Escola de Matemática Aplicada

Assignment II: Advanced simulation techniques.

Julie P. S. Santos
Programa: *Doutorado em Modelagem Matemática*
Disciplina: *Estatística Computacional*
Professor: *Luiz Max Carvalho*
Email: *julie.santos@fgv.edu.br*

Submission date: *December 13, 2021*

Background

We have by now hopefully acquired a solid theoretical understanding of simulation techniques, including Markov chain Monte Carlo (MCMC). In this assigment, we shall re-visit some of the main techniques in the field of Simulation. The goal is to broaden your knowledge of the field by implementing one of the many variations on the general theme of simulation algorithms.
Each method/paper brings its own advantages and pitfalls, and each explores a slightly different aspect of Computational Statistics. You should pick **one** of the listed papers and answer the associated questions.
In what follows, ESS stands for effective sample size, and is similar to n_{eff} we have encountered before: it measures the number of effectively uncorrelated samples in a given collection of random variates.

Blocked Gibbs sampling

The so-called Gibbs sampler is a work horse of Computational Statistics. It depends on decomposing a target distribution into conditional densities from which new values of a given coordinate can be drawn.
One of the difficulties one might encounter with the Gibbs sampler is that it might be slow to converge, specially in highly-correlated targets. In Statistics, multilevel models (also called hierarchical or random effects) are extremely useful in modelling data coming from stratified structures (e.g. individuals within a city and cities within a state) and typically present highly correlated posterior distributions.
One way to counteract the correlation between coordinates in the Gibbs sampler is to **block** them together, and sample correlated coordinates jointly.
For this assigment you are referred to the 2009 *Journal of Computational and Graphical Statistics* paper by Tan and Hobert.

Question 1

Precisely describe the so-called blocked Gibbs sampler; *Hint:* you do not need to describe theoretical properties of the algorithm given in this paper; a general description of the algorithm should suffice.

Solução:

O método de amostragem Gibbs é um método de Monte Carlo via Cadeias de Markov. Cadeias de Markov são sequências de variáveis independentes na qual o estado atual do sistema não guarda informações acerca de outros estados, um exemplo comum para as cadeias de Markov é o clássico "Andar do bêbado". É usualmente descrito na literatura como um caso particular da amostragem Metropolis-Hasting, porém o caso específico da amostragem Gibbs é geralmente usada para os casos que a distribuição é multivariada. Uma vez que a distribuição condicional é completa, é simples o processo de amostragem e o algoritmo atualiza um parâmetro por vez, ou um bloco de parâmetros no caso de Gibbs blocado, gerando candidatos pela distribuição condicional completa e aceitando, desta forma, todos os valores. [1]

O teorema de Hemmersley-Clifford nos trás luz a motivação do uso da distribuição condicional completa determinar a distribuição conjunta, sendo ela definida positiva.

Teorema de Hammersley-Clifford. Se $p(x) > 0$ para todo x, então a distribuição conjunta $p(x)$ é determinada unicamente pelas condicionais completas. Mais precisamente,

$$p(x) = p(y) \prod_{k=1}^{K} \frac{p_{jk}(x_{jk}|x_{j1}, \cdots, x_{jk-1}, y_{jk+1}, \cdots, y_{jK})}{p_{jk}(y_{jk}|x_{jk}, \cdots, x_{jk-1}, y_{jk+1}, \cdots, y_{jK})} \tag{1}$$

para cada permutação $j = j(k)$ em $\{1, \cdots, n\}$ e todo y.

Demonstração. [2]

Assumindo que a condição de positividade está previamente satisfeita, tome $Q(x) = log[p(x)/p(0)]$ que possui uma expansão única da forma

$$Q(x) = \sum_{1\leq i\leq n} x_i G_i(x_i) + \sum_{1\leq i<j\leq n} x_i x_j G_{i,j}(x_i, x_j) + \cdots x_1 x_2 \cdots x_n G_{1,2,\cdots,n}(x_1, x_2, \cdots, x_n)$$

em que

$$G_{i,j,\cdots,s}(x_i, x_j, \cdots, x_s) \neq 0 \text{ apenas se } \{1, j, \cdots, s\} \in cl(\mathcal{G}).$$

Em que $\mathcal{G}$ é um grafo definido como $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, s.t. $\mathcal{V} = \{X_1, \cdots, X_n\}$ e $\{X_i, X_j\} \in \mathcal{E}$ se somente se

$$p(x_i|\frac{\{x_1, \cdots, x_n\}}{\{x_i\}}) \neq p(x_i|\frac{\{x_1, \cdots, x_n\}}{\{x_i, x_j\}})$$

obs: Esta é uma simplificação da prova realizada por Besag (1974). □

[1] https://repositorio.unb.br/bitstream/10482/40488/3/2021˙DeboraCristianedosSantos.pdf

[2] Julian, Besag (1974). Spatial Interaction and the Statistical Analysis of Lattice Systems. Journal of the Royal Statistical Society. Series B (Methodological), Vol. 36, No. 2 (1974), pp. 192-236

Abaixo, segue o algoritmo do método de amostragem Gibbs (simples ou em blocos):

Seja $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_d)$ um vetor aleatório $\in \mathbb{R}^d$.

Algoritmo

Entrada:

$N \in \mathbb{N}$ representando o número de amostras desejado;

Um valor inicial $\theta^{(0)}$

Simulação

para $t = 1, 2, \ldots, N$ **faça**

Gere $\theta_1^{(t)} \sim f(\theta_1 \mid \theta_2^{(t-1)}, \theta_3^{(t-1)}, \ldots, \theta_d^{(t-1)})$

$\theta_2^{(t)} \sim f(\theta_2 \mid \theta_1^{t}, \theta_3^{(t-1)}, \ldots, \theta_d^{(t-1)})$

$\vdots$

$\theta_d^{(t)} \sim f(\theta_d \mid \theta_1^{t}, \theta_2^{t}, \ldots, \theta_{d-1}^{t})$;

fim

Saída:

A cadeia $\{\boldsymbol{\theta}^{(t)} \mid t = 1, \ldots, N\}$ com os valores gerados.

