

# Método de Monte Carlo via Cadeias de Markov com saltos reversíveis

Isabel Pereira, *isabel.pereira@ua.pt*

*CIDMA-Centro de Investigação e Desenvolvimento em Matemática e Aplicações  
Departamento de Matemática da Universidade de Aveiro*

## 1 Introdução

Os métodos de Monte Carlo via cadeias de Markov, referidos simplesmente como MCMC, constituem um dos mais notáveis avanços na estatística nos últimos 20 anos. As suas origens remontam aos trabalhos de Metropolis *et al* (1953) e Hastings (1970). No entanto, foi apenas nos finais dos anos oitenta que esta ideia começou a ser alvo de investigação. Inicialmente os métodos MCMC foram usados na estatística bayesiana, no entanto a sua utilização é mais alargada podendo ser usada na dita estatística clássica.

De uma forma geral aplicam-se os algoritmos MCMC para obter amostras da distribuição *a posteriori* dos parâmetros do modelo. Qual o procedimento a usar para comparar diferentes modelos considerando um determinado conjunto de dados? Na estatística clássica os modelos podem ser comparados usando critérios de informação, em particular o AIC (*Akaike's information criterion*) ou o BIC (*Bayesian information criterion*). Existe uma série de aproximações bayesianas para fazer o ajustamento e seleção de modelos, destacando-se o DIC (*Deviance information criterion*) e os Fatores de Bayes ( ver Spiegelhalter *et al*, 2002). No entanto o algoritmo de Monte Carlo via cadeias de Markov com saltos reversíveis - RJMCMC (*reversible Jump Monte Carlo Markov Chain*)- introduzido por Green (1995), permite tratar simultaneamente a questão da seleção do modelo e da estimação dos parâmetros. Para tal contribui o facto de ser possível a variação da dimensão do espaço do parâmetro entre as diferentes iterações da cadeia de Markov. Assim, pode ser visto como uma extensão do algoritmo de Metropolis-Hastings para um espaço de estados mais geral.

## 2 Como funciona o RJMCMC?

O algoritmo RJMCMC move-se entre espaços de parâmetros de uma coleção de modelos, em geral em número finito e na maioria das situações são modelos encaixados. Assim o método é uma combinação do algoritmo MCMC para um dado modelo (fixo) com o passo adicional que permite a movimentação entre diferentes modelos. Por uma questão de simplificação considera-se a situação de os modelos estarem encaixados, que é a situação mais frequentemente usada. Isto significa incluir ou retirar parâmetros do modelo atual. Suponham-se  $K$  modelos possíveis,  $\theta_k$  representa o conjunto de parâmetros do  $k$ -ésimo modelo,  $k = 1, \dots, K$ , i.e., de  $M_k$ .

### • Algoritmo RJMCMC, com $N$ iterações

1. Escolher um modelo inicial  $M_k$ , i.e., inicializar  $k$  e  $\theta_k$ , na iteração  $t = 1$ .
2. Para cada iteração  $t \geq 1$  fazer

- (a) movimento dentro do modelo: para o modelo fixo  $M_k$  - obtido no fim da iteração  $t - 1$  - atualizar os parâmetros  $\boldsymbol{\theta}_k$ , de acordo com a metodologia MCMC;
  - (b) movimento entre modelos: propôr a mudança para um novo modelo  $M_{k'}$ , calcular a probabilidade de aceitação para o movimento e decidir pela aceitação ou rejeição do novo modelo.
3. Incrementar a iteração  $t = t + 1$ ; se  $t < N$  voltar a 2.

