

Algoritmo ejeção-absorção metropolizado para segmentação de imagens

Alexandre Pitanguí Calixto

Orientador: Prof. Dr. Luís Aparecido Milan

Coorientador: Prof. Dr. Erlandson Ferreira Saraiva

Tese apresentada ao Departamento de Estatística da Universidade Federal de São Carlos - DEs/UFSCar, como parte dos requisitos para obtenção do título de Doutor em Estatística.

São Carlos

Dezembro de 2014

**Ficha catalográfica elaborada pelo DePT da
Biblioteca Comunitária/UFSCar**

C154ae Calixto, Alexandre Pitangui.
Algoritmo ejeção-absorção metropolizado para
segmentação de imagens / Alexandre Pitangui Calixto. --
São Carlos : UFSCar, 2015.
78 f.

Tese (Doutorado) -- Universidade Federal de São Carlos,
2014.

1. Estatística. 2. Segmentação de imagem. 3. Potts,
Modelo de. 4. Distribuição de Gibbs. 5. Algoritmo de
Swendsen-Wang. 6. *Reversible Jump*. I. Título.

CDD: 519.5 (20^a)



UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia
Programa de Pós-Graduação em Estatística

Folha de Aprovação

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a Defesa de Tese de Doutorado do candidato Alexandre Pitangui Calixto, realizada em 19/12/2014:

Prof. Dr. Luis Aparecido Milan
UFSCar

Prof. Dr. Alexandre Luis Magalhães Levada
UFSCar

Prof. Dr. Erandson Ferreira Saraiva
UFMS

Prof. Dr. Jose Galvao Leite
UFSCar

Prof. Dr. Ricardo Sandes Ehlers
USP

Profa. Dra. Roseli Aparecida Leandro
USP

Agradeço,

aos meus pais pelo apoio e incentivo e por proporcionar a mim e a meus irmãos a oportunidade de estudar e termos sucesso em nossas carreiras;

ao professor Dr. Luís Aparecido Milan pela orientação, pelas ideias e principalmente pelo exemplo de dedicação e disciplina de trabalho. Agradeço ainda, as suas palavras sábias de apoio, paciência, cuidado e incentivo à pesquisa para que o trabalho que realizamos fosse concluído com sucesso;

ao professor Dr. Erlandson Ferreira Saraiva pela coorientação e principalmente pelos bons momentos que passamos juntos, estudando e discutindo sobre nossas pesquisas ou apenas conversando sobre diversos assuntos no happy-hour;

às pessoas que durante o desenvolvimento deste trabalho, contribuíram de diferentes formas para sua realização. Dentre elas, destaco os colegas de Curso, todos os professores e os funcionários do DEs-UFSCar que de alguma forma auxiliaram na execução deste trabalho;

à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro.

Resumo

Nesta tese, modelamos uma imagem através de uma grade regular retangular e assumimos que esta grade é dividida em múltiplas regiões disjuntas de *pixels*. Quando duas ou mais regiões apresentam a mesma característica, a união dessas regiões forma um conjunto chamado de componente. Associamos a cada *pixel* da imagem uma variável indicadora não observável que indica a componente a que o *pixel* pertence. Estas variáveis indicadoras não observáveis são modeladas através da distribuição de probabilidade de Gibbs com parâmetro de regularização espacial β . Assumimos que β e o número de componentes K são desconhecidos. Para estimação conjunta dos parâmetros de interesse, propomos um algoritmo MCMC denominado de ejeção-absorção metropolizado (EAM). Algumas vantagens do algoritmo proposto são: (i) O algoritmo não necessita da especificação de uma função de transição para realização dos movimentos ejeção e absorção. Ao contrário do algoritmo *reversible jump* (RJ) que requer a especificação de “boas” funções de transição para ser computacionalmente eficiente; (ii) Os movimentos ejeção e absorção são desenvolvidos com base nos dados observados e podem ser rapidamente propostos e testados; (iii) Novas componentes são “criadas” com base em informações provenientes de regiões de observações e os parâmetros das novas componentes são gerados das distribuições *a posteriori*. Ilustramos o desempenho do algoritmo EAM utilizando conjuntos de dados simulados e reais.

Palavras chave: segmentação de imagem; modelo Potts; distribuição de Gibbs; algoritmo Swendsen-Wang; *reversible jump*; atualização *split-merge*.

Abstract

We proposed a new split-merge MCMC algorithm for image segmentation. We describe how an image can be subdivided into multiple disjoint regions, with each region having an associated latent indicator variable. The latent indicator variables are modeled with a prior Gibbs distribution governed by a spatial regularization parameter. Regions with same label define a component. Pixels within a component are distributed according to a Gaussian distribution. We treat the spatial regularization parameter and the number of components K as unknown. To estimate K , the spatial regularization parameter and the component parameters we propose the Metropolised split-merge (MSM) algorithm. The MSM comprises two type of moves. The first one, is a data-driven split-merge move. These movements change the number of components K in the neighborhood $K \pm 1$ and are accepted according to Metropolis-Hastings acceptance probability. After a split-merge step, the component parameters, the spatial regularization parameter and latent allocation variables are updated conditional on K by using the Gibbs sampling, the Metropolis-Hastings and Swendsen-Wang algorithm, respectively. The main advantage of the proposed algorithm is that it is easy to implement and the acceptance probability for split-merge movements depends only of the observed data. The performance of the proposed algorithm is verified using artificial datasets as well as real datasets.

Sumário

Resumo	ii
Abstract	iii
1 Introdução	1
1.1 Motivação	1
1.2 Revisão bibliográfica	2
1.3 Propostas e organização	4
2 Modelo de Potts	8
2.1 Grafos	8
2.2 Modelo de Potts	12
2.3 Algoritmo Swendsen-Wang	15
2.3.1 <i>Slice sampler</i> caso simples	16
2.3.2 <i>Slice sampler</i> caso geral	20
2.3.3 Algoritmo Swendsen-Wang	25
3 Modelo bayesiano para segmentação de imagens	28
3.1 Modelo bayesiano	28
3.1.1 Distribuições <i>a priori</i>	33
3.1.2 Distribuições <i>a posteriori</i>	35
3.2 Algoritmo ejeção-absorção metropolizado	36
3.2.1 Atualizações condicionadas em K	36
3.2.2 Atualização de K	39

3.2.3	Algoritmo EAM	47
4	Estudo de simulação e aplicação	49
4.1	Dados simulados	50
4.2	Aplicação	58
5	Considerações finais	60
	Referências bibliográficas	74

Capítulo 1

Introdução

1.1 Motivação

Com o avanço tecnológico, houve um aumento significativo do uso de computadores para obtenção de informações provenientes de imagens. Estas informações são utilizadas para tomada de decisões em diversas áreas, tais como: na engenharia, com a descoberta de fissuras em materiais; na biologia, com a contagem de microrganismos em amostras orgânicas; e na medicina, com a contagem de glóbulos brancos em amostras de sangue, identificação de regiões com tumores em imagens de ressonância magnética; entre outras.

O procedimento de obtenção e análise das informações provenientes de imagens é denominado na literatura de processamento digital de imagens (PDI). O PDI consiste de técnicas matemáticas utilizadas para extrair e analisar as informações contidas em uma imagem e pode ser dividido em quatro etapas:

- (i) Processamento geral: Nesta etapa, uma imagem digital é transformada em uma matriz de números, onde o valor de cada coordenada da matriz representa a cor ou a intensidade do *pixel*¹ de posição correspondente na imagem. Ou seja, esta matriz reproduz a imagem *pixel a pixel*;
- (ii) Segmentação: A segmentação consiste do particionamento da imagem em múltiplas regiões homogêneas.

¹*Pixel* é definido como sendo o menor elemento de uma imagem digital.

tiplas regiões (conjunto de *pixels* contíguos), que são consideradas homogêneas segundo algum critério;

- (iii) Detecção de características: Nesta etapa são obtidas informações, tais como forma e topologia das regiões identificadas em (ii);
- (iv) Interpretação: Nesta etapa é feita a interpretação dos dados quantitativos obtidos nas etapas (ii) e (iii). Ou seja, a informação da imagem é transformada em conhecimento.

O objetivo da pesquisa relatada neste texto é o desenvolvimento de um método estatístico, baseado em inferência bayesiana e métodos de Monte Carlo em cadeias de Markov (*Markov chain Monte Carlo*, MCMC), para segmentação de imagens. A expressão segmentação de imagens vem do termo em inglês “*image segmentation*”. Segundo Gonzalez e Woods (1992), a segmentação é uma das etapas mais difíceis e determina o sucesso de toda a análise. A análise se torna mais complicada quando o número de cores ou intensidades de tons cinza da imagem é desconhecido. Dessa forma, uma abordagem estatística que seja eficiente e computacionalmente viável se torna necessária.

1.2 Revisão bibliográfica

Em segmentação de imagem, os *pixels* pertencentes a uma mesma região são semelhantes com relação a alguma característica ou propriedade computadorizada, tal como cor, intensidade ou textura. Regiões adjacentes são significativamente diferentes em relação à mesma característica. Quando duas ou mais regiões apresentam a mesma característica, a união destas regiões forma um conjunto chamado de componente. Segundo Unnikrishnan *et al.* (2007), ao selecionar uma partição que seja capaz de destacar as propriedades de cada componente, obtemos uma representação compacta da imagem em termos das suas partes proveitosas, que ressaltam as informações presentes na imagem.

Na literatura, existem várias abordagens estatísticas aplicadas à segmentação de imagem, entre elas, *histogram clustering*, proposta por Puzicha *et al.* (1999),

segmentação *k-means*, proposta por Samson *et al.* (2000), *normalized cuts*, proposta por Shi e Malik (2000) e mistura de distribuições normais, proposta por Hermes *et al.* (2002). Embora simples e computacionalmente viáveis, estes métodos podem ser implementados somente para casos em que o número de componentes K pode ser previamente fixado. Isto pode ser um problema pois, na prática, o valor K nem sempre é conhecido e precisa ser estimado. Dessa forma, o objetivo é obter um método computacional que permita K ser desconhecido e que este seja estimado juntamente com os parâmetros das componentes.

Para situações em que K é desconhecido, Green e Richardson (2002) propõem um modelo markoviano oculto (do inglês *hidden Markov model*, HMM), com dependência espacial, e modelam a heterogeneidade dos dados através do modelo Potts (Potts, 1952). Para estimação conjunta de K e dos parâmetros das componentes, os autores utilizam o algoritmo *reversible-jump* (RJ) com movimentos denominados de *split-merge*, onde o movimento *split* divide uma componente em duas, aumentando o número total de componentes em uma unidade, $K + 1$, e o movimento *merge* faz a união entre duas componentes adjacentes obtendo uma nova componente, diminuindo o número total de componentes em uma unidade, $K - 1$. Estes movimentos mudam o valor de K na vizinhança $K \pm 1$, respectivamente, e preservam os momentos associados à(s) componente(s) escolhida(s) para a proposta de um *split* ou *merge*. No entanto, este método apresenta resultados satisfatórios somente em problemas de pequeno ou médio portes (Dobra *et al.*, 2004; Jones *et al.*, 2005; Hans *et al.*, 2007). Para análise de imagens, onde o número de *pixels* é grande, o método pode apresentar uma taxa de aceitação baixa e uma convergência muito lenta, inviabilizando sua utilização prática.

Utilizando as ideias de Green e Richardson (2002), alguns autores propuseram aplicações do algoritmo RJ para análise de imagens. Entre eles, Dryden *et al.* (2003) utilizam o algoritmo RJ para analisar imagens de ervas daninhas e plantações, em que um movimento *split* é realizado selecionando uma componente k entre as K componentes que formam a imagem, composta por pelo menos duas regiões. Em seguida, é “criada” uma nova componente k^* , vazia, com parâmetros θ_{k^*} gerados da

distribuição *a priori*. Então, um novo estado proposto é obtido alocando as regiões pertencentes à antiga componente k entre as novas componentes k e k^* com probabilidade $1/2$, ficando com um total de $K + 1$ componentes. Zhang *et al.* (2007) utilizam o algoritmo RJ para analisar imagens de ressonância magnética do cérebro de 11 indivíduos. Nesta proposta, Zhang *et al.* (2007) comparam o desempenho de sua proposta ao algoritmo EM (*Expectation Maximization*). Zhang *et al.* (2010) utilizam o mesmo procedimento para identificar alterações na permeabilidade vascular de tumores cerebrais. Johnson e Piert (2009) propõem um modelo bayesiano bivariado para análise de imagens e desenvolvem o procedimento de estimação utilizando o algoritmo RJ. Os movimentos *split* e *merge* são desenvolvidos utilizando um procedimento de alocação baseado no algoritmo Swendsen-Wang (Swendsen e Wang, 1987). Estes autores aplicam o modelo em um estudo sobre a relação entre hipóxia² e angiogênese³ em tumores. Askari *et al.* (2013) propõem um algoritmo RJ para segmentação de imagens de radar (*synthetic aperture radar*, SAR). Para identificar o número de regiões homogêneas, os autores utilizam o algoritmo RJ com os movimentos *split-merge*.

1.3 Propostas e organização

Nesta tese, desenvolvemos uma abordagem bayesiana hierárquica para segmentação de imagens em situações onde o número de componentes K é desconhecido. Para isto, modelamos uma imagem através de uma grade regular retangular e assumimos que esta grade é dividida em múltiplas regiões disjuntas de *pixels*. Associamos a cada *pixel* uma variável indicadora não observável que indica a componente à qual o *pixel* pertence. Estas variáveis indicadoras não observáveis são denotadas por $\mathbf{S} = (S_1, \dots, S_n)$ e são modeladas pela distribuição de Gibbs com parâmetro de regularização espacial β , onde n é a quantidade de *pixels* da imagem. Além disso,

²Trata-se de um estado de baixo teor de oxigênio nos tecidos orgânicos cuja ocorrência é atribuída a diversos fatores.

³É o termo usado para descrever o mecanismo de crescimento de novos vasos sanguíneos a partir dos já existentes.

assumimos que os *pixels* pertencentes a uma componente k são distribuídos de acordo com uma distribuição normal com média μ_k e variância σ_k^2 , para $k = 1, \dots, K$. Para os parâmetros (μ_k, σ_k^2) , consideramos distribuições *a priori* conjugadas, para $k = 1, \dots, K$. Assumimos que β e K são desconhecidos e consideramos distribuições *a priori* gama para β e uniforme discreta para K .

Para estimação conjunta dos parâmetros de interesse $(\mu_k, \sigma_k^2, \mathbf{S}, \beta, K)$, propomos um novo algoritmo MCMC denominado ejeção-absorção metropolizado (EAM). O algoritmo proposto é composto de dois tipos de movimentos:

- (i) Movimentos que não mudam o número de componentes: neste movimento, os parâmetros (μ_k, σ_k^2) , β e \mathbf{S} são atualizados condicionado a K , usando os algoritmos *Gibbs sampling*, Metropolis-Hastings e Swendsen-Wang, respectivamente, para $k = 1, \dots, K$;
- (ii) Movimentos que mudam o número de componentes: neste movimento, o número de componentes K é atualizado via movimentos denominados ejeção e absorção. Estes movimentos são realizados utilizando um procedimento de alocação probabilística, baseado nos dados observados e mudam o número K de componentes para $K \pm 1$.

Para manter válida a equação de balanceamento em relação à distribuição *a posteriori* quando propomos os movimentos ejeção e absorção, a probabilidade de aceitação para estes movimentos são calculadas de acordo com o procedimento proposto por Carlin e Chib (1995), Godsill (2001) and Dellaportas *et al.* (2002). Este procedimento considera uma parametrização alternativa, onde são utilizadas pseudo distribuições *a prioris* para completar o espaço paramétrico e assim desenvolver o procedimento em um espaço paramétrico de dimensão fixa. Dessa forma, a probabilidade de aceitação para os movimentos ejeção-absorção é dada pela probabilidade de aceitação do algoritmo Metropolis-Hastings.

Escolhendo a densidade *a posteriori* dos parâmetros das componentes como a densidade geradora de candidatos, a razão de verossimilhanças e a razão das densidades *a priori* (da probabilidade de aceitação Metropolis-Hastings) se cancelam

com o termo correspondente na densidade *a posteriori*, eliminando os parâmetros da probabilidade de aceitação. Assim, o algoritmo proposto realiza uma atualização Metropolis-Hastings com uma probabilidade de aceitação que depende somente dos dados associados com a componente(s) selecionada(s) para uma ejeção ou absorção.

A principal diferença do método proposto com relação ao algoritmo RJ é que quando uma componente nova é “criada”, através de uma proposta ejeção ou absorção, esta determina uma partição nos dados observados. Isto ocorre devido à forma que as propostas ejeção e absorção são desenvolvidas, baseadas nos dados e não nos parâmetros como no algoritmo RJ.

