988, Cooling schedules for optimal annealing, Math. of Operations Research 13, n°2, 311-329
- [51] Hastings W.K., 1970, Monte Carlo sampling methods using Markov chains and their applications, Biometrika 57, 97-109
- [52] Iosifescu M. et Theodorescu R., 1969, *Random processes and learning*, Springer
- [53] Isaacson D.L. et Madsen R.W., 1985, *Markov chains; Theory and applications*, 2-ème Ed., Krieger.
- [54] Jonhson et Kotz (1969-1972), *Distribution in statistics*, (4 volumes), Wiley
- [55] Johnson V.E., 1996, Studying convergence of Markov chain Monte Carlo algorithms using coupled sample paths, JASA, 91, 154-166

- [56] Johnson V. et Reutter A., 1995, General strategies for assesing convergence of MCMC algorithms using coupled sample paths, preprint
- [57] Karlin S., 1968, *A first course in stochastic processes*, Acad. Press.
- [58] Keilson J., 1979, *Markov chain models-Rarity and Exponentiality*; Springer
- [59] Kemeny G. et Snell J.L., 1960, *Finite Markov Chains*, D. Van Nostrand.
- [60] Kendall W.S., 1990, A spatial Markov property for nearest-neighbour Markov point processes, *J. Appl. Proba.*, 28, 767-778
- [61] Kendall W.S, 1997, Perfect simulation for the area-interaction point process, in *Probability towards 2000*, Eds Arcadi L. et Heyde C., 218-234, New-York: Springer
- [62] Kendall W. et Møller J., 1999, Perfect Metropolis-Hastings simulation of locally stable point processes, preprint
- [63] Kelly F.P. et Ripley B.D., 1976, A note on Strauss's model for clustering, *Biometrika* 63, 357-360
- [64] Kinderman et Snell J.L., 1980, *Markov random fields and their applications*, Contemporary Maths., Vol 1, A.M.S., Providence.
- [65] Kirkpatrick S., Gelatt C.D. et Vecchi M.P., 1983, Optimisation by simulated annealing, *Science*, 220, 671-680
- [66] Lavielle M., 1998, Optimal segmentation of a random process, *IEEE Trans. Signal Proc.*, Vol 46, n°5, 1365 – 1373
- [67] Li S., 1995, *Markov Random Field modeling in computer vision*, Springer
- [68] Lindvall T., 1992, *Lectures on coupling method*, Wiley
- [69] Luenberger D., 1979, *Introduction to dynamical systems : Theory and Applications*, Wiley.
- [70] Liu J.S., 1995, Metropolized Gibbs sampler: an improvement; Preprint, Univ. of Stanford
- [71] Marroquin J., Mitter S. et Poggio T., 1987, Probabilistic solution of ill-posed problems in computational vision, *JASA*, 82, 76-89
- [72] Martinez S. et Ycart B., 1998, Decaying rates and cutoff for convergence and hitting times of Markov chains, soumis
- [73] Matheron G., 1975, *Random sets and integral geometry*, Wiley
- [74] Metropolis N., Rosenbluth A., Teller A. et Teller E., 1953, Equations of state calculations by fast computing machines, *J. Chem. Phys.* 21, 1087-1092
- [75] Meyn S.P. et Tweedie R.L., 1993, *Markov chains and stochastic stability*, Springer.
- [76] Møller J., 1989, On the rate of convergence of spatial birth and death processes, *Ann. Inst. Stat. Math.*, 41, 565-581
- [77] Møller J., 1994, Markov chain Monte Carlo and spatial point processes, Tech. Report 293, Dept. Th. Stat., Univ. of Aarhus