- Movimento entre modelos

Veja-se o seguinte exemplo apresentado por Green (1995):

Suponha-se que se pretende fazer o movimento de um modelo 2 de parâmetros  $(\alpha, \beta)$  para o modelo 1 com um parâmetro. Esse parâmetro poderia ser estabelecido como  $\lambda = \frac{\alpha + \beta}{2}$ . No entanto, dado o modelo 1 de parâmetro  $\lambda$  que valores se deveriam escolher para  $(\alpha, \beta)$  no movimento contrário? Existe uma infinidade de possíveis escolhas para  $(\alpha, \beta)$  que conduzem a  $\lambda = \frac{\alpha + \beta}{2}$ . É pois necessária uma bijeção entre os parâmetros de ambos os modelos e tal só é possível se o número de parâmetros for o mesmo. Para tal adiciona-se uma variável auxiliar  $u$  ao modelo 1, que não tenha outro papel senão o de permitir que haja uma bijeção entre os dois modelos. Seja  $\alpha = \lambda + u$  e  $\beta = \lambda - u$ ; então  $\lambda = \frac{\alpha + \beta}{2}$  e  $u = \frac{\alpha - \beta}{2}$ .

- Probabilidade de aceitação do movimento entre modelos

De acordo com o algoritmo de Metropolis-Hastings a probabilidade de aceitação de um movimento dentro do modelo, dado  $\mathbf{x}$  e parâmetro  $\boldsymbol{\theta}$  é dada por:

$$\min \left\{ 1, \frac{\pi(\boldsymbol{\theta}'|\mathbf{x})}{\pi(\boldsymbol{\theta}|\mathbf{x})} \times \frac{q(\boldsymbol{\theta}|\boldsymbol{\theta}')}{q(\boldsymbol{\theta}'|\boldsymbol{\theta})} \right\},$$

onde  $\pi(\cdot|\mathbf{x})$  representa a distribuição *a posteriori* e  $q(\cdot|\cdot)$  a distribuição proponente para o parâmetro.

Admitam-se  $K$  modelos possíveis e  $\boldsymbol{\theta}_k$  os parâmetros do modelo  $k$ ,  $k = 1, \dots, K$ ; suponha-se que se está no modelo  $M_k$  de parâmetro  $\boldsymbol{\theta}_k$  e que se pretende fazer a mudança para o modelo  $M_{k'}$  de parâmetro  $\boldsymbol{\theta}'_{k'}$ . Sejam  $\mathbf{u}_k$  as variáveis auxiliares associadas ao modelo  $k$ ,  $k = 1, \dots, K$ , tais que para qualquer modelo  $k$ ,  $(\boldsymbol{\theta}_k, \mathbf{u})$  tem sempre o mesmo número de elementos, por exemplo  $L$ . Para  $1 \leq k, k' \leq K$ , seja  $g_{k \rightarrow k'}$  uma função invertível tal que  $(\boldsymbol{\theta}'_{k'}, \mathbf{u}') = g_{k \rightarrow k'}(\boldsymbol{\theta}_k, \mathbf{u})$  com  $g_{k' \rightarrow k} = g_{k \rightarrow k'}^{-1}$  e  $r_k(\mathbf{u})$  a proponente para as variáveis auxiliares  $\mathbf{u}$ . A probabilidade de aceitação do movimento do modelo  $k$  com parâmetro  $\boldsymbol{\theta}_k$ , para o modelo  $k'$  com parâmetro  $\boldsymbol{\theta}'_{k'}$ , é

$$\alpha(\boldsymbol{\theta}_k, \boldsymbol{\theta}'_{k'}) = \min \left\{ 1, \frac{\pi(k', \boldsymbol{\theta}'_{k'}|\mathbf{x})}{\pi(k, \boldsymbol{\theta}_k|\mathbf{x})} \times \frac{q(k|k')}{q(k'|k)} \times \frac{r_{k'}(\mathbf{u}')}{r_k(\mathbf{u})} \times J \right\},$$

onde  $\pi(k', \boldsymbol{\theta}'_{k'}|\mathbf{x})$  é a distribuição *a posteriori* dos parâmetros do modelo  $M_{k'}$ ,  $q(k'|k)$  é a probabilidade do movimento  $k \rightarrow k'$ ,  $r_k(\mathbf{u})$  a distribuição proponente para as variáveis auxiliares  $\mathbf{u}$  e  $J$  o jacobiano da transformação do movimento descrito.