As contribuições presentes no algoritmo proposto são:

- (i) O algoritmo não necessita da especificação de uma função de transição para realização dos movimentos ejeção e absorção, ao contrário do algoritmo RJ, que requer a especificação de “boas” funções de transição para ser computacionalmente eficiente;
- (ii) As propostas ejeção e absorção são desenvolvidas com base nos dados observados, ao contrário dos métodos usuais, onde o procedimento de atualização de K que é baseada na proposta de uma nova configuração para as variáveis não observáveis;
- (iii) Os movimentos ejeção e absorção podem ser rapidamente propostos e testados; Além disso, novas componentes são “criadas” com base em informações provenientes de regiões de observações e os parâmetros das novas componentes são gerados das distribuições *a posteriori*;
- (iv) O parâmetro de regularização espacial β é tratado como parâmetro, permitindo maior flexibilidade na escolha do melhor modelo e atualizado *offline* através de um procedimento conhecido como integração termal, deixando eficiente o processo computacional;
- (v) O uso do algoritmo Swendsen-Wang aumenta a eficiência computacional, pois os *pixels* são realocados por grupos e não individualmente;

Ilustramos o desempenho do algoritmo EAM utilizando conjuntos de dados simulados e reais. Para cada conjunto de dados, apresentamos as estimativas da probabilidade *a posteriori* e a média ergódica de K ao longo das iterações. Também verificamos a capacidade do método em “visitar” diferentes valores de K ao longo das iterações e a estimativa da autocorrelação. Finalizamos a análise de cada conjunto de dados apresentando a figura identificada pelo método.

O restante da tese está organizada da seguinte forma: no Capítulo 2, apresentamos alguns conceitos sobre grafos, descrevemos o modelo de Potts (Potts, 1952) e o algoritmo Swendsen-Wang (Swendsen e Wang, 1987). O interesse no modelo de Potts e no algoritmo Swendsen-Wang é devido ao fato de que estes serão utilizados para o desenvolvimento da abordagem bayesiana e do algoritmo proposto. No Capítulo 3, desenvolvemos uma abordagem bayesiana hierárquica para segmentação de imagens e desenvolvemos o algoritmo EAM. No Capítulo 4, verificamos a performance do algoritmo EAM utilizando dados simulados e reais. No Capítulo 5, fazemos as considerações finais sobre o algoritmo proposto.

Capítulo 2

Modelo de Potts

Este capítulo é dividido em três seções. Na primeira seção apresentamos alguns conceitos sobre grafos. Estes conceitos são importantes para a descrição do modelo de Potts (Potts, 1952), descrito na segunda seção. Na terceira seção descrevemos o algoritmo Swendsen-Wang (Swendsen e Wang, 1987). Este algoritmo é um caso especial do algoritmo *slice sampler*, proposto por Neal (2003). O interesse no modelo de Potts e no algoritmo Swendsen-Wang é devido ao fato de que estes serão utilizados na abordagem bayesiana para segmentação de imagens com o número de componentes desconhecido, proposta no próximo capítulo.

2.1 Grafos

Em muitas situações é conveniente descrever um problema em estudo utilizando um diagrama com pontos e segmentos de reta ligando tais pontos. Por exemplo, em uma imagem, os *pixels* podem ser representados por pontos, enquanto segmentos de reta ligando os pontos podem representar os vínculos entre esses *pixels*. A abstração matemática de situações desse tipo dá origem ao conceito de grafo.

Definição 1. *Um grafo finito $G = (\mathbf{V}, \mathbf{A})$ consiste de um conjunto finito de vértices $\mathbf{V} = \{1, \dots, n\}$ e um conjunto de arestas \mathbf{A} , unindo pares não ordenados desses vértices. Se $i, j \in \mathbf{V}$ são dois vértices ligados por uma aresta, então a aresta $\{i, j\} \in \mathbf{A}$ é denotada por ij .*

A Figura 2.1 mostra o grafo $G_1 = (\mathbf{V}_1, \mathbf{A}_1)$, com $\mathbf{V}_1 = \{1, 2, 3, 4, 5, 6\}$ e $\mathbf{A}_1 = \{14, 24, 34, 45, 46\}$, e o grafo $G_2 = (\mathbf{V}_2, \mathbf{A}_2)$, com $\mathbf{V}_2 = \{1, 2\}$ e $\mathbf{A}_2 = \{12, 22\}$. Note que laços¹, como no grafo G_2 , são permitidos por definição.

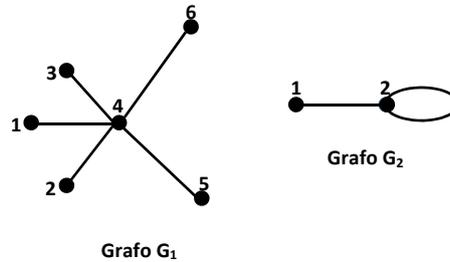


Figura 2.1: Exemplos de Grafos

Definição 2. *Dois vértices $i, j \in \mathbf{V}$ são ditos vizinhos (ou adjacentes) se existe uma aresta que os conecta. Esta situação é representada pela notação $i \sim j$.*

Por exemplo, no grafo G_1 da Figura 2.1, os vértices 1 e 4 são vizinhos, $1 \sim 4$, mas os vértices 2 e 5 não são vizinhos.

Definição 3. *O grau de um vértice i , denotado por $g(i)$, é o número de arestas que se conecta a esse vértice (i.e., o número de vizinhos).*

Definição 4. *O grau máximo de um grafo G , denotado por $\Delta(G)$, é definido por*

$$\Delta(G) = \max_{i \in \mathbf{V}} \{g(i)\}.$$

Definição 5. *O grau mínimo de um grafo G , denotado por $\delta(G)$, é definido por*

$$\delta(G) = \min_{i \in \mathbf{V}} \{g(i)\}.$$

Por exemplo, no grafo G_1 da Figura 2.1, $\Delta(G_1) = 5$ e $\delta(G_1) = 1$.

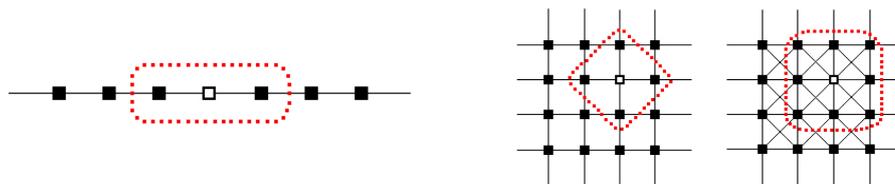
O conjunto de vértices $\{j \in \mathbf{V} : i \sim j\}$ é chamado de vizinhança de i e é denotado por $\mathcal{V}(i)$. Por exemplo, para o grafo G_1 da Figura 2.1, $\mathcal{V}(1) = \{4\}$ e $\mathcal{V}(4) = \{1, 2, 3, 5, 6\}$.

Definição 6. *Uma coleção $\mathbb{V} = \{\mathcal{V}(i) : i \in \mathbf{V}\}$ de subconjuntos de \mathbf{V} é chamada de sistema de vizinhanças se:*

¹Um laço (*loop*) é uma aresta que conecta um vértice a ele mesmo.

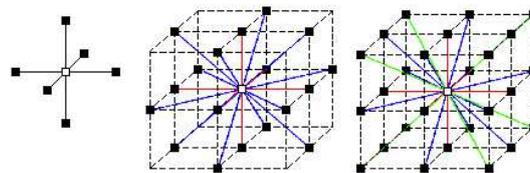
- (i) Em todo vértice i não há um laço, i.e., o vértice i não pertence ao seu conjunto de vizinhos, $i \notin \mathcal{V}(i)$;
- (ii) Quaisquer que sejam os vértices $i, j \in \mathbf{V}$, com $i \sim j$, a equivalência $i \in \mathcal{V}(j) \Leftrightarrow j \in \mathcal{V}(i)$ é verdadeira, i.e., o grafo $G = (\mathbf{V}, \mathbf{A})$ é não direcionado².

A Figura 2.2 mostra diferentes estruturas de sistemas de vizinhanças em uma, duas e três dimensões. A Figura 2.2(a) mostra uma estrutura unidimensional onde é possível construir apenas um sistema de vizinhanças com 2 vizinhos; a Figura 2.2(b) mostra uma estrutura bidimensional onde é possível construir sistemas de vizinhanças com 4 e 8 vizinhos; a Figura 2.2(c) mostra uma estrutura tridimensional onde é possível construir sistemas de vizinhanças com 6, 18 e 26 vizinhos.



(a) 1D

(b) 2D



(c) 3D

Figura 2.2: Exemplos de sistemas de vizinhanças

²Um grafo $G = (\mathbf{V}, \mathbf{A})$ é não direcionado se o conjunto \mathbf{A} é definido em termos de pares não ordenados de vértices, que são normalmente chamados de arestas.

Em algumas situações, é conveniente assumir que o grafo tem uma estrutura regular, *i.e.*, uma disposição geométrica igualmente espaçados de pontos sobre uma região plana ou espacial. Este tipo particular de grafo é denominado de grade. A Figura 2.3 mostra três tipos de grades regulares: retangular, triangular e hexagonal.

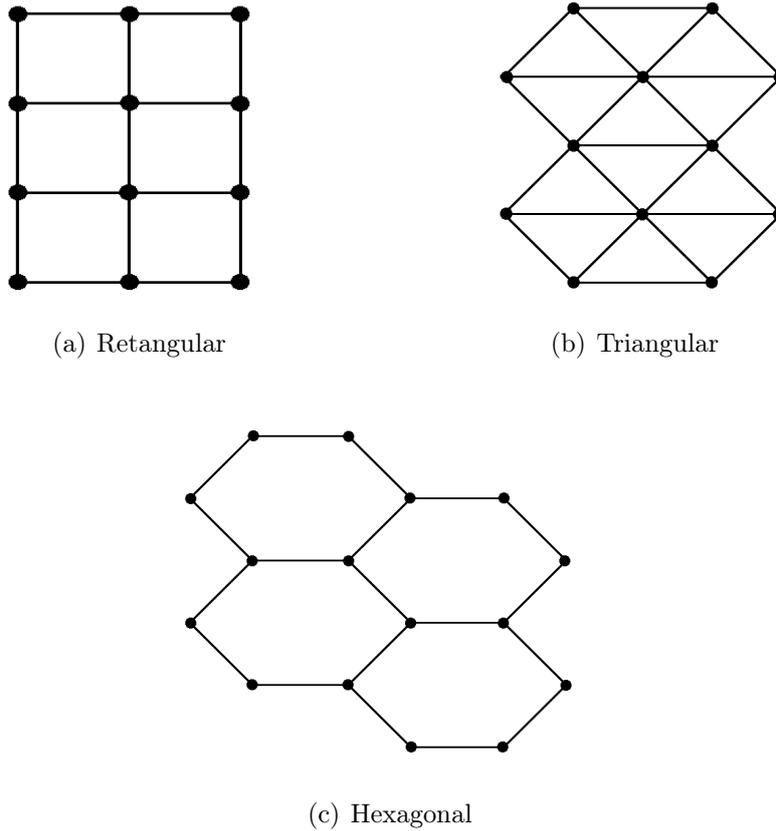


Figura 2.3: Exemplos de grades regulares

Para finalizar esta seção, considere uma grade regular $\mathcal{G} = (\mathbf{V}, \mathbf{A})$ do tipo retangular com $n_1 \times n_2$ vértices e um conjunto de arestas \mathbf{A} . Representando cada vértice desta grade por (p, q) , então o conjunto de vértices \mathbf{V} é dado por $\mathbf{V} = \{(p, q) : 1 \leq p \leq n_1, 1 \leq q \leq n_2\}$. Para simplificar a notação, considere que os vértices sejam numerados de 1 até $n = n_1 \times n_2$ com $i = p + (q - 1) \cdot n_1$ sendo o índice do vértice (p, q) , para $i = 1, \dots, n$. Esta notação será utilizada no Capítulo 3 para representação de uma imagem.

2.2 Modelo de Potts

O modelo de Potts (Potts, 1952) é um modelo matemático usado para estudar o comportamento de um grafo $G = (\mathbf{V}, \mathbf{A})$.

Para a descrição do modelo de Potts, considere, sem perda de generalidade, que o grafo $G = (\mathbf{V}, \mathbf{A})$ é uma grade regular $\mathcal{G} = (\mathbf{V}, \mathbf{A})$ de dimensão $n_1 \times n_2$.

Seja Λ um conjunto de características que cada um dos vértices da grade pode assumir. Por exemplo, em análise de imagens com cores branco e preto, $\Lambda = \{\text{branco, preto}\}$. Cada vértice da grade $\mathcal{G} = (\mathbf{V}, \mathbf{A})$ pode assumir apenas uma das duas cores. A Figura 2.4 mostra duas configurações para a grade $\mathcal{G} = (\mathbf{V}, \mathbf{A})$ com $\mathbf{V} = \{1, \dots, 16\}$ e $\Lambda = \{\text{branco, preto}\}$. O símbolo “o” representa um vértice de cor branca e o símbolo “●” representa um vértice de cor preta. A existência de uma aresta entre dois vértices é representada por um segmento de reta contínua e a não existência de uma aresta é representada por um segmento de reta tracejada. A existência de uma aresta entre dois vértices vizinhos indica que estes vértices são relacionados e assumem a mesma cor. Caso contrário, os dois vértices não são relacionados e assumem cores diferentes.

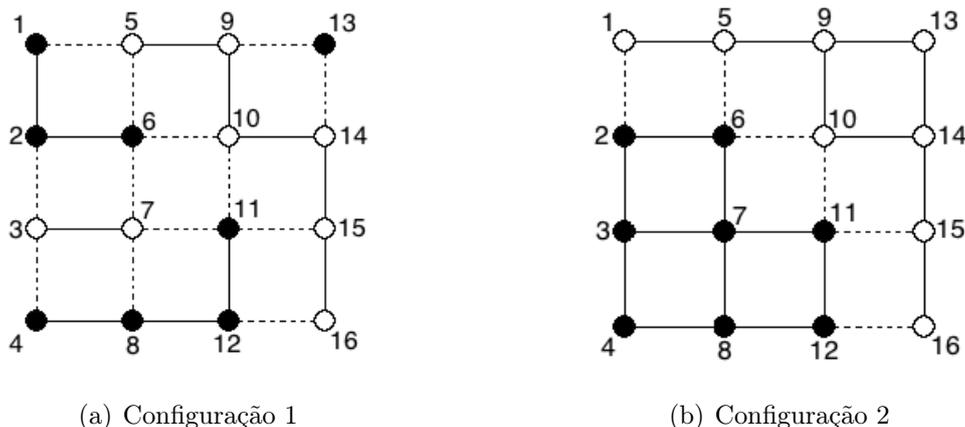


Figura 2.4: Configurações para a grade $\mathcal{G} = (\mathbf{V}, \mathbf{A})$ com $\Lambda = \{\text{branco, preto}\}$

Outros exemplos de conjuntos Λ são: carga elétrica ($\Lambda = \{\text{positivo, negativo}\}$), temperatura ($\Lambda = \{\text{quente, frio, gelado}\}$), direção ($\Lambda = \{\text{cima, baixo, esquerda, direita}\}$), cores ($\Lambda = \{\text{preto, azul, verde, vermelho, branco}\}$), entre outros.

Quando o conjunto Λ é composto por apenas duas características, temos o cha-

mado modelo de Ising (Ising, 1925).

O modelo desenvolvido por Potts é uma generalização do modelo Ising e considera que o conjunto Λ é composto por K características, para $K > 2$. Assim, de forma genérica, considere $\Lambda = \{1, \dots, K\}$, onde K é a quantidade total de características. Cada valor $k \in \Lambda$ é denominado rótulo (característica).

Para cada vértice $i \in \mathbf{V}$ da grade $\mathcal{G} = (\mathbf{V}, \mathbf{A})$, $i = p + (q - 1) \cdot n_1$, associamos uma variável indicadora S_i , de forma que $s_i = k$ indica que o vértice i assume o rótulo k , para $i = 1, \dots, n$ e $k = 1, \dots, K$.

Definição 7. *O estado $\mathbf{S} = (S_1, \dots, S_n)$ de uma grade $\mathcal{G} = (\mathbf{V}, \mathbf{A})$ é a escolha de um rótulo para cada um dos vértices da grade.*

O conjunto $\Omega_{\mathbf{s}}$ de todos os possíveis estados de uma grade $\mathcal{G} = (\mathbf{V}, \mathbf{A})$ é chamado de espaço de estados. O número total de configurações $\mathbf{s} \in \Omega_{\mathbf{s}}$ é $n(\Omega_{\mathbf{s}}) = K^n$.

Por exemplo, o estado \mathbf{s}' para a grade $\mathcal{G} = (\mathbf{V}, \mathbf{A})$ da Figura 2.4(a) é $\mathbf{s}' = (P, P, B, P, B, P, B, P, B, B, P, P, P, B, B, B)$, com B sendo branco e P preto. Para a Figura 2.4(b), o estado é $\mathbf{s}'' = (B, P, P, P, B, P, P, P, B, B, P, P, B, B, B, B)$. O número total de possíveis configurações é $n(\Omega_{\mathbf{s}}) = 2^{16} = 65.536$.

Cada estado $\mathbf{s} \in \Omega_{\mathbf{s}}$ particiona a grade $\mathcal{G} = (\mathbf{V}, \mathbf{A})$ em M regiões, R_1, \dots, R_M , formadas por conjuntos de vértices conectados, *i.e.*, para todo vértice $i, j \in R_m$, existe um conjunto de arestas conectando estes vértices, para $m = 1, \dots, M$. Por exemplo, o estado \mathbf{s}' da Figura 2.4(a), particiona a grade em $M = 5$ regiões, onde $R_1 = \{1, 2, 6\}$, $R_2 = \{3, 7\}$, $R_3 = \{4, 8, 11, 12\}$, $R_4 = \{5, 9, 10, 14, 15, 16\}$ e $R_5 = \{13\}$. O estado \mathbf{s}'' da Figura 2.4(b), particiona a grade em $M = 2$ regiões, onde $R_1 = \{1, 5, 9, 10, 13, 14, 15, 16\}$ e $R_2 = \{2, 3, 4, 6, 7, 8, 11, 12\}$.