Question 2

Explain the advantages – both theoretical and practical – of a clever blocking scheme;

Solution:
De acordo com o artigo de Tan e Hobert (2009) o método pode ser aplicável a distribuições à posteriori intratáveis oferencendo o mesmo nível de confiança de um método de Monte-Carlo clássico. Funciona bem em diversos casos, contudo não em todos. Para os casos de interesse, é possível implementar a simulação de forma relativamente simples. Também produz erros padrões assintoticamente válidos para as médias ergóticas usados para estimar a esperança a posteriori intratável.O desvio padrão pode ser diretamente aplicado com o Teorema Central do Limite e determinar o tamanho da amostra que irá oferecer o intervalo de confiança escolhido. Esse fato é importante pois está diretamente relacionado a ergodicidade da cadeia de Markov gerada pelo método de amostragem em blocos.

Intuitivamente falando, é natural pensar que amostrar em blocos (apropriadamente escolhidos) é computacionalmente mais efetivo no tempo do que atualizar um parâmetro por vez. De fato, apresenta uma certa vantagem de tempo a um maior custo computacional (no free lunch). Contudo, acredito fortemente que a vantagem do método reside apenas sobre a amostragem de Gibbs simples, sendo a em blocos geometricamente ergódica e com isso possuir sua convergência garantida teoricamente.

□

Question 3

Would it be possible to apply the “simple” Gibbs sampler in this example? Why?

Solution:
Não existe impossibilidade técnica ou teórica para a aplicação do amostrador de Gibbs simples. Contudo, a atualização dos parâmetros em bloco levam a propriedades de convergência melhores do que a versão simples do amostrador de Gibbs.

Question 4

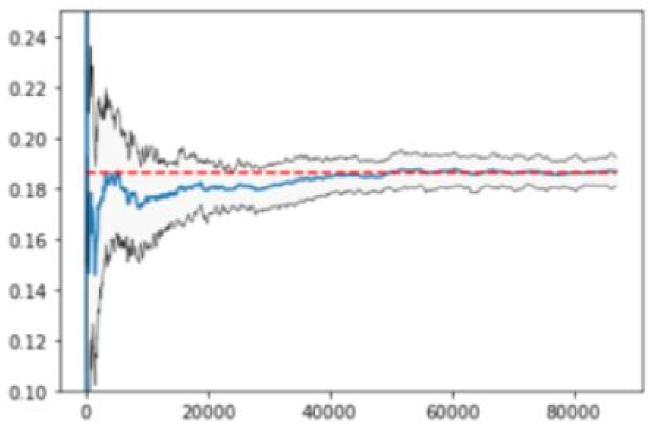
Implementation:
a) Implement the blocked Gibbs sampler discussed in the paper in order to fit the model of Section 1 of the article to the data described in Section 5 therein.
b) Assess convergence (or lack thereof) and mixing of the resulting chain.
c)Confirm your results agree with those given by the original authors up to Monte Carlo error.

Solution:
A implementação foi feita com base no código disponibilizado pelos autores nos arquivos suplementares do artigo. O código foi disponibilizado na linguagem R e extremamente útil na replicação da amostragem Gibbs em blocos na linguagem python. O código da implementação em python se encontra no formato notebook disponível no sítio *Git Hub* no link https://github.com/juliezousa/Estatistica-Computacional-FGV-PhD/blob/main/Final_Assignment_Est_Comp/Final_Assignment_Julie.ipynb .

Para 5,000 regenerations foram obtidos os seguintes resultados:

	g_R	γ^2	$\sqrt{(\gamma^2/R)}$	$g_R \pm 2\sqrt{(\gamma^2/R)}$
σ_θ^2	0.18634	0.03458	0.00263	(0.18024, 0.19245)
σ_e^2	0.62112	0.01080	0.00147	(0.61818, 0.62406)
$\sigma_\theta^2/(\sigma_\theta^2+\sigma_e^2)$	0.20881	0.03539	0.00302	(0.20277, 0.21484)

que se mostraram muito próximos dos resultados obtidos no artigo base. A simulação para 40,000 regenerations também foi implementada e analogamente apresentou resultado concordantes com do artigo. Seu cálculo, por sua vez, foi mais custoso computacionalmente devido ao elevado número de iterações (da ordem de centenas de milhares). A plotagem gráfica para essa ordem não foi suportada para minha máquina, por conta disso apresento os dados para 5,000 regenerations que também apresentam uma excelente visualização da convergência do método.



Question 5

Comment on the significance of geometric ergodicity for the blocked Gibbs sampler proposed in the article.

Solution:

A geometricidade ergótica do amostrador de Gibbs em blocos se apresenta como uma das vantagens do método em blocos em relação ao simples. Isso se deve ao fato de que cadeias de Markov ergóticas são regenarativas. Da hipótese ergótica, em um espaço suficientemente grande de tempo todos os estados do conjunto revisitam suas posições anteriores. Também implica que o o estimador é fortemente consistente, tal implicação permite a construção de um intervalo de confiança assintótico da aplicação do Teorema Central do Limite. A teoria de ergoticidade também assume que a cadeia de Markov irá convergir para uma distribuição estacionária, isso quer dizer que os valores da cadeia giram em torno da média e a convergência é garantida. Logo, a convergênia quase certa do método se torna um aspecto significante do amostrador de Gibbs proposto por Gibbs e Hobert.