- [78] Møller J., 1998, Markov chain Monte Carlo and spatial point processes, in Proc. Sm. Eur. de Stat., *Current trend in stochastic geometry with applications*, Eds. Barndorff-Nielsen et Kendall W.S., Chapman et Hall.
- [79] Møller J., 1999, Perfect simulation of conditionally specified models, J.R.S.S. B, 61, Part 1, 251-264
- [80] Murdoch D.J. et Green P.J., 1998, Exact sampling from continuous state space, Scan. J. Stat., 25, 483-502
- [81] O'Ruanaidh J. et Fitzgerald W., 1996, *Numerical Bayesian methods applied to Signal Processing*, Springer
- [82] Peskun P., 1973, Optimun Monte-Carlo sampling using Markov chains; Biometrika 60, 607-612
- [83] Pitman J.W., 1976, On coupling of Markov chains, Z. fur W., 35, 315-322
- [84] Propp J.G. et Wilson D.B., 1996, Exact sampling with coupled Markov chains and applications to statistical mechanics, Random Structures and Algorithms, 9(1), 223-251
- [85] Prum B., 1986, *Processus sur un rseau et mesure de Gibbs*, Masson (existe en version anglaise)
- [86] Raftery A.E. et Lewis S.M., 1996, Implementing MCMC, in *Markov Chain Monte-Carlo in Practice*, Gilks, Richardson et Spiegelhalter Eds., Chapman et Hall
- [87] Ripley B.D. et Kelly F.P., 1977, Markov point processes, J. London Math. Soc. 15, 188-192
- [88] Robert Ch., 1996, *Mthodes de simulation en statistique*, Economica (version anglaise chez Springer)
- [89] Robert C.P. et Cellier D., 1998, Convergence control of MCMC algorithms, 27-46, in *Discretization and MCMC convergence assesement*, Ed. Robert C., L.N.S. n°135, Springer
- [90] Robert C.P. (Editeur), 1998, *Discretization and MCMC convergence assesement*, L.N.S. n°135, Springer
- [91] Saloff-Coste L., 1997, *Lectures on finite Markov chains*, Ecole d't de Probabilit de St. Flour XXVI, L.N.M. n° 1665, Springer.
- [92] Seneta E., 1975, *Non-negative matrices*, Halsted Press.
- [93] Sinclair A., 1993, *Algorithms for random generation and counting : a Markov chain approach*, Birkhäuser
- [94] Stoyan D., Kendall W.S. et Mecke J, 1995, *Stochastic Geometry and its Applications*, 2nd Ed., Wiley
- [95] Strauss D.J., 1975, A model for clustering, Biometrika 62, 467-475
- [96] Strauss D.J., 1977, Clustering on coloured lattice, J. Appl. Proba., 14, 135-143
- [97] Thorisson T.H.T, 1995, Coupling methods in probability theory, Scand. Jour. of Stat., Vol 22, 159-182.

- [98] Tierney L., 1994, Markov chains for exploring posterior distributions, *Ann. of Stat.* 22, 1701-1762
- [99] Tierney L., 1996, Introduction to general state-space Markov chain theory, p.59-74, in *Markov Chain Monte Carlo in practice*, Eds. Gilks, Richardson et Spiegelhalter, Chapman & Hall.
- [100] Tierney L., 1995, A note on Metropolis-Hastings kernels for general state space, Preprint 606, School of Statistics, Univ. of Minnesota.
- [101] Trouv A., 1993, Parallisation massive du recuit simul, Thèse de Doctorat, Univ. d'Orsay.
- [102] Tweedie R.L., *Markov Chains : Structure and Applications*, à paraître, Handbook of Statistics
- [103] van Laarhoven P.J. et Aarts E., 1987; *Simulated annealing : Theory and Applications*, Reidel
- [104] van Lieshout M.N.M., 1994, Stochastic annealing for nearest-neighbour point processes with application to object recognition, *Adv. Appl. Prob.*, 26, 281-300
- [105] van Lieshout M.N.M., 1995, Markov point processes and their application in high-level imaging, *Bull. ISI*, 559-576
- [106] Winkler G., 1995; *Image analysis, Randoms Fields and Dynamic Monte Carlo Methods: a mathematical introduction*, Springer
- [107] Yao J.F., 1999, On constrained simulation and optimisation by Metropolis chains, à paraître in *Statistics and Probability letters* (disponible à <http://www.univ-paris1.fr/SAMOS>)
- [108] Ycart B., 1998, Cutoff for samples of Markov chains, *ESAIM Probability and Statistics*, à paraître
- [109] Ycart B., 1999, Stopping tests for Monte-Carlo Markov chains methods, soumis
- [110] Ycart B., 1998, *Cutoff for Markov chains : some example and applications*, Summer school on complex systems, Santiago du Chilú, Dec. 1998
- [111] Younes L., 1988, *Problèmes d'estimation paramétrique pour les champs de Gibbs Markoviens. Applications au traitement d'Images*, Thèse de Doctorat, Univ. d'Orsay.