Como diferentes estados de uma grade $\mathcal{G} = (\mathbf{V}, \mathbf{A})$ definem diferentes rótulos para os elementos da grade, então é importante ter uma medida que forneça o grau de relacionamento total de um estado \mathbf{S} . A função que mede este grau de relacionamento é o hamiltoniano. A definição do hamiltoniano dada a seguir é uma forma simplificada, obtida sob a hipótese de que o campo magnético externo é nulo. Para maiores detalhes, ver Potts (1952) e Feng and Tierney (2011).

Definição 8. O hamiltoniano é uma função $H : \Omega_{\mathbf{s}} \rightarrow \mathbb{R}$, onde um estado $\mathbf{S} = (S_1, \dots, S_n)$ da grade $\mathcal{G} = (\mathbf{V}, \mathbf{A})$ é transformado em um valor real da seguinte forma

$$H(\mathbf{s}) = -\beta \sum_{i \sim j} \mathbb{I}_{s_i}(s_j),$$

em que β é uma constante que determina a força da interação entre os vértices i e j vizinhos, $\mathbf{s} = \{s_1, \dots, s_n\}$ e $\mathbb{I}_{s_i}(s_j) = 1$ se $s_i = s_j$ e $\mathbb{I}_{s_i}(s_j) = 0$ caso contrário, quaisquer que sejam $i, j \in \mathbf{V}$.

O parâmetro β é um parâmetro de regularização espacial que controla a dependência espacial entre os vértices vizinhos. Quando $\beta = 0$, os vértices são independentes. Por outro lado, quanto maior o valor de β , maior a dependência entre os vértices vizinhos.

Para ilustrar o cálculo do hamiltoniano, considere as grades da Figura 2.4 (pag.12). Esta figura apresenta os estados \mathbf{s}' e \mathbf{s}'' de uma grade $\mathcal{G} = (\mathbf{V}, \mathbf{A})$, de dimensão $n = 4 \times 4 = 16$, para um conjunto de características $\Lambda = \{\text{branco, preto}\}$. Assuma que a constante β de interação do sistema seja igual a 1. A existência de uma relação entre os vértices i e j , $i \sim j$, é representada por um segmento de reta contínuo com o valor da aresta ij igual a 1. A não existência de uma relação é representada por um segmento de reta tracejado com o valor da aresta ij sendo 0. Assim, o valor do hamiltoniano para o estado \mathbf{s}' , Figura 2.4(a), é $H(\mathbf{s}') = -11$. O hamiltoniano do estado \mathbf{s}'' , Figura 2.4(b), é $H(\mathbf{s}'') = -18$.

Como o valor de $H(\cdot)$ diminui à medida que a quantidade de ligações existente entre os vértices aumenta, então quanto menor é o valor de $H(\cdot)$, mais forte é a estrutura de relacionamento do estado \mathbf{S} . Para o exemplo dado acima, $H(\mathbf{s}'') < H(\mathbf{s}')$, indicando que o estado \mathbf{s}'' apresenta uma estrutura de relacionamento mais forte do que o estado \mathbf{s}' . Visualmente, isto significa que a partição da grade $\mathcal{G} = (\mathbf{V}, \mathbf{A})$ induzida pelo estado \mathbf{s}'' é mais organizada do que a partição induzida pelo estado \mathbf{s}' , como pode ser observado na Figura 2.4.

Dada uma grade $\mathcal{G} = (\mathbf{V}, \mathbf{A})$, um conjunto de características Λ e o hamiltoniano $H(\mathbf{S})$ para um estado \mathbf{S} , é interessante transformar o grau de relacionamento deste estado em uma medida que seja interpretada como quão provável é o estado \mathbf{S} .

Uma distribuição de probabilidades que preserva a dependência entre os vértices vizinhos da grade $\mathcal{G} = (\mathbf{V}, \mathbf{A})$ é a distribuição de Gibbs, dada por

$$\pi(\mathbf{s}|\beta, K) = \frac{\exp \left\{ \beta \sum_{i \sim j} \mathbb{I}_{s_i}(s_j) \right\}}{\sum_{\mathbf{s} \in \Omega_{\mathbf{s}}} \exp \left\{ \beta \sum_{i \sim j} \mathbb{I}_{s_i}(s_j) \right\}}. \quad (2.1)$$

O denominador na Equação (2.1) é a constante de normalização, também chamada de função partição e é denotada por $g(\beta, K)$, $g(\beta, K) = \sum_{\mathbf{s} \in \Omega_{\mathbf{s}}} \exp \left\{ \beta \sum_{i \sim j} \mathbb{I}_{s_i}(s_j) \right\}$. O cálculo de $g(\beta, K)$ pode ser computacionalmente intratável, pois sua soma possui K^n parcelas. Por exemplo, para uma grade $\mathcal{G} = (\mathbf{V}, \mathbf{A})$ de dimensão 9×9 com $K = 3$, existem $K^n = 3^{81} = 4,434265 \times 10^{38}$ parcelas a serem somadas. Alguns autores, como Chalmond (1989) e Won e Derin (1992), evitam o cálculo dessa constante de normalização usando uma pseudo-verossimilhança como uma aproximação de $\pi(\mathbf{S} = \mathbf{s}|\beta, K)$, que é dada por

$$\tilde{\pi}(\mathbf{s}|\beta, K) = \prod_{i=1}^n \frac{\exp \left\{ \beta \sum_{i \sim j} \mathbb{I}_{s_i}(s_j) \right\}}{\sum_{k=1}^K \exp \left\{ \beta \sum_{i \sim j} \mathbb{I}_{s_i}(s_j) \right\}}.$$

Porém, segundo Barker *et al.* (1997), sob certas circunstâncias, esta pseudo-verossimilhança pode fazer com que a distribuição *a posteriori* seja imprópria. Gelman e Meng (1998), Green e Richardson (2002) e Zhang *et al.* (2007) usam uma aproximação para $\pi(\mathbf{S} = \mathbf{s}|\beta, K)$ obtida via método da integração termal, proposto por Ogata (1989). Os detalhes deste método de integração são apresentados no Apêndice A₁.

2.3 Algoritmo Swendsen-Wang

Nesta seção, apresentamos o algoritmo Swendsen-Wang (Swendsen e Wang, 1987). Este algoritmo é um caso especial do algoritmo *slice sampler*, proposto por Neal (2003). Devido a isto, iniciamos esta seção descrevendo o *slice sampler* caso simples; em seguida descrevemos o *slice sampler* caso geral e finalizamos a seção descrevendo o algoritmo Swendsen-Wang.

2.3.1 *Slice sampler* caso simples

Suponha que desejamos obter amostras de uma distribuição de probabilidade de uma variável aleatória Y , definida em um subconjunto $\mathcal{X} \subset \mathbb{R}$. Considere que a densidade de Y , $\pi(y)$, seja proporcional a uma função $f(y)$, *i.e.*, $\pi(y) = \frac{f(y)}{\eta}$, onde η é a constante normalizadora.

Como a distribuição de Y não é totalmente conhecida, então para gerar valores de Y , considere o seguinte procedimento. Seja U uma variável auxiliar com distribuição uniforme no intervalo $(0, f(y))$, $U \sim \mathcal{U}(0, f(y))$. Assim, a distribuição conjunta de Y e U é dada pela distribuição uniforme na região bidimensional $\mathbb{U}_{yu} = \{(y, u) : 0 < u < f(y)\}$ abaixo da curva de $f(y)$.

A densidade conjunta de (Y, U) é dada por

$$\pi(y, u) = \begin{cases} 1/\eta, & \text{se } 0 < u < f(y) \\ 0, & \text{caso contrário} \end{cases},$$

onde $\eta = \int f(y) dy$. Para maiores detalhes dos cálculos, ver Apêndice A_2 .

A densidade marginal de Y é dada por

$$\pi(y) = \int_0^{f(y)} \pi(y, u) du = \int_0^{f(y)} 1/\eta du = f(y)/\eta,$$

como desejado.

Para gerar os valores y de interesse, geramos valores (y, u) da distribuição conjunta $\pi(y, u)$ e ignoramos os valores u gerados. Porém, gerar valores da distribuição conjunta de (Y, U) pode não ser simples. Uma alternativa é desenvolver um algoritmo MCMC que tenha como distribuição de equilíbrio esta distribuição uniforme. Por exemplo, podemos usar o algoritmo *Gibbs sampling* gerando valores iterativamente da distribuição condicional de $U|y$, que é dada pela distribuição uniforme no intervalo $(0, f(y))$; e da distribuição condicional de $Y|u$, que é dada pela distribuição uniforme na região $\mathbb{S} = \{y : f(y) > u\}$. Neal (2003) denomina \mathbb{S} de *slice* definido por y .

- **Algoritmo 1: *Slice sampler* caso simples.** Sejam Y uma variável aleatória, com densidade $\pi(y)$ proporcional a uma função $f(y)$, e U uma variável auxiliar,

com distribuição $\mathcal{U}(0, f(y))$. Inicie o algoritmo selecionando aleatoriamente um valor $y^{(1)}$ pertencente ao domínio da função $f(y)$. Para a l -ésima iteração, $l = 2, \dots, L$, proceda da seguinte maneira:

(1) gere $u^{(l)}$ da distribuição condicional

$$U|y^{(l-1)} \sim \mathcal{U}(0, f(y^{(l-1)}));$$

(2) gere $y^{(l)}$ da distribuição condicional

$$Y|u^{(l)} \sim \mathcal{U}(\mathbb{S}^{(l)}),$$

onde $\mathbb{S}^{(l)} = \{y : f(y) > u^{(l)}\}$.

Ou seja, o método *slice sampler* gera valores de U e Y utilizando movimentos uniformes alternados em direção aos eixos das ordenadas e das abscissas, respectivamente. A Figura 2.5 ilustra esses movimentos, para a função $f(\cdot)$ unidimensional. As semirretas na vertical, abaixo da curva, representam os movimentos na direção u , e as semirretas na horizontal representam os movimentos na direção y .

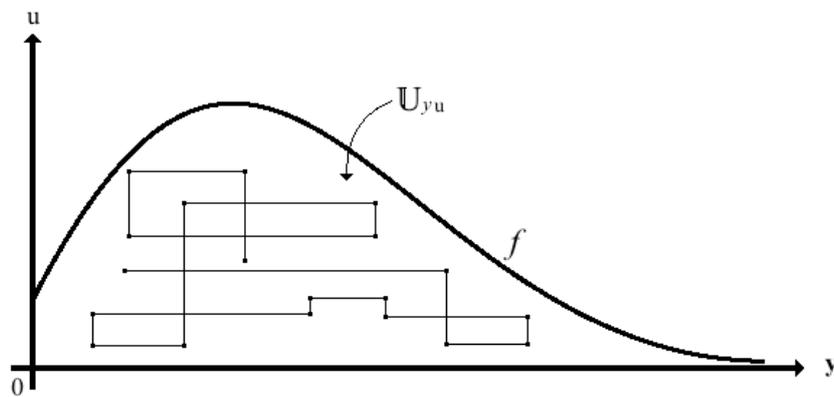


Figura 2.5: Movimentos no *slice sampler*

A seguir apresentamos dois exemplos em que utilizamos o *slice sampler* para gerar valores da variável Y desejada.

Exemplo 2.1. Considere uma variável aleatória Y , com densidade $f(y) = \frac{1}{2} \exp\{-y^{\frac{1}{2}}\}$ para $y > 0$. Utilizando o *slice sampler*, como descrito acima, iniciamos o algoritmo selecionando aleatoriamente um valor $y^{(1)}$ pertencente a \mathbb{R}^+ . Para a l -ésima iteração, $l = 2, \dots, L$, geramos valores $u^{(l)}$ e $y^{(l)}$ das seguintes distribuições condicionais

$$U|y^{(l-1)} \sim \mathcal{U}\left(0, \frac{1}{2} \exp\left\{-\left(y^{(l-1)}\right)^{\frac{1}{2}}\right\}\right) \quad \text{e} \quad Y|u^{(l)} \sim \mathcal{U}(\mathbb{S}^{(l)}),$$

respectivamente, onde $\mathbb{S}^{(l)} = \left\{y : \frac{1}{2} \exp\left\{-y^{\frac{1}{2}}\right\} > u^{(l)}\right\}$. Como a função $f(\cdot)$ possui inversa podemos reescrever o conjunto $\mathbb{S}^{(l)}$ utilizando a $f^{-1}(\cdot)$, *i.e.*, $\mathbb{S}^{(l)} = \{y : y < 2 \log(2u^{(l)})\}$, onde $u^{(l)} > 0$. Então, a distribuição de $Y|u^{(l)}$, é dada por

$$Y|u^{(l)} \sim \mathcal{U}(0, 2 \log(2u^{(l)})).$$

A Figura 2.6(a) mostra os gráficos dos valores gerados para as variáveis U e Y com $L = 5.000$ iterações. A Figura 2.6(b) mostra o histograma dos valores de Y gerados, sobreposto pelo gráfico da função $f(y)$ proporcional a densidade alvo.

Exemplo 2.2. Neste exemplo, nosso interesse é gerar valores da distribuição normal, com média $\mu = 3$ e variância $\sigma^2 = 1$, truncada no intervalo $(0, 1)$. Assim, considere Y a variável de interesse com densidade

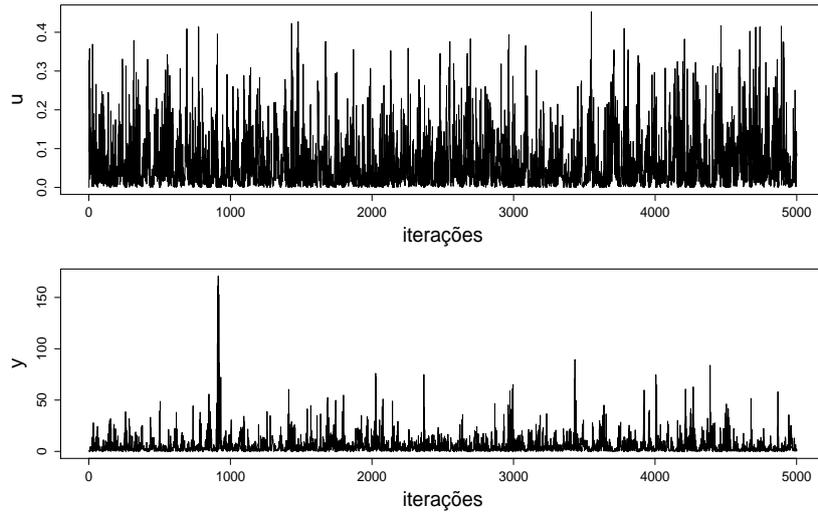
$$\pi(y) \propto f(y) = \exp\left\{-\frac{(y-3)^2}{2}\right\} \mathbb{I}_{[0,1]}(y)$$

e U a variável auxiliar com distribuição uniforme no intervalo $\left(e^{-\frac{9}{2}}, f(y)\right)$, pois como y é truncado no intervalo $(0, 1)$ o menor valor que u pode assumir é $e^{-\frac{9}{2}}$. Iniciamos selecionando aleatoriamente $y^{(1)} \in (0, 1)$. Para a l -ésima iteração, $l = 2, \dots, L$, geramos valores $u^{(l)}$ e $y^{(l)}$ das seguintes distribuições condicionais

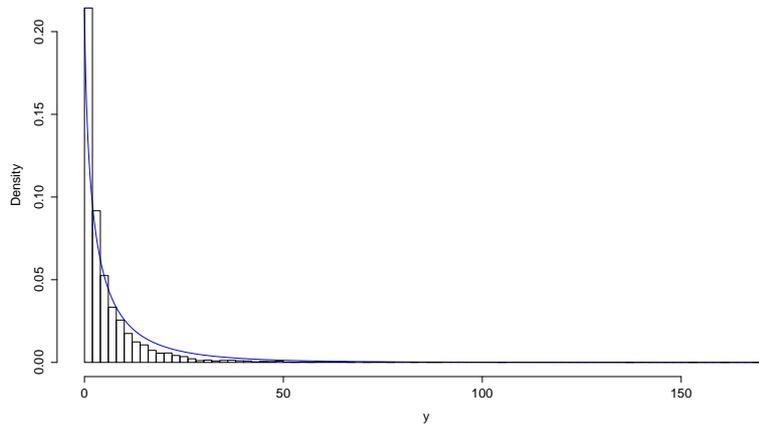
$$U|y^{(l-1)} \sim \mathcal{U}\left(e^{-\frac{9}{2}}, \exp\left\{-\frac{(y^{(l-1)}-3)^2}{2}\right\} \mathbb{I}_{[0,1]}(y^{(l-1)})\right)$$

e

$$Y|u^{(l)} \sim \mathcal{U}(\mathbb{S}^{(l)}),$$



(a) 5.000 valores gerados das variáveis u (superior) e Y (inferior)