É necessário calcular o jacobiano da transformação sempre que se faz um movimento entre modelos com diferentes parâmetros. Este jacobiano é o determinante da matriz  $V$  de dimensão  $L \times L$  definida por:

$$V = \left[ \frac{\partial g_{k \rightarrow k'}(\boldsymbol{\theta}_k, \mathbf{u})}{\partial(\boldsymbol{\theta}_k, \mathbf{u})} \right].$$

Evidentemente que o movimento inverso de  $M_{k'}$  para  $M_k$  é feito deterministicamente e é aceite com probabilidade

$$\alpha(\boldsymbol{\theta}'_{k'}, \boldsymbol{\theta}_k) = \left( \frac{1}{\alpha(\boldsymbol{\theta}_k, \boldsymbol{\theta}'_{k'})} \right)^{-1}.$$

Admita-se que o novo modelo a propor  $M_{k'}$  tem dimensão superior, i.e.,  $n_{k'} > n_k$ . Seguem-se os seguintes passos:

- propor o movimento  $M_k \rightarrow M_{k'}$  com probabilidade  $q(k'|k)$ ;
- gerar vetor  $\mathbf{u}$ ,  $n_{k'} - n_k$  dimensional, a partir de  $r_k(\mathbf{u})$ ;
- fazer  $(\boldsymbol{\theta}'_{k'}, \mathbf{u}') = g_{k \rightarrow k'}(\boldsymbol{\theta}_k, \mathbf{u})$ ;
- aceitar  $(\boldsymbol{\theta}'_{k'}, \mathbf{u}')$  com probabilidade  $\alpha(\boldsymbol{\theta}_k, \boldsymbol{\theta}'_{k'})$ .

### 3 Exemplo: Problema de mudança na localização

Considere-se o conjunto de dados obtidos por Jarrett (1979) que traduzem o número de acidentes em minas de carvão por ano, no período de 15 de março de 1851 a 22 de março de 1962. Os dados sugerem que houve uma redução no número de acidentes, por ano nesse período e que houve uma mudança no valor médio da distribuição, por volta do ano 1891.

Carlin et al (1992) sugeriram o modelo

$$\begin{aligned} X_i &\sim Po(\theta), & i = 1, \dots, k \\ X_i &\sim Po(\lambda), & i = k + 1, \dots, n (= 112). \end{aligned}$$

Este é tipicamente um problema de mudança de localização, i.e., existe uma mudança de comportamento em termos de localização, num instante desconhecido. Veja-se como é que o RJMCMC pode ser usado para analisar este problema:

#### Modelo 1:

$$\begin{aligned} X_i &\sim Po(\lambda), \quad i = 1, \dots, n \\ \text{apriori: } \lambda &\sim Gama(a, b) \end{aligned}$$

#### Modelo 2:

Existe  $k \in \{1, 2, \dots, n - 1\}$  tal que

$$\begin{aligned} X_i &\sim Po(\theta_1), & i = 1, \dots, k \\ X_i &\sim Po(\theta_2), & i = k + 1, \dots, n \end{aligned}$$

$$\text{a priori's: } \theta_i \sim Gama(a, b), i = 1, 2; \quad k \sim U\{1, 2, \dots, n - 1\}.$$

A atualização de parâmetros em cada um dos modelos faz-se usando as condicionais completas, correspondendo ao movimento dentro do modelo. São as seguintes para cada um dos modelos:

#### Modelo 1:

$$\lambda | \mathbf{x} \sim Gama(a + \sum_{i=1}^n x_i, b + n).$$

**Modelo 2:**

$$\theta_1 | \mathbf{x}, \theta_2, k \sim \text{Gama} \left( a + \sum_{i=1}^k x_i, b + n \right);$$

$$\theta_2 | \mathbf{x}, \theta_1, k \sim \text{Gama} \left( a + \sum_{i=k+1}^n x_i, b + n \right);$$

$$p(k | \mathbf{x}, \theta_1, \theta_2) = \frac{e^{k(\theta_2 - \theta_1)} \left( \frac{\theta_1}{\theta_2} \right)^{\sum_{i=1}^k x_i}}{\sum_{j=1}^n e^{j(\theta_2 - \theta_1)} \left( \frac{\theta_1}{\theta_2} \right)^{\sum_{i=1}^j x_i}}.$$

Relativamente ao movimento entre modelos: suponha-se a passagem do modelo 1 para o modelo 2. Uma vez que o 2º modelo tem dois parâmetros a mais que o 1º são necessárias duas variáveis auxiliares. Sejam  $u_1$  e  $u_2$  tais que:

- $u_1 \sim N(0, \sigma^2)$
- $u_2 \sim U\{1, 2, \dots, n-1\}$ .