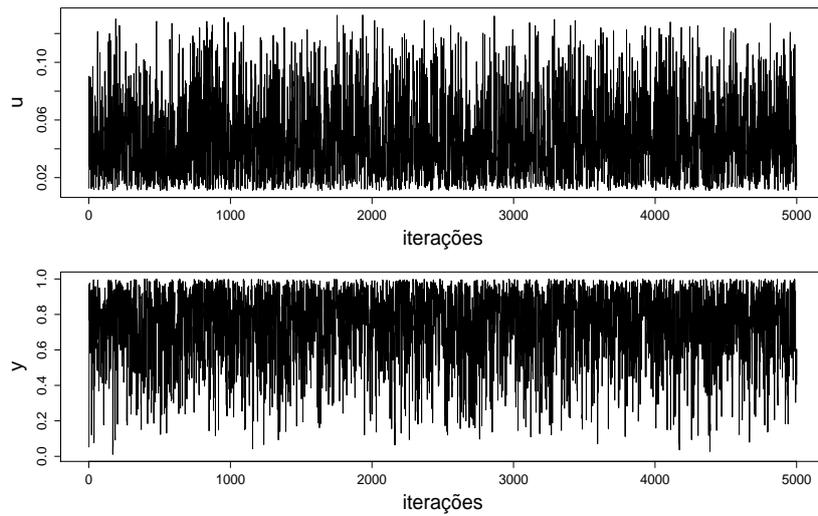
(b) Histograma dos valores de Y gerados, sobreposto pela função $f(y)$

Figura 2.6: *Slice sampler* para $\pi(y) \propto \frac{1}{2} \exp \left\{ -y^{\frac{1}{2}} \right\}$, $y > 0$

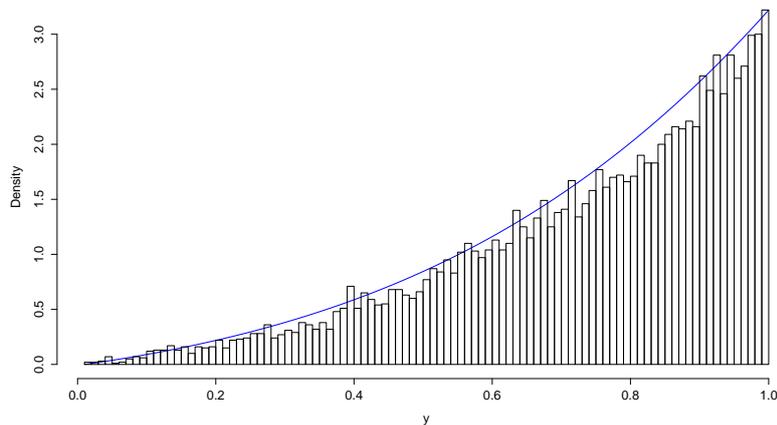
onde $\mathbb{S}^{(l)} = \left\{ y : y < 3 - (-2 \log(u^{(l)}))^{\frac{1}{2}} \right\}$ e $u^{(l)} > e^{-\frac{9}{2}}$. Então, a distribuição de $Y|u^{(l)}$ é dada por

$$Y|u^{(l)} \sim \mathcal{U} \left(0, 3 - (-2 \log(u^{(l)}))^{\frac{1}{2}} \right)$$

A Figura 2.7(a) mostra os gráficos dos valores gerados de u (superior) e de Y (inferior) para $L = 5.000$ iterações. A Figura 2.7(b) mostra o histograma dos valores de Y gerados, sobreposto pelo gráfico da função proporcional à função alvo.



(a) 5.000 valores gerados das variáveis u (superior) e Y (inferior)



(b) Histograma dos valores de Y gerados, sobreposto pela função $f(y)$

Figura 2.7: *Slice sampler* para a distribuição normal truncada em $(0, 1)$

2.3.2 *Slice sampler* caso geral

Considere agora que a função densidade (ou função de probabilidade) $\pi(y)$ seja proporcional ao produto de K funções, isto é,

$$\pi(y) \propto f_1(y)f_2(y) \dots f_K(y). \quad (2.2)$$

Para este caso, temos a generalização do *slice sampler*, proposto por Edwards e Sokal (1988). Nesta generalização, os autores introduzem K variáveis auxiliares U_1, \dots, U_K e definem a distribuição conjunta de (Y, U_1, \dots, U_K) , que é dada pela

distribuição uniforme no hiper-retângulo

$$\mathbb{S} = \{(u_1, \dots, u_K) : 0 < u_k < f_k(y), \forall k = 1, \dots, K\}.$$

Dessa forma, repetindo o que foi feito para o *slice sampler* caso simples, uma amostra da distribuição conjunta de (Y, U_1, \dots, U_K) pode ser obtida utilizando o algoritmo *Gibbs sampling*.

- **Algoritmo 2: *Slice sampler* caso geral.** Considere uma variável aleatória Y com densidade $\pi(y)$, como em (2.2), e as variáveis auxiliares U_1, \dots, U_K , com distribuição uniforme $\mathcal{U}(0, f_k(y))$. Sejam $\mathcal{D}_1, \dots, \mathcal{D}_K$ o domínio das funções $f_1(y), \dots, f_K(y)$, respectivamente. Inicie o algoritmo selecionando aleatoriamente $y^{(1)} \in \bigcap_{k=1}^K \mathcal{D}_k$. Para a l -ésima iteração, $l = 2, \dots, L$, proceda da seguinte maneira:

- (1) gere $u_k^{(l)}$ da distribuição condicional

$$U_k | y^{(l-1)} \sim \mathcal{U}(0, f_k(y^{(l-1)}))$$

para $k = 1, \dots, K$. Assim, obtemos o ponto $(u_1^{(l)}, \dots, u_K^{(l)})$;

- (2) gere $y^{(l)}$ da distribuição condicional

$$Y | u_1^{(l)}, \dots, u_K^{(l)} \sim \mathcal{U}(\mathbb{S}^{(l)}),$$

onde $\mathbb{S}^{(l)} = \{y : f_k(y) \geq u_k^{(l)}, \forall k = 1, \dots, K\}$.

Abaixo apresentamos dois exemplos, em que geramos valores da variável de interesse utilizando o *slice sampler* caso geral. No primeiro exemplo, geramos valores de uma distribuição gama truncada, e no segundo de uma distribuição proporcional ao produto de três funções.

Exemplo 2.3. Considere uma variável aleatória Y com distribuição Gama de parâmetros α e β , truncada no intervalo (a, b) , ou seja, $\pi(y) \propto y^{\alpha-1} e^{-\beta y} \mathbb{I}_{(a,b)}(y)$, com $\alpha > 1$, $\beta > 0$, $a, b \in \mathbb{R}$ e $a < b$. Sejam, $f_1(y) = y^{\alpha-1}$ e $f_2(y) = e^{-\beta y}$.

Inicie o *slice sampler*, selecionando aleatoriamente $y^{(1)} \in (a, b)$. Para a l -ésima iteração, $l = 2, \dots, L$, proceda da seguinte maneira:

(1) gere $u_1^{(l)}$ e $u_2^{(l)}$ das distribuições condicionais

$$U_1|y^{(l-1)} \sim \mathcal{U}(0, f_1(y^{(l-1)})) \quad \text{e} \quad U_2|y^{(l-1)} \sim \mathcal{U}(0, f_2(y^{(l-1)})),$$

respectivamente. Assim, obtemos o ponto $(u_1^{(l)}, u_2^{(l)})$;

(2) gere $y^{(l)}$ da distribuição condicional

$$Y|u_1^{(l)}, u_2^{(l)} \sim \mathcal{U}(\mathbb{S}^{(l)}),$$

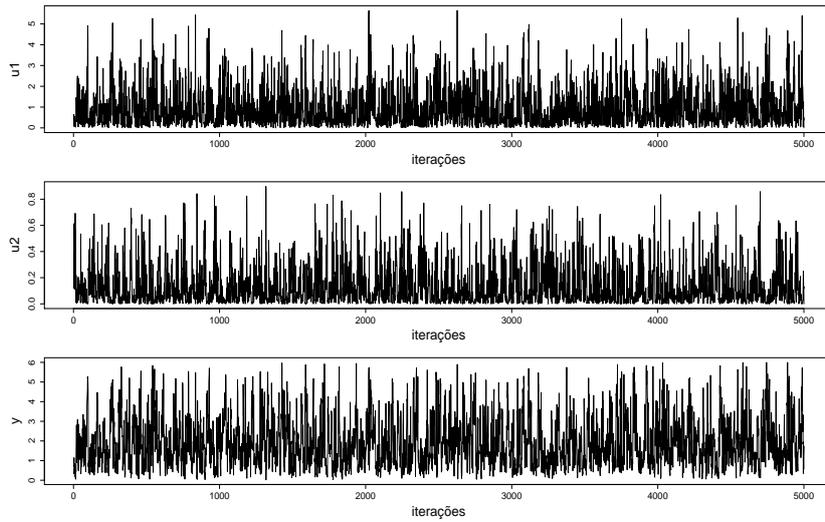
onde $\mathbb{S}^{(l)} = \{y : f_1(y) \geq u_1^{(l)}, f_2(y) \geq u_2^{(l)}\} \cap (a, b)$. Como as funções $f_1(\cdot)$ e $f_2(\cdot)$ possuem inversa podemos reescrever o conjunto $\mathbb{S}^{(l)}$ da seguinte forma, $\mathbb{S}^{(l)} = \left\{y : y \geq \left(u_1^{(l)}\right)^{\frac{1}{\alpha-1}}, y \leq -\frac{1}{\beta} \log\left(u_2^{(l)}\right)\right\} \cap (a, b)$, onde $u_1^{(l)} \geq 0$ e $u_2^{(l)} > 0$. Para este caso, após algumas manipulações algébricas (ver Apêndice A₃), obtemos $\mathbb{S} = (a', b')$, onde

$$a' = \max\left\{a, \left(u_1^{(l)}\right)^{\frac{1}{\alpha-1}}\right\} \quad \text{e} \quad b' = \min\left\{b, -\frac{1}{\beta} \log\left(u_2^{(l)}\right)\right\}.$$

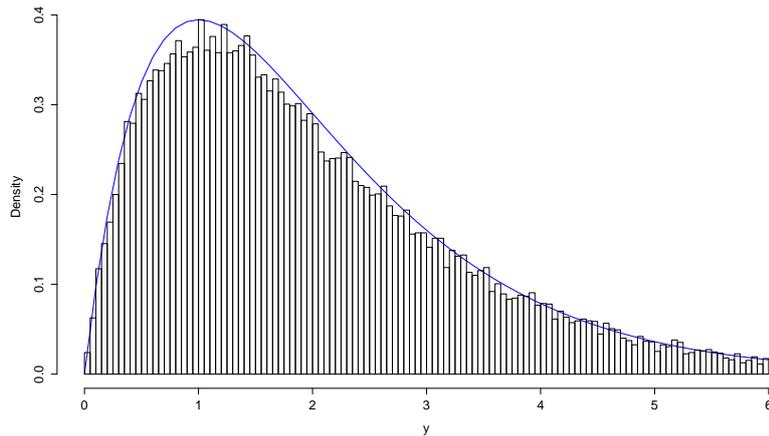
Portanto,

$$Y|u_1^{(l)}, u_2^{(l)} \sim \mathcal{U}(a', b')$$

A Figura 2.8(a) apresenta os gráficos dos valores gerados de u_1 (superior), u_2 (centro) e de Y (inferior) para $L = 5.000$ iterações, da distribuição Gama truncada no intervalo $(0, 6)$. A Figura 2.8(b) mostra histograma dos valores Y gerados.



(a) 5.000 valores gerados das variáveis u_1 (superior), u_2 (centro) e Y (inferior)



(b) Histograma dos valores de Y gerados, sobreposto pela função $f(y)$

Figura 2.8: *Slice sampler* para distribuição gama truncada no intervalo $(1, 6)$

Exemplo 2.4. Considere uma variável aleatória Y com função densidade de probabilidade $\pi(y) \propto f(y) = (1 + \sin^2(3y))(1 + \cos^4(5y)) \exp\left\{-\frac{y^2}{2}\right\}$. Note que, a função $f(y)$ pode ser decomposta no produto das funções $f_1(y) = 1 + \sin^2(3y)$, $f_2(y) = 1 + \cos^4(5y)$ e $f_3(y) = \exp\left\{-\frac{y^2}{2}\right\}$.

Vamos gerar valores da variável Y utilizando o *slice sampler*. Para isto, inicie o algoritmo selecionando aleatoriamente $y^{(1)} \in \mathbb{R}$. Para a l -ésima iteração, $l = 2, \dots, L$, proceda da seguinte maneira:

(1) gere $u_1^{(l)}$, $u_2^{(l)}$ e $u_3^{(l)}$ das seguintes distribuições uniformes condicionais

$$\begin{aligned} U_1|y^{(l-1)} &\sim \mathcal{U}(1, 1 + \sin^2(3y^{(l-1)})), \\ U_2|y^{(l-1)} &\sim \mathcal{U}(1, 1 + \cos^4(5y^{(l-1)})), \\ U_3|y^{(l-1)} &\sim \mathcal{U}\left(0, \exp\left\{-\frac{(y^{(l-1)})^2}{2}\right\}\right). \end{aligned}$$

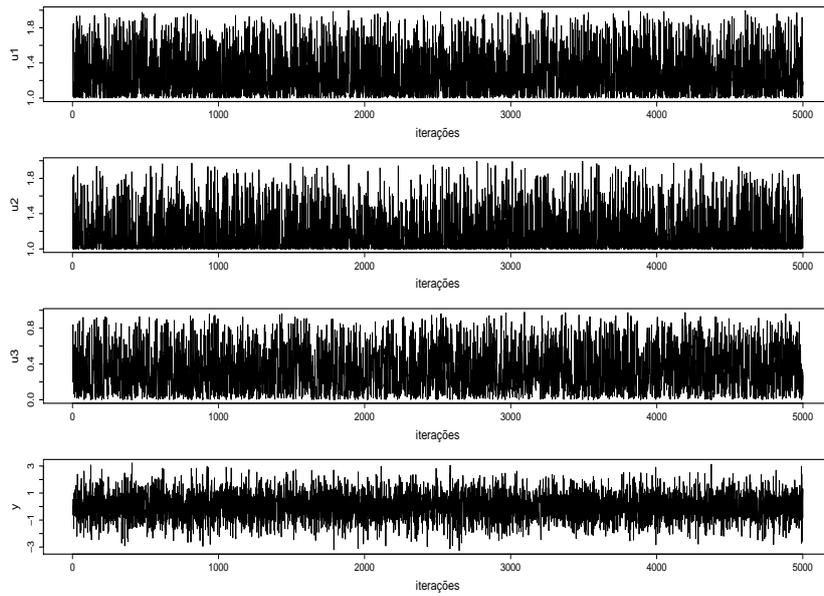
Assim, obtemos o ponto $(u_1^{(l)}, u_2^{(l)}, u_3^{(l)})$, onde $u_1 \geq 1$, $u_2 \geq 1$ e $0 < u_3 \leq 1$.

(2) gere $y^{(l)}$ da distribuição condicional $Y|u_1^{(l)}, u_2^{(l)}, u_3^{(l)} \sim \mathcal{U}(\mathbb{S}^{(l)})$, onde $\mathbb{S}^{(l)} = \{y : f_1(y) \geq u_1^{(l)}, f_2(y) \geq u_2^{(l)}, f_3(y) \geq u_3^{(l)}\}$. Calculando a inversa das funções $f_1(\cdot)$, $f_2(\cdot)$ e $f_3(\cdot)$, podemos reescrever o conjunto $\mathbb{S}^{(l)}$ como

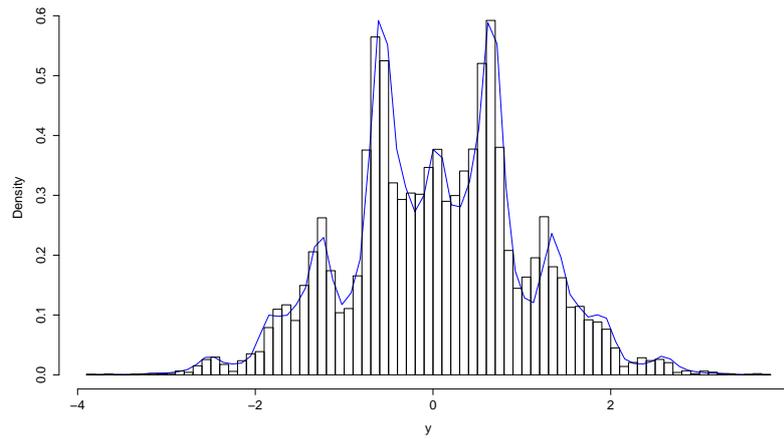
$$\begin{aligned} \mathbb{S}^{(l)} = &\left\{y : |\sin(3y)| \geq (u_1^{(l)} - 1)^{\frac{1}{2}}\right\} \\ &\cap \left\{y : |\cos(5y)| \geq (u_2^{(l)} - 1)^{\frac{1}{4}}\right\} \\ &\cap \left\{y : |y| \leq \left(-2 \log(u_3^{(l)})\right)^{\frac{1}{2}}\right\}. \end{aligned}$$

Os cálculos de podem ser encontrados no Apêndice A4. Para maiores detalhes ver Robert e Casela (2004).

Na Figura 2.9(a) apresentamos os gráficos dos valores gerados de u_1 , u_2 , u_3 e de Y para $L = 5.000$ iterações. Na Figura 2.9(b), mostramos o histograma dos valores de Y gerados, sobreposto pelo gráfico da função $f(y)$ proporcional a densidade alvo.



(a) 5.000 valores gerados das variáveis u_1 , u_2 , u_3 e Y



(b) Histograma dos valores de Y gerados, sobreposto pela função $f(y)$

Figura 2.9: *Slice sampler* para $\pi(y) \propto (1 + \sin^2(3y)) (1 + \cos^4(5y)) \exp\left\{-\frac{y^2}{2}\right\}$

2.3.3 Algoritmo Swendsen-Wang

Suponha que temos interesse em gerar valores da distribuição de Gibbs

$$\pi(\mathbf{S} = \mathbf{s} | \beta, K) = g^{-1}(\beta, K) \exp \left\{ \beta \sum_{i \sim j} \mathbb{I}_{s_i}(s_j) \right\},$$

para $\mathbf{s} \in \Omega_{\mathbf{s}}$, onde

$$g(\beta, K) = \sum_{\mathbf{s} \in \Omega_{\mathbf{s}}} \exp \left\{ \beta \sum_{i \sim j} \mathbb{I}_{s_i}(s_j) \right\}$$

é a constante de normalização, também chamada de função partição.

Para simular valores de \mathbf{S} com distribuição $\pi(\mathbf{S} = \mathbf{s} | \beta, K)$, vamos utilizar o método *slice sampler*, como descrito na seção anterior. Para isto, considere $\mathbf{u} = \{u_{ij} : i, j = 1, \dots, n \text{ e } i \sim j\}$ um conjunto de variáveis auxiliares e assumamos que $u_{ij} \sim \mathcal{U}(0, \exp\{\beta \mathbb{I}_{s_i}(s_j)\})$, para todo $i, j \in \{1, \dots, n\}$ e $i \neq j$, onde $\mathcal{U}(A)$ é a distribuição uniforme contínua no intervalo A . Ou seja, dado \mathbf{s} , as variáveis u_{ij} são mutuamente independentes com distribuição uniforme, *i.e.*,

$$\pi(\mathbf{u} | \mathbf{s}) = \prod_{i \sim j} \exp\{-\beta \mathbb{I}_{s_i}(s_j)\} \mathbb{I}(0 \leq u_{ij} \leq \exp\{\beta \mathbb{I}_{s_i}(s_j)\}).$$

A distribuição conjunta de (\mathbf{S}, \mathbf{u}) é dada por

$$\pi(\mathbf{S}, \mathbf{u}) = \pi(\mathbf{S}) \pi(\mathbf{u} | \mathbf{s}) \propto \exp \left\{ \beta \sum_{i \sim j} \mathbb{I}_{s_i}(s_j) \right\} \mathbb{I}_{u_{ij}} \left(0, \exp \left\{ \beta \sum_{i \sim j} \mathbb{I}_{s_i}(s_j) \right\} \right).$$

Note que, como β é um parâmetro positivo, $\beta > 0$, então ao gerar u_{ij} da distribuição uniforme $\mathcal{U}(0, \exp\{\beta \mathbb{I}_{s_i}(s_j)\})$, temos duas possibilidades: (i) se $\mathbb{I}_{s_i}(s_j) = 0$, então $u_{ij} \sim \mathcal{U}(0, 1)$; (ii) caso contrário, $\mathbb{I}_{s_i}(s_j) = 1$, e $u_{ij} \sim \mathcal{U}(0, e^\beta)$. Quando o valor $u_{ij} \in [1, e^\beta]$, os vértices i e j assumem o mesmo rótulo na configuração atual \mathbf{s} , $s_i = s_j$ com $i \sim j$, e estes dois vértices estão conectados por uma aresta.

Assim, dado que os vértices i e j assumem o mesmo rótulo na configuração atual \mathbf{s} , então a probabilidade de $u_{ij} \in [1, e^\beta]$ é $P(u_{ij} > 1 | \mathbf{s}) = 1 - e^{-\beta}$. Este fato pode ser facilmente verificado integrando a densidade no intervalo considerado, *i.e.*,

$$\begin{aligned} P(u_{ij} > 1 | \mathbf{s}) &= \int_1^{\exp\{\beta \mathbb{I}_{s_i}(s_j)\}} \frac{1}{\exp\{\beta \mathbb{I}_{s_i}(s_j)\}} du_{ij} \\ &= \frac{\exp\{\beta \mathbb{I}_{s_i}(s_j)\} - 1}{\exp\{\beta \mathbb{I}_{s_i}(s_j)\}} \\ &= 1 - \exp\{-\beta \mathbb{I}_{s_i}(s_j)\}. \end{aligned} \tag{2.3}$$

Ou seja, dado que $s_i = s_j$ no estado atual \mathbf{s} , então conectamos os vértices i e j com probabilidade $1 - e^{-\beta}$. Este procedimento de ligação entre os vértices i

e j vizinhos, $i \sim j$, define M regiões, R_1, \dots, R_M , de vértices conectados, para $i, j = 1, \dots, n$, $i \neq j$ e $i \sim j$.

Para atualizar $\mathbf{S}|\mathbf{u}$, proceda da seguinte maneira:

- (1) A partir da configuração atual \mathbf{s} , obtenha uma configuração de ligações \mathbf{B} , que é um conjunto de “arestas” conectando vértices vizinhos. Esta transição, $\mathbf{s} \rightarrow \mathbf{B}$, é obtida da seguinte forma: para todo $i \sim j$, se $s_i = s_j$, então “crie” uma aresta entre os vértices i e j com probabilidade $1 - e^{-\beta}$. Caso contrário, nada é feito;
- (2) Converta o resultado da configuração de ligações \mathbf{B} em uma nova configuração \mathbf{s}' como segue:
 - (i) identifique as regiões R_1, \dots, R_M de vértices conectados através de arestas;
 - (ii) para cada região R_m estabeleça um rótulo k com probabilidade $1/K$, para $k \in \{1, \dots, K\}$;
 - (iii) se R_m é rotulado como k , então faça $s_i = k$ para todo vértice $i \in R_m$, para $i = 1, \dots, n$, $m = 1, \dots, M$, $k = 1, \dots, K$ e $K \leq M$. Faça $\mathbf{s} = \mathbf{s}'$.

Ou seja, o algoritmo Swendsen-Wang particiona estocasticamente a configuração \mathbf{S} em regiões, de forma que elas possam ser atualizadas de forma independente.

Capítulo 3

Modelo bayesiano para segmentação de imagens

Neste capítulo, apresentamos uma abordagem bayesiana para segmentação de imagens em situações em que o número de componentes é desconhecido. Para fazer inferências em relação aos parâmetros de interesse, propomos um novo algoritmo MCMC. O algoritmo proposto é composto de dois tipos de movimentos: movimentos que não mudam o número de componentes e movimentos que mudam o número de componentes na vizinhança $K \pm 1$. Os movimentos que mudam o número K de componentes nas vizinhanças $K + 1$ e $K - 1$ são denominados de ejeção e absorção, respectivamente. Estes movimentos são realizados fundamentados em um procedimento de alocação probabilística baseado nos dados observados. A probabilidade de aceitação é dada pela probabilidade de aceitação Metropolis-Hastings.

3.1 Modelo bayesiano

Seja $\mathcal{G} = (\mathbf{V}, \mathbf{A})$ uma grade regular com $n_1 \times n_2$ vértices e um conjunto \mathbf{A} de arestas. Denotando cada vértice desta grade por (p, q) , então o conjunto \mathbf{V} de vértices é dado por $\mathbf{V} = \{(p, q) : 1 \leq p \leq n_1, 1 \leq q \leq n_2\}$. A Figura 3.1(a) mostra um exemplo de grade \mathcal{G} com $n_1 = n_2 = 8$, e todos os vértices sendo conectados. A Figura 3.1(b) mostra a mesma grade \mathcal{G} com alguns vértices não sendo conectados. Além disso,

assuma que esta grade apresente uma estrutura de vizinhança com quatro vizinhos, com exceção dos vértices das bordas com três vizinhos e dos vértices dos cantos com dois vizinhos, como mostrado na Figura 3.1(c).

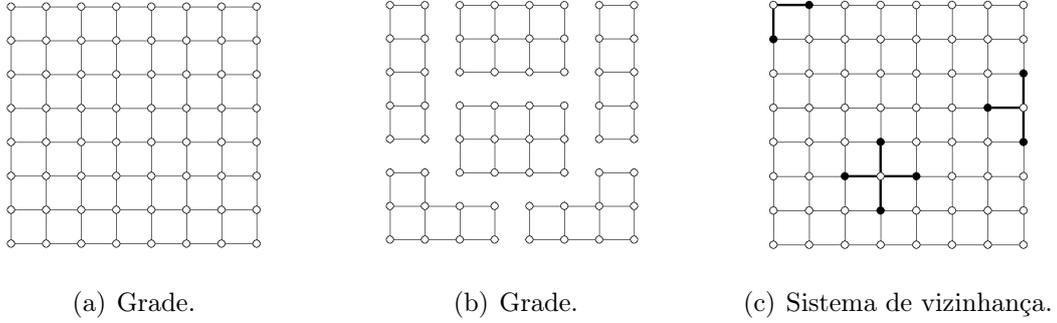


Figura 3.1: Exemplos de grades e sistema de vizinhança assumido

Considere $\mathbf{y} = (y_1, \dots, y_n)$ uma imagem definida em \mathcal{G} , onde $n = n_1 \cdot n_2$ é o número de *pixels* da imagem, *i.e.*, número de vértices $(p, q) \in \mathcal{G}$. Para fazer a ligação entre os elementos do vetor \mathbf{y} com os vértices da grade \mathcal{G} , considere $i = p + (q - 1) \cdot n_1$ o índice do *pixel* representado pelo vértice (p, q) , sendo y_i a intensidade ou nível de cinza observado do *pixel* i , para $i = 1, \dots, n$. Seja K a quantidade de tons de cinza (ou cores) da imagem.

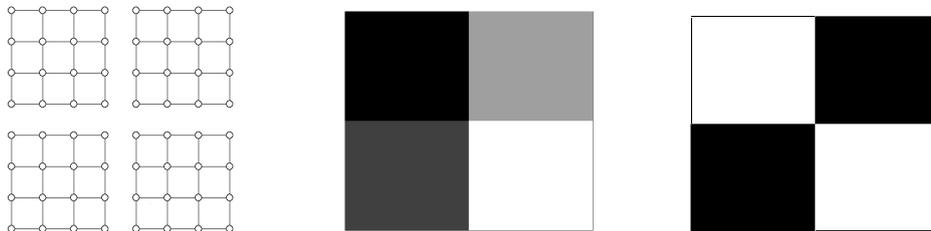
Assuma que y_i é uma realização de uma variável aleatória Y_i e que a imagem observada, representada pela grade \mathcal{G} , seja composta por M regiões disjuntas R_1, \dots, R_M , *i.e.*,

$$\mathcal{G} = \bigcup_{m=1}^M R_m \quad \text{e} \quad R_m \cap R_{m'} = \emptyset,$$

para $m, m' \in \{1, \dots, M\}$ e $m \neq m'$. Cada região R_m é formada por um conjunto de *pixels* conectados, *i.e.*, para quaisquer $y_i, y_j \in R_m$, existe um conjunto de arestas conectando y_i e y_j . Por outro lado, se $y_i \in R_m$ e $y_j \in R_{m'}$, para $m \neq m'$, então não existe um conjunto de arestas conectando os *pixels* i e j . Nesta representação, a existência de uma aresta conectando os *pixels* vizinhos, i e j , $i \sim j$, indica que estes *pixels* assumem a mesma cor.

A Figura 3.2(a) mostra a representação \mathcal{G} de uma imagem com quatro regiões. Na Figura 3.2(b) temos o exemplo de uma imagem que pode ser representada por \mathcal{G} da Figura 3.2(a). Nesta figura, temos $M = 4$ regiões, com cada região assumindo

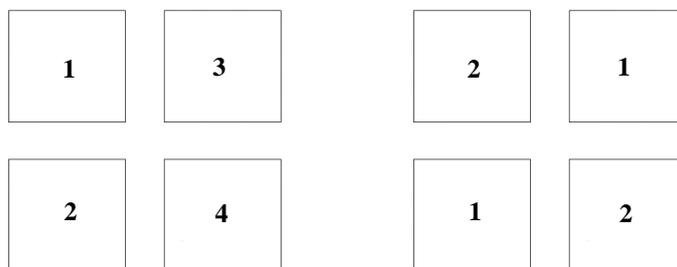
uma cor diferente, *i.e.*, $K = 4$. Na Figura 3.2(c) temos as mesmas quatro regiões, porém apenas duas cores, *i.e.*, $K = 2$. Logo, temos $M \geq K$ e diferentes regiões podem assumir a mesma cor.



(a) Grade com 4 regiões. (b) Fig. com $M = 4$ e $K = 4$. (c) Fig. com $M = 4$ e $K = 2$.

Figura 3.2: Grade com $M = 4$ e exemplos de imagens com $K = 4$ e $K = 2$

Suponha que as K possíveis cores da imagem sejam representadas pelos elementos do conjunto $\Lambda = \{1, \dots, K\}$. Por exemplo, na Figura 3.2(b), as cores preto, cinza escuro, cinza claro e branco são representadas pelo conjunto $\Lambda = \{1, 2, 3, 4\}$, respectivamente; e as cores da Figura 3.2(c), preto e branco, pelo conjunto $\Lambda = \{1, 2\}$, respectivamente. A Figura 3.3 mostra um exemplo de rotulação das regiões das imagens da Figura 3.2(b,c).



(a) Rotulação, $M = 4$ e $K = 4$. (b) Rotulação, $M = 4$ e $K = 2$.

Figura 3.3: Exemplo de rotulação das regiões da Figura 3.2(b,c), respectivamente

Assim, considere que cada região R_m assume um rótulo k , para $k \in \Lambda$ e $K \leq M$. Se R_m é rotulada como k , então associe a cada $y_i \in R_m$ uma variável indicadora não observável $s_i = k$, para $i = 1, \dots, n$, $k = 1, \dots, K$ e $m = 1, \dots, M$. A partir deste ponto considere o termo componente para denotar o conjunto de todas as observações y_i que possuem o mesmo valor de s_i . Denote a componente k por D_k , onde $D_k = \{y_i : s_i = k\}$, para $k \in \Lambda$ e $i = 1, \dots, n$. Note que uma componente

k pode ser composta por observações pertencentes a uma ou mais regiões disjuntas R_m com mesmo rótulo k , para $k = 1, \dots, K$, $m = 1, \dots, M$ e $K \leq M$. Seja $n_k = |\{i : s_i = k\}|$ o número de observações pertencentes a k -ésima componente, para $i = 1, \dots, n$, $k = 1, \dots, K$ e $\sum_{k=1}^K n_k = n$.

Considere que as variáveis não observáveis $\mathbf{S} = (S_1, \dots, S_n)$ apresentem uma estrutura de dependência como descrito na Figura 3.1(c). Condicional em \mathbf{s} , assumamos que as variáveis aleatórias Y_1, \dots, Y_n sejam independentes. A Figura 3.4 ilustra o modelo utilizado. Na grade \mathcal{G} , cada símbolo “o” representa um *pixel* i , que está associado a uma variável não observável S_i . A dependência entre as variáveis não observáveis é representada pelas arestas conectando os *pixels*. A relação entre S_i e Y_i é representada pela flecha de S_i para Y_i . A independência condicional entre os Y_i 's é representada pela não ligação entre os símbolos “□”.

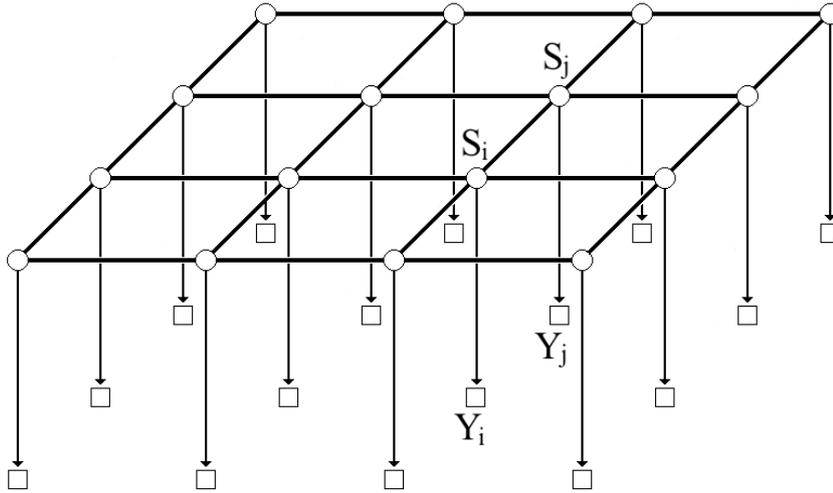


Figura 3.4: Representação do modelo utilizado

Assim, dados K e $s_i = k$, assumamos que

$$Y_i = \mu_k + \varepsilon_k, \quad \text{com} \quad \varepsilon_k \sim \mathcal{N}(0, \sigma_k^2), \quad (3.1)$$

onde μ_k representa a verdadeira intensidade do tom de cinza k (não observado), ε_k é o ruído e $\mathcal{N}(0, \sigma_k^2)$ representa a distribuição normal com média zero e variância σ_k^2 , para $i = 1, \dots, n$ e $k = 1, \dots, K$. O ruído pode ser proveniente do processo de medição das intensidades dos *pixels* que está sujeito à precisão do equipamento,

temperatura ambiente, entre outros (López-Rubio, 2009). A Figura 3.5 mostra a representação do modelo em (3.1)

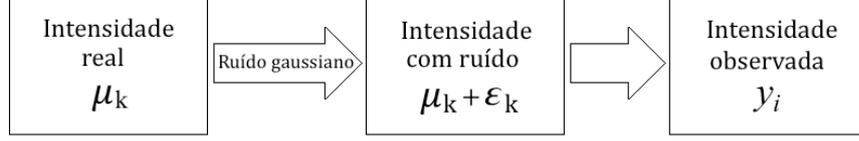


Figura 3.5: Representação do modelo utilizado

Portanto, de (3.1),

$$Y_i | K, s_i = k, \mu_k, \sigma_k^2 \sim \mathcal{N}(\mu_k, \sigma_k^2), \quad (3.2)$$

para $i = 1, \dots, n$ e $k = 1, \dots, K$. Denote por $\boldsymbol{\theta}_K = (\theta_1, \dots, \theta_k)$ todos os parâmetros do modelo, onde $\theta_k = (\mu_k, \sigma_k^2)$, para $k = 1, \dots, K$. A suposição de normalidade é usual em análise de imagens, ver por exemplo, Richardson e Green (1997); Zhang *et al.* (2007); Ferreira (2009); Zhang *et al.* (2010); Askari *et al.* (2013) e outros.