Usando a transformação bijetiva entre  $(\lambda, u_1, u_2)$  e  $(\theta_1, \theta_2, k)$  definida por

$$\begin{cases} \theta_1 = \lambda e^{u_1} \\ \theta_2 = \lambda e^{-u_1} \\ k = u_2 \end{cases},$$

tem-se reciprocamente

$$\begin{cases} \lambda = \sqrt{\theta_1 \theta_2} \\ \theta_2 = \frac{1}{2} \log(\theta_1 / \theta_2) \\ u_2 = k \end{cases}.$$

Para calcular a probabilidade de aceitação do movimento do modelo 1 para o modelo 2 determinam-se os seguintes quocientes

**q.veros:**

$$\frac{L(\theta_1, \theta_2, k | \mathbf{x}, M_2)}{L(\lambda | \mathbf{x}, M_1)} = \left( \frac{\theta_1}{\lambda} \right)^{\sum_{i=1}^k x_i} e^{-k(\theta_1 - \lambda)} \left( \frac{\theta_2}{\lambda} \right)^{\sum_{i=k+1}^n x_i} e^{(n-k)(\theta_2 - \lambda)}.$$

**q.prioris:**

$$\frac{\pi(\theta_1, \theta_2, k | M_2)}{\pi(\lambda | M_1)} = \frac{b^a}{\gamma(a)} (\lambda \theta_1 \theta_2)^{a-1} e^{-b(\theta_1 + \theta_2 - \lambda)}.$$

**q.propon:** Como o movimento do modelo 2 para o modelo 1 é determinístico, apenas se tem

$$\frac{q(1|2)}{q(2|1)} = \left\{ \frac{1}{n-1} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{u_1}{2\sigma^2}\right) \right\}^{-1}.$$

**jacobiano:**

$$J = \begin{vmatrix} e^{u_1} & \lambda e^{u_1} & 0 \\ e^{-u_1} & -\lambda e^{-u_1} & 0 \\ 0 & 0 & 1 \end{vmatrix} = 2\lambda$$

**Probabilidade de aceitação 1 → 2:**

$$\alpha(1 \rightarrow 2) = \min \{1, q.\text{veros} \times q.\text{prioris} \times q.\text{propon} \times \text{jacobiano}\}.$$

## 4 Áreas de aplicação

Algumas áreas onde mais se tem destacado a metodologia RJMCMC:

- Modelos com mudança na localização

Uma das primeiras aplicações do RJMCMC foi em problemas de mudança de localização bayesianos, considerando desconhecidos *a priori* o seu número e a sua localização. Green (1995) analisou também o número de acidentes de minas usando um processo de Poisson considerando os dados medidos em cada dia em vez de ser por ano, em que a intensidade do processo é definida através de uma função em escada com um número desconhecido de saltos e com localização também desconhecida. Fan e Brooks (2000) aplicaram este método de saltos reversíveis para modelar a forma de túmulos pré-históricos, em que a curvatura da abóbada muda um número desconhecido de vezes.

- Modelos de mistura finita

Considerando os dados  $\mathbf{x}$ , o vetor parâmetro  $\boldsymbol{\theta}_k = (\phi_1, \dots, \phi_k)$  e  $k$  componentes, os modelos de mistura são definidos por

$$f(\mathbf{x}|\boldsymbol{\theta}_k) = \sum_{i=1}^k w_i f_i(\mathbf{x}|\phi_i),$$

onde  $w_i$  representa o peso da  $i$ -ésima componente de mistura  $f_i$  e tal que  $\sum_{i=1}^k w_i = 1$ . O número das componentes de mistura também é desconhecido. Richardson e Green (1997) analisaram dados, de natureza multimodal, referentes à atividade enzimática no sangue de 245 indivíduos, usando um modelo de mistura de normais com o objetivo de identificar subgrupos de indivíduos com velocidade elevada ou fraca de metabolização. Tadesse et al (2005) estenderam este estudo tratando o problema da agregação em dados de natureza multidimensional (dados de *microarrays* referentes ao estudo de cancro endometrial), usando um modelo de normais multivariadas com um número desconhecido de componentes e selecionando as variáveis que melhor discriminam os diferentes grupos.