Condicional em $\mathbf{s} = (s_1, \dots, s_n)$, a densidade de y_i é dada por

$$f(y_i | \theta_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp \left\{ -\frac{1}{2\sigma_k^2} (y_i - \mu_k)^2 \right\}, \quad (3.3)$$

para $i = 1, \dots, n$ e $k = 1, \dots, K$. Logo,

$$P(\mathbf{Y} = \mathbf{y} | \mathbf{s}, \boldsymbol{\theta}_K, K) = \prod_{i=1}^n \prod_{k=1}^K f(y_i | s_i = k, \boldsymbol{\theta}_K)^{\mathbb{I}_{s_i}(k)},$$

onde $\mathbb{I}_{s_i}(k) = 1$, se $s_i = k$ e $\mathbb{I}_{s_i}(k) = 0$, caso contrário. Como $f(y_i | s_i = k, \boldsymbol{\theta}_K) = f(y_i | \theta_k)$, então

$$P(\mathbf{Y} = \mathbf{y} | \mathbf{s}, \boldsymbol{\theta}_K, K) = \prod_{i=1}^n \prod_{k=1}^K f(y_i | \theta_k)^{\mathbb{I}_{s_i}(k)} = \prod_{k=1}^K \left(\prod_{y_i \in D_k} f(y_i | \theta_k) \right) \quad (3.4)$$

A probabilidade conjunta dos dados observados \mathbf{y} e das variáveis indicadoras \mathbf{S} , condicional nos parâmetros $\boldsymbol{\theta}_K$ e no número de componentes K , é dada por

$$\begin{aligned} P(\mathbf{y}, \mathbf{s} | \boldsymbol{\theta}_K, K) &= P(\mathbf{Y} = \mathbf{y} | \mathbf{s}, \boldsymbol{\theta}_K, K) \pi(\mathbf{S} = \mathbf{s} | K) \\ &= \left[\prod_{k=1}^K \left(\prod_{y_i \in D_k} f(y_i | \theta_k) \right) \right] \pi(\mathbf{S} = \mathbf{s} | K). \end{aligned} \quad (3.5)$$

3.1.1 Distribuições *a priori*

Para completar o modelo bayesiano que está sendo proposto, consideramos que a distribuição *a priori* conjunta para $(\boldsymbol{\theta}_K, \mathbf{S}, K)$ é condicionalmente independente, *i.e.*,

$$\pi(\boldsymbol{\theta}_K, \mathbf{S}, K) = \pi(\boldsymbol{\theta}_K|K)\pi(\mathbf{S}|K)\pi(K).$$

Assumindo que não há informação *a priori* disponível sobre K , uma escolha natural como distribuição *a priori* é a distribuição uniforme discreta $\mathcal{U}(1, K_m)$, onde K_m é um valor previamente fixado que indica o valor máximo que K pode assumir (Richardson e Green, 1997; Bhattacharya, 2008). Uma outra possibilidade seria considerar uma distribuição *a priori* Poisson de parâmetro λ_K truncada no intervalo $[1, K_m]$, $K \sim \mathcal{Poi}(\lambda_K)$, como descrito em Zhang *et al.* (2007); Johnson e Piert (2009). Para o modelo que está sendo proposto, assumimos $K \sim U(1, K_m)$. Porém, isto não impede que o algoritmo MCMC, que será proposto na próxima seção, seja utilizado com outras distribuições *a priori* para K .

Para a variável indicadora não observável \mathbf{S} , consideramos a distribuição *a priori* de Gibbs com parâmetros β e K ,

$$\pi(\mathbf{S} = \mathbf{s}|\beta, K) = g^{-1}(\beta, K) \exp \left\{ \beta \sum_{i \sim j} \mathbb{I}_{s_i}(s_j) \right\}, \quad (3.6)$$

onde $\mathbb{I}_{s_i}(s_j) = 1$, se $s_i = s_j$ e $\mathbb{I}_{s_i}(s_j) = 0$, caso contrário; $i \sim j$ indica que os *pixels* i e j são vizinhos e $g^{-1}(\cdot)$ é a constante de normalização. O parâmetro β é um parâmetro de regularização espacial que controla a dependência entre os *pixels* vizinhos. Quando $\beta = 0$, os *pixels* são independentes. Por outro lado, quanto maior é o valor de β , maior será a dependência entre os *pixels* vizinhos. Neste ponto, considere mais um nível hierárquico, assumindo para β uma distribuição *a priori* Gama de parâmetros a e b , $\beta \sim \Gamma(a, b)$, com $a, b > 0$, e uma parametrização cuja média é a/b e cuja variância é a/b^2 .

Para o parâmetro $\theta_k = (\mu_k, \sigma_k^2)$, considere as seguintes distribuições *a priori*,

$$\mu_k | \sigma_k^2, \mu_0, \lambda \sim \mathcal{N} \left(\mu_0, \frac{\sigma_k^2}{\lambda} \right) \quad \text{e} \quad \sigma_k^2 | \alpha, \gamma \sim \mathcal{IG} \left(\frac{\alpha}{2}, \frac{\gamma}{2} \right), \quad (3.7)$$

onde μ_0, λ, α e γ são hiperparâmetros conhecidos.

De forma resumida, temos as seguintes distribuições *a priori*:

$$\begin{aligned}\mu_k | \mu_0, \lambda, \sigma_k^2 &\sim \mathcal{N}\left(\mu_0, \frac{\sigma_k^2}{\lambda}\right) \\ \sigma_k^2 | \alpha, \gamma &\sim \mathcal{IG}\left(\frac{\alpha}{2}, \frac{\gamma}{2}\right) \\ \beta | a, b &\sim \Gamma(a, b) \\ K | K_m &\sim \mathcal{U}(1, K_m),\end{aligned}$$

onde $\mu_0, \lambda, \alpha, \gamma, a, b$ e K_m são hiperparâmetros conhecidos. A Figura 3.6 mostra um grafo, chamado de *directed acyclic graph* (DAG), representando a estrutura hierárquica do modelo bayesiano proposto, onde as constantes são colocados dentro dos retângulos, as variáveis dentro das elipses e $\theta' = (\mu_0, \lambda, \alpha, \gamma)$. A escolha dos valores para os hiperparâmetros, geralmente, depende do caso sendo analisado e das informações disponibilizadas pelos especialistas da área. Neste ponto, não atribuímos valores aos hiperparâmetros, e o desenvolvimento do método é feito supondo que sejam valores conhecidos. No Capítulo 4 discutimos uma forma de especificação destes valores.

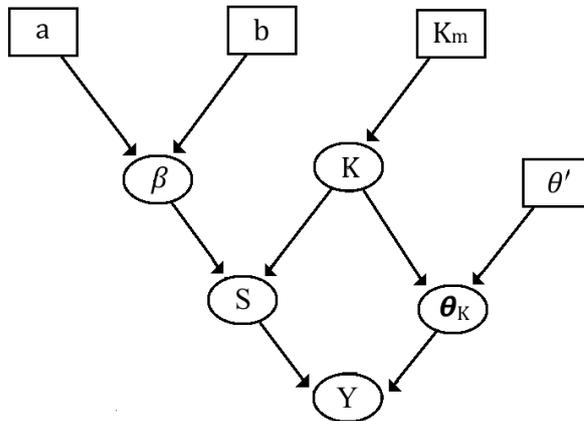


Figura 3.6: DAG do modelo bayesiano proposto

3.1.2 Distribuições *a posteriori*

Utilizando o teorema de Bayes, a distribuição *a posteriori* conjunta para $(\boldsymbol{\theta}_K, \mathbf{S}, \beta, K)$ é dada por

$$\pi(\boldsymbol{\theta}_K, \mathbf{S}, \beta, K | \mathbf{y}) \propto P(\mathbf{y} | \mathbf{s}, \boldsymbol{\theta}_K, K) \pi(\mathbf{S} = \mathbf{s} | \beta, K) \pi(\boldsymbol{\theta}_K | K) \pi(\beta) \pi(K). \quad (3.8)$$

Dado K , as distribuições *a posteriori* condicionais para μ_k , σ_k^2 , β e \mathbf{S} são dadas, respectivamente, por

$$\sigma_k^2 | \alpha, \gamma, \lambda, \mathbf{s}, K, \mathbf{y} \sim \mathcal{IG} \left(\frac{\alpha + n_k + 1}{2}, \frac{\gamma}{2} + \frac{\sum_{y_i \in D_k} y_i^2 + \lambda \mu_0^2}{2} - \frac{\left(\sum_{y_i \in D_k} y_i + \lambda \mu_0 \right)^2}{2(n_k + \lambda)} \right) \quad (3.9)$$

$$\mu_k | \mu_0, \sigma_k^2, \lambda, K, \mathbf{y} \sim \mathcal{N} \left(\frac{\sum_{y_i \in D_k} y_i + \lambda \mu_0}{n_k + \lambda}, \frac{\sigma_k^2}{n_k + \lambda} \right) \quad (3.10)$$

$$\pi(\beta | \mathbf{s}, a, b, K) \propto g^{-1}(\beta, K) \exp \left\{ \beta \sum_{i \sim j} \mathbb{I}_{s_i}(s_j) \right\} \beta^{a-1} \exp\{-b\beta\} \quad (3.11)$$

$$\pi(\mathbf{S} = \mathbf{s} | \mathbf{y}, \boldsymbol{\theta}_K, \beta, K) \propto \exp \left\{ \beta \sum_{i \sim j} \mathbb{I}_{s_i}(s_j) \right\} \left[\prod_{k=1}^K \left(\prod_{y_i \in D_k} f(y_i | \theta_k) \right) \right], \quad (3.12)$$

para $i = 1, \dots, n$ e $k = 1, \dots, K$. Os cálculos para obtenção destas distribuições condicionais são apresentados no Apêndice A₅.

Porém, a distribuição condicional para K , $\pi(K | \mathbf{s}, \boldsymbol{\theta}_K, \beta, \mathbf{y})$, não é analiticamente conhecida e um procedimento de amostragem de K não é dado de forma simples e direta. Devido a isto, desenvolvemos um procedimento de atualização de K , utilizando movimentos denominados ejeção e absorção. Estes movimentos propõem mudanças no número K de componentes nas vizinhanças $K + 1$ e $K - 1$, respectivamente, e são aceitos de acordo com a probabilidade de aceitação Metropolis-Hastings.

Neste ponto, algumas considerações sobre a rotulação das componentes são necessárias. Note que qualquer permutação dos rótulos $\{1, \dots, K\}$ produz a mesma função de verossimilhança. Esta invariância da função de verossimilhança em relação aos rótulos das componentes é conhecida na literatura pelo termo *label switching* (Celeux *et al.*, 2000; Stephens, 2000b). Na abordagem bayesiana, se não há informações *a priori* para distinguir as K componentes, então a distribuição *a priori* será

a mesma para toda permutação de $\boldsymbol{\theta}_K$, e a distribuição *a posteriori* também será simétrica. Esta simetria pode causar problemas quando tentamos estimar quantidades de uma componente individual, pois não conseguimos distinguir a componente de interesse das demais. Além disso, as probabilidades de alocação são inúteis para o agrupamento das observações (Stephens, 2000a). Entretanto, um modelo que seja identificável pode ser obtido assumindo uma restrição sobre os parâmetros do modelo. Dessa forma, para desenvolver o algoritmo proposto, adotamos uma escolha tradicional e impomos restrições sobre as médias das componentes. Ou seja, assumimos que $\mu_1 < \mu_2 < \dots < \mu_K$. O algoritmo proposto pode ser aplicado considerando-se outras restrições. Para uma discussão mais ampla sobre *label switching*, ver Celeux *et al.* (2000); Stephens (2000b); Frühwirth-Schnatter (2001); Jasra *et al.* (2005).

3.2 Algoritmo ejeção-absorção metropolizado

Para gerar amostras da distribuição *a posteriori* conjunta em (3.8), utilizando as distribuições condicionais em (3.9) a (3.12) e os movimentos ejeção e absorção, propomos o algoritmo MCMC denominado de ejeção-absorção metropolizado (EAM). Este algoritmo, primeiramente, atualiza $\boldsymbol{\theta}_K$, β e \mathbf{S} condicionado a K usando os algoritmos *Gibbs sampling*, Metropolis-Hastings e Swendsen-Wang, respectivamente. Em seguida, o número de componentes K é atualizado via movimentos ejeção e absorção. Para manter válida a equação de balanceamento em relação à distribuição *a posteriori* quando propomos os movimentos ejeção e absorção, a probabilidade de aceitação para estes movimentos são calculadas de acordo com procedimento proposto por Carlin e Chib (1995); Godsill (2001); Dellaportas *et al.* (2002).

3.2.1 Atualizações condicionadas em K

Dado K , os parâmetros das componentes, $\boldsymbol{\theta}_K = (\theta_1, \dots, \theta_K)$, onde $\theta_k = (\mu_k, \sigma_k^2)$ para $k = 1, \dots, K$, são atualizados via algoritmo *Gibbs sampling*, como descrito a seguir.

Algoritmo Gibbs sampling. Considere uma cadeia de Markov consistindo de $(\boldsymbol{\theta}_K, \mathbf{S}, \beta, K)$. Condicional em (\mathbf{s}, K) , atualize $\theta_k = (\mu_k, \sigma_k^2)$ gerando valores das distribuições condicionais (3.9) e (3.10), para $k = 1, \dots, K$.

O parâmetro de regularização espacial β é atualizado via algoritmo Metropolis-Hastings utilizando como distribuição geradora de novos candidatos a distribuição *a priori* Gama de parâmetros a e b , $\Gamma(a, b)$. Assim, a probabilidade de aceitação para o novo valor β^* , com $\beta^* \sim \Gamma(a, b)$, é dada por $\Psi(\beta^*|\beta) = \min\{1, A_\beta\}$, onde

$$A_\beta = \frac{g^{-1}(\beta^*, K) \exp\left\{\beta^* \sum_{i \sim j} \mathbb{I}_{s_i}(s_j)\right\}}{g^{-1}(\beta, K) \exp\left\{\beta \sum_{i \sim j} \mathbb{I}_{s_i}(s_j)\right\}}. \quad (3.13)$$

Para estimar a constante de normalização $g^{-1}(\cdot)$, usamos o método da integração termodinâmica, proposto por Ogata (1989). Detalhes dos cálculos são apresentados no Apêndice A₁. O algoritmo Metropolis-Hastings é dado como segue:

Algoritmo Metropolis-Hastings. Considere uma cadeia de Markov consistindo de $(\boldsymbol{\theta}_K, \mathbf{S}, \beta, K)$. Gere $\beta^* \sim \Gamma(a, b)$ e aceite este novo valor com probabilidade $\Psi(\beta^*|\beta) = \min(1, A_\beta)$, onde A_β é dado em (3.13). Caso contrário, mantenha o valor atual β na próxima iteração.

As variáveis indicadoras não observáveis \mathbf{S} são atualizadas via algoritmo Swendsen-Wang. Como descrito na Seção 2.3.3 do Capítulo 2, este algoritmo utiliza variáveis auxiliares \mathbf{u} para simplificar a identificação das regiões e construir uma distribuição conjunta $\pi(\mathbf{S}, \mathbf{u})$ de forma que a distribuição marginal de \mathbf{S} seja dada por (3.6). As variáveis \mathbf{u} devem ser especificadas de forma que seja simples obter amostras das distribuições condicionais $\pi(\mathbf{u}|\mathbf{s})$ e $\pi(\mathbf{S}|\mathbf{u})$. Assim, para obter uma amostra da distribuição conjunta $\pi(\mathbf{S}, \mathbf{u})$, basta fazer iterações entre $\pi(\mathbf{u}|\mathbf{s})$ e $\pi(\mathbf{S}|\mathbf{u})$.

Assim, seja $\mathbf{u} = \{u_{ij} : i, j = 1, \dots, n \text{ e } i \sim j\}$ um conjunto de variáveis auxiliares e assumamos que $u_{ij} \sim \mathcal{U}(0, \exp\{\beta \mathbb{I}_{s_i}(s_j)\})$. A distribuição condicional de \mathbf{u} dado \mathbf{s} é

$$\pi(\mathbf{u}|\mathbf{s}) = \prod_{i \sim j} \exp\{-\beta \mathbb{I}_{s_i}(s_j)\} \mathbb{I}(0 \leq u_{ij} \leq \exp\{\beta \mathbb{I}_{s_i}(s_j)\}).$$

A distribuição condicional de \mathbf{S} dados \mathbf{u} e todos os outros parâmetros é dada por

$$\pi(\mathbf{S}|\mathbf{u}, \mathbf{y}, \boldsymbol{\theta}_K, \beta, K) \propto \left[\prod_{k=1}^K \left(\prod_{y_i \in D_k} f(y_i|\theta_k) \right) \right] \prod_{i \sim j} \mathbb{I}(0 \leq u_{ij} \leq \exp\{\beta \mathbb{I}_{s_i}(s_j)\}). \quad (3.14)$$

Na atualização $\mathbf{u}|\mathbf{s}$, dado que os vértices i e j assumem o mesmo rótulo na configuração atual \mathbf{s} , então eles são conectados com probabilidade $1 - e^{-\beta}$ (ver expressão (2.3), pag.25). Este procedimento de ligação entre os *pixels* i e j vizinhos, $i \sim j$, define M regiões, R_1, \dots, R_M , de *pixels* conectados, para $i = 1, \dots, n$ e $j = 1, \dots, k$.