- Seleção de variáveis

O problema de seleção de variáveis surge quando se pretende modelar a relação entre a variável resposta  $Y$  e  $p$  potenciais variáveis explicativas  $x_1, \dots, x_p$ . O problema fundamental na construção do modelo de regressão múltipla é o da seleção dos preditores a incluir. Considere-se, por exemplo, um modelo de regressão com erros gaussianos

$$Y = X_{\boldsymbol{\gamma}}\boldsymbol{\beta}_{\boldsymbol{\gamma}} + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2 I),$$

onde  $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_p)$  é um vetor binário que indexa o subconjunto de variáveis  $x_1, \dots, x_p$  a incluir no modelo linear,  $X_{\boldsymbol{\gamma}}$  é a matriz de delineamento cujas colunas correspondem ao subconjunto indexado por  $\boldsymbol{\gamma}$  e  $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$  é o correspondente subconjunto de coeficientes de regressão. Realçam-se os trabalhos de Smith e Kohn (1996) e de Nott e Leonte (2004).

- Abordagem não paramétrica bayesiana

A metodologia RJMCMC tem sido usada com sucesso na estatística não paramétrica com o objetivo de automatizar o processo de seleção dos nós, o seu número e a sua localização, quando se usa o modelo *spline* de ordem  $p$  para o ajustamento da curva. Para estimar

$$f(x) = \alpha_0 + \sum_{j=1}^p \alpha_j x^j + \sum_{i=1}^k \eta_i (x - k_i)_+^p, \quad x \in [a, b[, \quad w_+ = \max(o, w),$$

é necessário calcular o valor desconhecido de *knots*  $k$ , as suas localizações  $k_i$  e os correspondentes coeficientes de regressão  $\alpha_i$  e  $\eta_j$ ,  $i = 0, \dots, p$  e  $j = 1, \dots, k$ .

Este procedimento envolve movimentos que adicionam, retiram ou recolocam os *knots* no modelo. Vejam-se os trabalhos de Denison *et al* (1998) e DiMatteo *et al* (2001).

- Modelos de séries temporais

Existem múltiplos modelos, lineares e não lineares, sugeridos na literatura para modelarem dados  $x_1, \dots, x_n$  temporalmente dependentes. O modelo autorregressivo de ordem  $k$ , encarada como desconhecida, foi tratado na literatura por diversos autores nomeadamente Elhers e Brooks (2003), Vermaak *et al* (2004) e os modelos limiars foram tratados por Campbell (2004). Pode-se destacar o trabalho de Enciso-Mora (2008) aplicado a modelos de contagem usando o operador *thinning* binomial, designados por INARMA (Integer AutoRegressive Moving Average), que usam o salto reversível conjugado com a ampliação dados, e onde em cada passo são aumentadas ou diminuídas as ordens anteriores apenas numa unidade e é mantida fixa a média do processo.

Uma das maiores dificuldades na implementação dos saltos reversíveis tem a ver com a arbitrariedade na escolha da função  $g_{k \rightarrow k'}$  e das distribuições proponentes  $r_k(\mathbf{u})$  para o vetor  $\mathbf{u}$ . O bom desempenho da metodologia está intimamente relacionada com a adequabilidade da escolha de  $g_{k \rightarrow k'}$  e de  $r_k(\mathbf{u})$ , dependendo particularmente da forma e parâmetros de  $r_k(\mathbf{u})$ . Uma possibilidade explorada por alguns autores consiste na utilização do movimento de nascimento e morte feito em simultâneo, mantendo fixa a dimensão do modelo; com o objetivo de reduzir a dimensionalidade/complexidade do problema faz-se a marginalização de componentes ou o aumento de dados/aumento do espaço de estados da distribuição *a posteriori* alvo. Para aprofundar esta metodologia sugere-se consultar Sisson (2005) e Fan e Sisson (2011).