Como uma componente D_k é composta por uma ou mais regiões, então podemos reescrever (3.14), em termos das regiões, da seguinte forma

$$\pi(\mathbf{S}|\mathbf{u}, \mathbf{y}, \boldsymbol{\theta}_K, \beta, K) \propto \left[\prod_{m=1}^M \left(\prod_{y_i \in R_m} f(y_i|\theta_k) \right) \right] \prod_{i \sim j} \mathbb{I}(0 \leq u_{ij} \leq \exp\{\beta \mathbb{I}_{s_i}(s_j)\}).$$

Dessa forma, podemos atualizar cada região R_m de forma independente. Logo, na atualização $\mathbf{S}|\mathbf{u}, \mathbf{y}, \boldsymbol{\theta}_K, \beta, K$, é atribuído um novo rótulo k a cada região R_m com probabilidade

$$\pi(R_m = k|\mathbf{u}, \mathbf{y}, \boldsymbol{\theta}_K, \beta, K) \propto \prod_{y_i \in R_m} \sigma_k^{-1} \exp \left\{ -\frac{1}{2\sigma_k^2} (y_i - \mu_k)^2 \right\}, \quad (3.15)$$

para $m = 1, \dots, M$, $k \in \{1, \dots, K\}$ e $K \leq M$. A rotulação das regiões define a nova configuração \mathbf{s} .

O algoritmo Swendsen-Wang pode ser sintetizado da seguinte forma:

Algoritmo Swendsen-Wang. Considere uma cadeia de Markov consistindo de $(\boldsymbol{\theta}_K, \mathbf{S}, \beta, K)$. Dado $(\boldsymbol{\theta}_K, \beta, K)$, atualize \mathbf{S} , através dos seguintes passos:

- (1) A partir da configuração atual, \mathbf{s} , obtenha uma configuração de ligações \mathbf{B} , que é um conjunto de “arestas” conectando *pixels* vizinhos. Esta transição, $\mathbf{s} \rightarrow \mathbf{B}$, é obtida da seguinte forma: para todo $i \sim j$, se $s_i = s_j$, então “crie” uma aresta entre os *pixels* i e j com probabilidade $1 - e^{-\beta}$. Caso contrário, nada é feito;
- (2) Converta o resultado da configuração de ligações \mathbf{B} em uma nova configuração \mathbf{s}' como segue:

- (i) Identifique as regiões R_1, \dots, R_M de *pixels* conectados através de arestas;
- (ii) Para cada região R_m estabeleça um rótulo de acordo com a probabilidade em (3.15);
- (iii) Se R_m é rotulado por k , então faça $s_i = k$, para $i = 1, \dots, n$, $m = 1, \dots, M$ e $k = 1, \dots, K$. Faça $\mathbf{S} = \mathbf{s}'$.

3.2.2 Atualização de K

Para atualizar o número de componentes K utilizamos os movimentos ejeção (ej) e absorção (ab). Estes movimentos mudam o número de componentes K nas vizinhanças $K + 1$ e $K - 1$, respectivamente. No movimento ejeção, as observações de uma componente são divididas em duas novas componentes, aumentando K em uma unidade, $K^{ej} = K + 1$. No movimento absorção, as observações de duas componentes são unidas em uma nova componente e K diminui uma unidade, $K^{ab} = K - 1$.

Para o desenvolvimento destes dois movimentos, considere $P_{ej|K}$ e $P_{ab|K}$ as probabilidades de propor uma ejeção e uma absorção, respectivamente, com $P_{ej|K} + P_{ab|K} = 1$. Estas probabilidades dependem do número de componentes K , pois se $K = 1$ (ou $K = K_m$), é permitido fazer apenas uma ejeção (ou absorção) com probabilidade 1, enquanto que para $2 \leq K \leq K_m - 1$, é permitido fazer qualquer um dos dois movimentos com probabilidade $1/2$, *i.e.*,

$$P_{ej|K} = \begin{cases} 0, & \text{se } K = K_m \\ 1, & \text{se } K = 1 \\ 1/2, & \text{caso contrário} \end{cases}, \quad P_{ab|K} = \begin{cases} 1, & \text{se } K = K_m \\ 0, & \text{se } K = 1 \\ 1/2, & \text{caso contrário} \end{cases}.$$

Considere $\phi = (\boldsymbol{\theta}_K, \mathbf{s}, K)$ o estado atual, $\phi^{ej} = (\boldsymbol{\theta}_{K+1}^{ej}, \mathbf{s}^{ej}, K + 1)$ e $\phi^{ab} = (\boldsymbol{\theta}_{K-1}^{ab}, \mathbf{s}^{ab}, K - 1)$ os estados propostos obtidos pelos movimentos ejeção e absorção, respectivamente. Para manter válida a equação de balanceamento quando um dos movimentos é proposto, considere a seguinte parametrização alternativa, obtida aumentando $\boldsymbol{\theta}_K$ para

$$\boldsymbol{\theta}_{K_m} = (\theta_1, \dots, \theta_K, \theta_{K+1}, \dots, \theta_{K_m}) = (\boldsymbol{\theta}_K, \boldsymbol{\theta}_0),$$

onde $k = 1, \dots, K$ representam as K componentes não vazias e $k = K + 1, \dots, K_m$ representam as $K_m - K$ componentes vazias, com $\boldsymbol{\theta}_0 = (\theta_{K+1}, \dots, \theta_{K_m})$ sendo interpretada como “potenciais” componentes, porém ainda não utilizadas. Considere a mesma definição para a variável indicadora não observável \mathbf{S} com distribuição *a priori* dada em (3.6). Além disso, assuma que $\theta_{K+1}, \dots, \theta_{K_m}$ são independentes entre si e também de $\boldsymbol{\theta}_K$.

Assim, a distribuição *a priori* para $\boldsymbol{\theta}_0$ dado K é

$$\pi(\boldsymbol{\theta}_0|K) = \prod_{k=K+1}^{K_m} \pi(\theta_k).$$

Carlin e Chib (1995), Godsill (2001) e Dellaportas *et al.* (2002) chamam $\pi(\boldsymbol{\theta}_0|K)$ de *pseudo-prior*, pois, dado K , cada $\pi(\theta_k)$, para $k = K + 1, \dots, K_m$, não representa efetivamente uma distribuição *a priori*, sendo somente conveniente para especificar um modelo ampliado com um espaço paramétrico de dimensão fixa.

Dado um valor de K , a função verossimilhança aumentada é dada por

$$L(\boldsymbol{\theta}_{K_m}, K|\mathbf{y}, \mathbf{s}) = P(\mathbf{y}|\mathbf{s}, \boldsymbol{\theta}_{K_m}, K)\pi(\mathbf{S}|K) = P(\mathbf{y}|\mathbf{s}, \boldsymbol{\theta}_K, K)\pi(\mathbf{S}|K) = L(\boldsymbol{\theta}_K, K|\mathbf{y}, \mathbf{s}),$$

pois

$$P(\mathbf{y}|\mathbf{s}, \boldsymbol{\theta}_{K_m}, K) = \prod_{i=1}^n f(y_i)^{\mathbb{I}_{s_i}(k)} = P(\mathbf{y}|\mathbf{s}, \boldsymbol{\theta}_K, K),$$

devido às componentes $K + 1, \dots, K_m$ serem componentes vazias.

A distribuição *a posteriori* é dada por

$$\pi(\boldsymbol{\theta}_{K_m}, \mathbf{S} = \mathbf{s}, \beta, K|\mathbf{y}) \propto P(\mathbf{y}|\mathbf{s}, \boldsymbol{\theta}_K, K)\pi(\mathbf{S} = \mathbf{s}|\beta, K)\pi(\boldsymbol{\theta}_K|K) \left(\prod_{k=K+1}^{K_m} \pi(\theta_k) \right) \pi(\beta)\pi(K).$$

Dessa forma, a probabilidade de aceitação para os movimentos ejeção e absorção é dada pela probabilidade de aceitação Metropolis-Hastings (Hastings, 1970; Chib e Greenberg, 1995). Ou seja, dado o estado atual $\boldsymbol{\phi} = (\boldsymbol{\theta}_K \cup \boldsymbol{\theta}_0, \mathbf{s}, K)$, a probabilidade de aceitação para o estado proposto $\boldsymbol{\phi}^* = (\boldsymbol{\theta}_K^* \cup \boldsymbol{\theta}_0^*, \mathbf{s}^*, K^*)$ é dada por

$\Psi(\phi^*|\phi) = \min(1, A^*)$, onde

$$A^* = \frac{P(\mathbf{y}|\mathbf{s}^*, \boldsymbol{\theta}_{K^*}^*, K^*)}{P(\mathbf{y}|\mathbf{s}, \boldsymbol{\theta}_K, K)} \frac{\pi(\mathbf{s}^*|\beta, K^*)}{\pi(\mathbf{s}|\beta, K)} \frac{\pi(\boldsymbol{\theta}_{K^*}^*|K^*)}{\pi(\boldsymbol{\theta}_K|K)} \frac{\prod_{k=K^*+1}^{K_m} \pi(\theta_k^*)}{\prod_{k=K+1}^{K_m} \pi(\theta_k)} \frac{q[\phi|\phi^*]}{q[\phi^*|\phi]}, \quad (3.16)$$

o sinal “*” indica a nova proposta ejeção (ou absorção) e $q(\cdot|\cdot)$ é a proposta de transição em que ϕ^* é obtido, seja através de um movimento ejeção (ou absorção), descrito a seguir.

Movimento ejeção

Para o desenvolvimento do movimento ejeção, considere y_a e y_b duas observações pertencentes a uma componente D_k , para $k \in \{1, \dots, K\}$. Seja

$$P(y_a, y_b) = \frac{B_{ab}}{1 + B_{ab}}$$

uma medida de dissimilaridade entre as observações y_a e y_b , $y_a, y_b \in D_k$, onde $B_{ab} = \frac{\mathbf{I}(y_a)\mathbf{I}(y_b)}{\mathbf{I}(y_a, y_b)}$ para $\mathbf{I}(y_a) = \int f(y_a|\theta_{k_1})\pi(\theta_{k_1})d\theta_{k_1}$, $\mathbf{I}(y_b) = \int f(y_b|\theta_{k_2})\pi(\theta_{k_2})d\theta_{k_2}$ e $\mathbf{I}(y_a, y_b) = \int f(y_a|\theta_k)f(y_b|\theta_k)\pi(\theta_k)d\theta_k$. Note que B_{ab} pode ser interpretada como sendo o fator de Bayes do modelo que assume $y_a \sim F(\theta_{k_1})$ e $y_b \sim F(\theta_{k_2})$, com $\theta_{k_1} \neq \theta_{k_2}$, em relação ao modelo que assume $y_a, y_b \sim F(\theta_k)$. Como estamos utilizando um modelo normal com distribuições *a priori* conjugadas, então a expressão para $\mathbf{I}(\cdot)$ é analiticamente conhecida e é dada por

$$\mathbf{I}(D^*) = \gamma^* \lambda^* \Gamma^* \left[1 + \frac{\sum_{y^* \in D^*} y^{*2} + \lambda \mu_0^2}{\gamma} - \frac{\left(\sum_{y^* \in D^*} y^* + \lambda \mu_0 \right)^2}{\gamma(n^* + \lambda)} \right]^{-\frac{\alpha+n^*}{2}}, \quad (3.17)$$

onde $\gamma^* = \left[\frac{1}{\gamma\pi} \right]^{\frac{n^*}{2}}$, $\lambda^* = \left[\frac{\lambda}{n^* + \lambda} \right]^{\frac{1}{2}}$ e $\Gamma^* = \frac{\Gamma(\frac{\alpha+n^*}{2})}{\Gamma(\frac{\alpha}{2})}$, para $D^* \in \{\{y_a\}, \{y_b\}, \{y_a, y_b\}\}$ e $n^* = |D^*|$. Os cálculos para obtenção de $\mathbf{I}(\cdot)$ estão apresentados no Apêndice A6.

Considere que as observações $y_{v'}$, $y_{v''} \in D_k$ apresentem máxima medida de dissimilaridade, *i.e.*,

$$P(y_{v'}, y_{v''}) = \max_{\forall y_a, y_b \in D_k} (P(y_a, y_b)). \quad (3.18)$$

Assim, dada a escolha por um movimento ejeção, considere K_2 o número de componentes, tal que $n_k \geq 2$, para $k = 1, \dots, K$. Escolha uma componente k , com probabilidade $P^{ej}(k|K_2) = 1/K_2$. Proponha um movimento ejeção da seguinte maneira:

- (1) Inicialize escolhendo duas observações, $y_{v'}$ e $y_{v''}$ de D_k , de acordo com (3.18);
- (2) Se $y_{v'}$ e $y_{v''}$ pertencem a regiões diferentes, *i.e.*, $y_{v'} \in R_{m'}$ e $y_{v''} \in R_{m''}$, para $m', m'' \in \{1, \dots, M\}$ e $m' \neq m''$, faça $R_{m'}^{ej} = R_{m'}$, $R_{m''}^{ej} = R_{m''}$, porém com rótulos k_1 e k_2 , respectivamente. Então, crie duas novas componentes, denotadas por $D_{k_1}^{ej}$ e $D_{k_2}^{ej}$, como segue:

- (a) Faça $D_{k_1}^{ej} = \{R_{m'}^{ej}\}$, $D_{k_2}^{ej} = \{R_{m''}^{ej}\}$, $n_{k_1}^{ej} = \|R_{m'}^{ej}\|$ e $n_{k_2}^{ej} = \|R_{m''}^{ej}\|$;

- (b) Se existir outra região R_m rotulada como k , para $m \in \{1, \dots, M\}$ e $m \neq m', m''$, então faça $R_m^{ej} = R_m$ e rotule esta região como k_1 com probabilidade

$$P(R_m = k_1) = \frac{1}{1 + B_{mk_1}},$$

onde $B_{mk_1} = \frac{\mathbf{I}(D_{k_2}^{ej} \cup R_m)}{\mathbf{I}(D_{k_1}^{ej} \cup R_m)}$ e $\mathbf{I}(\cdot)$ é dada em (3.17). A probabilidade de rotular R_m^{ej} como k_2 é $1 - P(R_m = k_1)$. Se R_m^{ej} é rotulada como k_1 , faça $D_{k_1}^{ej} = \{D_{k_1}\} \cup \{R_m^{ej}\}$. Caso contrário, rotule R_m^{ej} como k_2 e faça $D_{k_2}^{ej} = \{D_{k_2}\} \cup \{R_m^{ej}\}$;

- (c) Obtenha uma nova configuração para as variáveis não observáveis, denotada por \mathbf{s}^{ej} , onde $s_i^{ej} = k_1$, para todo $y_i \in D_{k_1}$ e $s_i^{ej} = k_2$, para todo $y_i \in D_{k_2}$;

- (3) Se $y_{v'}$ e $y_{v''}$ pertencem à mesma região, *i.e.*, $y_{v'}, y_{v''} \in R_{m'}$, então proponha uma ejeção da seguinte maneira:

- (a) Faça $D_{k_1}^{ej} = \{y_{v'}\}$, $D_{k_2}^{ej} = \{y_{v''}\}$ e $n_{k_1}^{ej} = n_{k_2}^{ej} = 1$;

- (b) Para todo $y_i \in D_k$ tal que $y_i \neq \{y_{v'}, y_{v''}\}$, calcule a medida de similaridade

$$P(y_i, y_{v'}) = \frac{1}{1 + B_{iv'}},$$

onde $B_{iv'} = \frac{\mathbf{I}(y_{v''}, y_i)}{\mathbf{I}(y_{v'}, y_i)}$ e $\mathbf{I}(\cdot)$ é dada em (3.17);

- (c) Gere $U_{ref} \sim \text{Beta}(2, 2)$. Se $P(y_i, y_{v'}) > u_{ref}$, aloque y_i em $D_{k_1}^{ej}$, i.e., faça $D_{k_1}^{ej} = \{D_{k_1}^{ej}\} \cup \{y_i\}$. Caso contrário, aloque y_i em $D_{k_2}^{ej}$;
- (d) Remova a ligação (aresta) entre os *pixels* vizinhos i e j se $y_i \in D_{k_1}^{ej}$ e $y_j \in D_{k_2}^{ej}$;
- (e) Proceda como nos itens (2-b) e (2-c) acima.

Para completar a proposta da nova configuração $\mathbf{s}^{ej} = (s_1^{ej}, \dots, s_n^{ej})$, fixe os novos rótulos da seguinte forma: (a) no estado proposto, faça $k_1 = k$ e $k_2 = k + 1$; (b) para todo $k' > k$, no estado atual, faça $k' = k' + 1$ no estado proposto, para $k \in \{1, \dots, K\}$. Dessa forma, obtemos uma nova configuração \mathbf{s}^{ej} com $K^{ej} = K + 1$ componentes, onde:

- (i) $s_i^{ej} = s_i$ se $s_i = k'$ para $k' = 1, \dots, k - 1$;
- (ii) $s_i^{ej} = s_i + 1$ se $s_i = k'$ para $k' = k + 1, \dots, K$;
- (iii) $s_i^{ej} = k$ ou $s_i^{ej} = k + 1$ se $s_i = k$, de acordo com os passos (2) ou (3) acima;

para $i = 1, \dots, n$.