**Agradecimentos** Este trabalho foi parcialmente financiado por *fundos FEDER* através do *COMPETE*–Programa Operacional Factores de Competitividade e por *fundos Portugueses* através do *Centro de Investigação e Desenvolvimento em Matemática e Aplicações* e FCT- Fundação para a Ciência e a Tecnologia, dentro do projeto PEst-C/MAT/UI4106/2011 e FCOMP-01-0124-FEDER- 022690.

## Referências

- [1] Campbell, E.P.(2004). Bayesian selection of threshold autoregressive models. *J. of Time Series Analysis* **25**, 467-482.
- [2] Carlin, B.P., Gelfand, A.E. e Smith, A.F.M. (1992) Hierarchical Bayesian Analysis of Change-point Problems. *J. of the Royal Statistical Society, C* **41**, 389-405.
- [3] Denison, D.G.T., Mallick, B.K. e Smith, A.F.M. (1998). Bayesian curve-fitting with free-knot splines. *J. of the Royal Statistical Society, B* **60**, 330-350.

- [4] DiMatteo, Genovese, C.R. e Kass, R.E. (2001). Automatic Bayesian curve fitting. *Biometrika* **88**, 1055-1071.
- [5] Enciso-Mora, V., Neal, P. e Subba Rao, T. (2008). Efficient order selection algorithms for integer-valued ARMA processes. *J. of Time Series Analysis* **30**, 1-18.
- [6] Ehlers, R.S. e Brooks, S.P. (2003). Constructing general efficient proposals for reversible jump MCMC. Technical Report, Department of Statistics, Federal University of Paraná.
- [7] Fan, Y. e Brooks, S.P. (2000). Bayesian modelling prehistoric corbelled domes. *J. of the Royal Statistical Society, D* **49**, 339-354.
- [8] Fan, Y. e Sisson, S.A. (2000). Reversible Jump MCMC. In Brooks, S., Gelman, A., Jones, C.L e Meng, X-L editors, *Handbook of Markov Chain Monte Carlo*, Taylor and Francis Group, USA, pp. 67 - 87.
- [9] Green, P.J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**, 711-732.
- [10] Hastings, W.K. (1970). Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika* **57**, 59-109.
- [11] Jarrett, R.G. (1979). A note on the intervals between coal-mining disasters. *Biometrika* **66**, 191-1092.
- [12] Metropolis, N., Rosenbluth, A.H., Rosenbluth, M.N. e Teller, A.H. (1953). Equation of State Calculations by Fast Computing Machines. *J. of Chemical Physics* **21**, 1087-1092.
- [13] Nott, D.J. e Leonte, D. (2004). Sampling schemes for Bayesian variable selection in generalised linear models. *J. Computational and Graphical Statistics*. **13**, 362-382.
- [14] Richardson, S. e Green, P.J. (1997). On Bayesian analysis of mixtures with an unknown number of components (with discussion). *J. of the Royal Statistical Society, B* **59**, 731-792.
- [15] Sisson, S.A. (2005). Trans-dimensional Markov chains: A decade of progress and future perspectives. *J. of American Statistical Association*. **100**, 1077-1089.
- [16] Smith, M e Kohn, R. (1996). Nonparametric regression using Bayesian variable selection. *J. of Econometrics* **75**, 317-344.
- [17] Spiegelhalter, D.J., Best, N.G., Carlin, B.P. e van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *J. of the Royal Statistical Society, B* **64**, 583-639.
- [18] Tadesse, M., Sha, N. e Vannucci, M. (2005). Bayesian variable selection in clustering high-dimensional data. *J. of American Statistical Association* **100**, 602-617.
- [19] Vermaak, J., Andrieu, C., Doucet, A. e Godsill, S.J. (2004). Reversible jump Markov chain Monte Carlo strategies for Bayesian model selection in autoregressive processes. *J. of Time Series Analysis* **25**, 785-809.