De (3.6), a probabilidade da configuração \mathbf{s}^{ej} é dada por

$$\pi(\mathbf{s}^{ej} | \beta, K + 1) = g^{-1}(\beta, K + 1) \exp \left\{ \beta \sum_{i \sim j} \mathbb{I}_{s_i^{ej}}(s_j^{ej}) \right\}.$$

Condicional em D_k^{ej} e D_{k+1}^{ej} , gere valores candidatos θ_k^{ej} e θ_{k+1}^{ej} das distribuições condicionais dadas em (3.9) e (3.10), respectivamente, obtendo novos vetores paramétricos $\boldsymbol{\theta}_{K+1}^{ej} = (\theta_1^{ej}, \dots, \theta_k^{ej}, \theta_{k+1}^{ej}, \dots, \theta_{K+1}^{ej})$ e $\boldsymbol{\theta}_0^{ej} = (\theta_{K+2}^{ej}, \dots, \theta_{K_m}^{ej})$, onde:

- (i) Para todo $\theta_{k'}^{ej} \in \boldsymbol{\theta}_{K+1}^{ej}$, tal que $k' < k$, faça $\theta_{k'}^{ej} = \theta_{k'}$, para $\theta_{k'} \in \boldsymbol{\theta}_K$;
- (ii) Para todo $\theta_{k'}^{ej} \in \boldsymbol{\theta}_{K+1}^{ej}$, tal que $k' > k + 1$, faça $\theta_{k'}^{ej} = \theta_{k'-1}$, para $\theta_{k'} \in \boldsymbol{\theta}_K$ e $k' \in \{1, \dots, K + 1\} \setminus \{k, k + 1\}$;
- (iii) Para todo $k \in \{K + 2, \dots, K_m\}$, faça $\theta_k^{ej} \sim \pi(\theta_k)$ e $\boldsymbol{\theta}_0^{ej} = (\theta_{K+2}^{ej}, \dots, \theta_{K_m}^{ej})$.

Devido ao tipo de rotulação assumido na Seção 3.1.2, considerando a ordem das médias, neste ponto é necessário verificar se a condição de adjacência é satisfeita,

i.e., se $\mu_{k-1}^{ej} < \mu_k^{ej} < \mu_{k+1}^{ej} < \mu_{k+2}^{ej}$. Caso a condição não seja satisfeita, então a proposta deve ser rejeitada, pois o movimento pode não ser reversível pelo movimento absorção.

Se a condição de adjacência é satisfeita, então temos uma nova configuração \mathbf{s}^{ej} , um novo conjunto de parâmetros $\boldsymbol{\theta}_{K_m}^{ej} = (\boldsymbol{\theta}_{K+1}^{ej}, \boldsymbol{\theta}_0^{ej})$, e o número de componentes K aumenta uma unidade. Esta proposta de transição é denotada por $\boldsymbol{\phi}^{ej}|\boldsymbol{\phi}$, onde $\boldsymbol{\phi}^{ej} = (\boldsymbol{\theta}_{K_m}^{ej}, \mathbf{s}^{ej}, K + 1)$, e sua probabilidade é dada por

$$q[\boldsymbol{\phi}^{ej}|\boldsymbol{\phi}] = P_{ej|K} P^{ej}(k|K_2) \pi(\mathbf{s}^{ej}|\beta, K + 1) \pi(\theta_k^{ej}|D_k^{ej}) \pi(\theta_{k+1}^{ej}|D_{k+1}^{ej}) \prod_{k=K+2}^{K_m} \pi(\theta_k^{ej}).$$

Movimento absorção

No movimento absorção, uma componente rotulada como k absorve a componente adjacente $k + 1$, para $k \in \{1, \dots, K - 1\}$. A probabilidade de selecionar as componentes k e $k + 1$ para serem unidas é $P^{ab}(k, k + 1) = \frac{1}{K-1}$.

Escolhidas as duas componentes, “crie” uma nova componente $D_k^{ab} = \{D_k\} \cup \{D_{k+1}\}$. Conecte através de arestas todos os vizinhos i e j com $y_i, y_j \in D_k^{ab}$. Assim, obtemos uma nova configuração $\mathbf{s}^{ab} = (s_1^{ab}, \dots, s_n^{ab})$ com $K^{ab} = K - 1$ componentes, onde:

- (i) $s_i^{ab} = s_i$ se $s_i = k'$ para $k' = 1, \dots, k$ e
- (ii) $s_i^{ab} = s_i - 1$ se $s_i = k'$ para $k' = k + 1, \dots, K$, $i = 1, \dots, n$.

De (3.6), a probabilidade da configuração \mathbf{s}^{ab} é

$$\pi(\mathbf{s}^{ab}|\beta, K - 1) = g^{-1}(\beta, K - 1) \exp \left\{ \beta \sum_{i \sim j} \mathbb{I}_{s_i^{ab}}(s_j^{ab}) \right\}.$$

Condicional em D_k^{ab} , gere um valor candidato θ_k^{ab} das distribuições condicionais em (3.9) e (3.10). Este procedimento determina novos vetores de parâmetros $\boldsymbol{\theta}_{K-1}^{ab} = (\theta_1^{ab}, \dots, \theta_{K-1}^{ab})$ e $\boldsymbol{\theta}_0^{ab} = (\theta_K^{ab}, \dots, \theta_{K_m}^{ab})$, onde:

- (i) Para todo $\theta_{k'}^{ab} \in \boldsymbol{\theta}_{K-1}^{ab}$, tal que $k' < k$, faça $\theta_{k'}^{ab} = \theta_{k'}$, para $\theta_{k'} \in \boldsymbol{\theta}_K$;
- (ii) Para todo $\theta_{k'}^{ab} \in \boldsymbol{\theta}_{K-1}^{ab}$, tal que $k' > k$, faça $\theta_{k'}^{ab} = \theta_{k'+1}$, para $\theta_{k'} \in \boldsymbol{\theta}_K$;

(iii) Para todo $k \in \{K, \dots, K_m\}$, faça $\theta_k^{ab} \sim \pi(\theta_k)$ e $\boldsymbol{\theta}_0^{ab} = (\theta_K^{ab}, \dots, \theta_{K_m}^{ab})$.

Como feito no movimento ejeção, neste ponto é necessário verificar se a condição de adjacência é satisfeita, *i.e.*, se $\theta_{k-1}^{ab} < \theta_k^{ab} < \theta_{k+1}^{ab}$. Caso não seja, a proposta deve ser rejeitada, pois o movimento pode não ser reversível.

Se a condição de adjacência é satisfeita, então temos uma nova configuração \mathbf{s}^{ab} , um novo vetor de parâmetros $\boldsymbol{\theta}_{K_m}^{ab} = (\boldsymbol{\theta}_{K-1}^{ab}, \boldsymbol{\theta}_0^{ab})$, e o número K de componentes diminui uma unidade. Esta proposta de transição é denotada por $\boldsymbol{\phi}^{ab}|\boldsymbol{\phi}$, onde $\boldsymbol{\phi}^{ab} = (\boldsymbol{\theta}_{K_m}^{ab}, \mathbf{s}^{ab}, K-1)$, e sua probabilidade é dada por

$$q[\boldsymbol{\phi}^{ab}|\boldsymbol{\phi}] = P_{ab|K} P^{ab}(k, k+1) \pi(\mathbf{s}^{ab}|\beta, K-1) \pi(\theta_k^{ab}|D_k^{ab}) \prod_{k=K}^{K_m} \pi(\theta_k).$$

Além disso, observe que, dado o estado atual $\boldsymbol{\phi}$, a probabilidade de propor uma ejeção na componente D_k , obtendo as componentes D_k^{ej} , D_{k+1}^{ej} e o estado proposto $\boldsymbol{\phi}^{ej}$, é equivalente a estar no estado $\boldsymbol{\phi}^{ab}$ com as componentes D_k e D_{k+1} juntas em uma componente $D_k^{ab} = \{D_k\} \cup \{D_{k+1}\}$, e propor o movimento de volta para o estado atual $\boldsymbol{\phi}$. Ou seja,

$$\begin{aligned} q[\boldsymbol{\phi}^{ej}|\boldsymbol{\phi}] &= q[\boldsymbol{\phi}|\boldsymbol{\phi}^{ab}] \\ &= P_{ej|K-1} P^{ej}(k|K_2) \pi(\mathbf{s}|\beta, K) \pi(\theta_k|D_k) \pi(\theta_{k+1}|D_{k+1}) \left(\prod_{k=K+1}^{K_m} \pi(\theta_k) \right). \end{aligned} \quad (3.19)$$

Análogo à (3.19),

$$\begin{aligned} q[\boldsymbol{\phi}^{ab}|\boldsymbol{\phi}] &= q[\boldsymbol{\phi}|\boldsymbol{\phi}^{ej}] \\ &= P_{ab|K+1} P^{ab}(k, k+1) \pi(\mathbf{s}|\beta, K) \pi(\theta_k|D_k) \left(\prod_{k=K+1}^{K_m} \pi(\theta_j) \right). \end{aligned}$$

Probabilidade de aceitação

De (3.16), a probabilidade de aceitação para uma proposta ejeção é dada por $\Psi(\boldsymbol{\phi}^{ej}|\boldsymbol{\phi}) = \min(1, A^{ej})$, onde

$$A^{ej} = \frac{P(\mathbf{y}|\mathbf{s}^{ej}, \boldsymbol{\theta}_{K+1}^{ej}, K+1) \pi(\mathbf{s}^{ej}|\beta, K+1) \pi(\boldsymbol{\theta}_{K+1}^{ej}|K+1) \prod_{k=K+2}^{K_m} \pi(\theta_k^{ej})}{P(\mathbf{y}|\mathbf{s}, \boldsymbol{\theta}_K, K) \pi(\mathbf{s}|\beta, K) \pi(\boldsymbol{\theta}_K|K) \prod_{k=K+1}^{K_m} \pi(\theta_k)} \frac{q[\boldsymbol{\phi}|\boldsymbol{\phi}^{ej}]}{q[\boldsymbol{\phi}^{ej}|\boldsymbol{\phi}]}$$

e a probabilidade $P(\mathbf{y}|\cdot)$ é dada em (3.4). Logo,

$$\frac{P(\mathbf{y}|\mathbf{s}^{ej}, \boldsymbol{\theta}_{K+1}^{ej}, K+1)}{P(\mathbf{y}|\mathbf{s}, \boldsymbol{\theta}_K, K)} = \frac{\left(\prod_{y_i \in D_k^{ej}} f(y_i|\theta_k^{ej}) \right) \left(\prod_{y_i \in D_{k+1}^{ej}} f(y_i|\theta_{k+1}^{ej}) \right)}{\prod_{y_i \in D_k} f(y_i|\theta_k)} \quad (3.20)$$

e

$$\frac{\pi(\mathbf{s}^{ej}|\beta, K+1)}{\pi(\mathbf{s}|\beta, K)} = \frac{g^{-1}(\beta, K+1) \exp \left\{ \beta \sum_{i \sim j} \mathbb{I}_{s_i} (s_j^{ej}) \right\}}{g^{-1}(\beta, K) \exp \left\{ \beta \sum_{i \sim j} \mathbb{I}_{s_i} (s_j) \right\}}. \quad (3.21)$$

A razão das distribuições *a priori* para a proposta ejeção é dada por

$$\frac{\pi(\boldsymbol{\theta}_{K_m}^{ej})}{\pi(\boldsymbol{\theta}_{K_m})} = \frac{\pi(\theta_k^{ej}) \pi(\theta_{k+1}^{ej}) \left(\prod_{k=K+2}^{K_m} \pi(\theta_k^{ej}) \right)}{\pi(\theta_k) \left(\prod_{k=K+1}^{K_m} \pi(\theta_k) \right)}. \quad (3.22)$$

A razão da probabilidade de transição para a proposta ejeção é dada por

$$\begin{aligned} \frac{q[\boldsymbol{\phi}|\boldsymbol{\phi}^{ej}]}{q[\boldsymbol{\phi}^{ej}|\boldsymbol{\phi}]} &= \frac{P_{ab|K+1} P^{ab}(k, k+1)}{P_{ej|K} P^{sp}(k|K_2)} \frac{\pi(\mathbf{s}|\beta, K)}{\pi(\mathbf{s}^{ej}|\beta, K+1)} \frac{\pi(\theta_k|D_k)}{\pi(\theta_k^{ej}|D_k^{ej})\pi(\theta_{k+1}^{ej}|D_{k+1}^{ej})} \\ &\quad \times \frac{\left(\prod_{k=K+1}^{K_m} \pi(\theta_k) \right)}{\left(\prod_{k=K+2}^{K_m} \pi(\theta_k^{ej}) \right)} \\ &= Q^{ej} \frac{\pi(\mathbf{s}|\beta, K)}{\pi(\mathbf{s}^{ej}|\beta, K+1)} \frac{\prod_{y_i \in D_k} f(y_i|\theta_k)}{\left(\prod_{y_i \in D_k^{ej}} f(y_i|\theta_k^{ej}) \right) \left(\prod_{y_i \in D_{k+1}^{ej}} f(y_i|\theta_{k+1}^{ej}) \right)} \\ &\quad \times \frac{\pi(\theta_k)}{\pi(\theta_k^{sp})\pi(\theta_{k+1}^{sp})} \frac{\mathbf{I}(D_k^{sp})(D_{k+1}^{sp})}{\mathbf{I}(D_k)} \frac{\left(\prod_{k=K+1}^{K_m} \pi(\theta_k) \right)}{\left(\prod_{k=K+2}^{K_m} \pi(\theta_k^{ej}) \right)}. \quad (3.23) \end{aligned}$$

onde

$$Q^{ej} = \frac{P_{ab|K+1} P^{ab}(k, k+1)}{P_{sp|K} P^{ej}(k|K_2)} = \begin{cases} \frac{1}{2} & , \text{ se } k = 1 \\ 2^{\mathbb{1}_{K(K_m-1)}} \frac{K_2}{K-1} & , \text{ se } 2 \leq k \leq K_m - 1 \end{cases},$$

para $\mathbb{I}_K(K_m - 1) = 1$ se $K = K_m - 1$ e $\mathbb{I}_K(K_m - 1) = 0$ caso contrário; e $\mathbf{I}(D^*)$ é a constante normalizadora, dada em (3.17), para $D^* \in \{D_k, D_k^{ej}, D_{k+1}^{ej}\}$.

Portanto, multiplicando (3.20), (3.21), (3.22) e (3.23), a probabilidade de aceitação para a proposta ejeção é $\Psi[\phi^{ej}|\phi] = \min(1, A^{ej})$, onde

$$A^{ej} = Q^{ej} \frac{\mathbf{I}(D_k^{ej}) \mathbf{I}(D_{k+1}^{ej})}{\mathbf{I}(D_k)}.$$

A probabilidade de aceitação para a proposta absorção é $\Psi[\phi^{ab}|\phi] = \min(1, A^{ab})$, onde $A^{ab} = \frac{1}{A^{ej}}$, porém com mudanças nas notações.

3.2.3 Algoritmo EAM

A seguir apresentamos os passos do algoritmo EAM:

- (1) Inicialize com um valor $K = K^{(1)}$, uma configuração s válida e parâmetros $\theta_k^{(1)}$;
- (2) Para a l -ésima iteração, $l = 2, \dots, L$, faça como segue:
 - (i) Escolha entre os movimentos ejeção e absorção com probabilidades $P_{ej|K}$ e $P_{ab|K}$, respectivamente;
 - (ii) Aceite a proposta com probabilidade $\Psi[\phi^*|\phi]$, onde $*$ denota ej ou ab ;
 - (a) Se o movimento ejeção é aceito, faça $K^{(l)} = K^{(l-1)} + 1$;
 - (b) Se o movimento absorção é aceito, faça $K^{(l)} = K^{(l-1)} - 1$;
 - (c) Caso contrário, mantenha $K^{(l)} = K^{(l-1)}$.
 - (iii) Dado $K^{(l)}$, atualize $\theta_{K^{(l)}}$, β e \mathbf{S} utilizando os algoritmos *Gibbs sampling*, *Metropolis-Hastings* e *Swendsen-Wang*, respectivamente, como descrito na Seção 3.2.1.

Para estimar o número de componentes K , considere um *burn in* de tamanho L_{burn} e calcule $N_{K=k}$, que denota o número de vezes em que $K = k$ nas $L - L_{burn}$ iterações, para $k = 1, \dots, K_m$. Considerando $P(K = k|\cdot) = N_{K=k}/(L - L_{burn})$ como sendo a estimativa da probabilidade *a posteriori* para $K = k$, $k = 1, \dots, K_m$, então $\tilde{K} = \arg \max_{1 \leq k \leq K_m} (P(K = k|\cdot))$ é a estimativa para o número de componentes.

Como o modelo mais provável é um com \tilde{K} componentes, então nossa inferência deve ser condicional em \tilde{K} . Assim, dado \tilde{K} , considere

- (i) $L_{\tilde{K}}$ é o número de iterações onde $K = \tilde{K}$ nas $L - L_{burn}$ iterações, *i.e.*,

$$L_{\tilde{K}} = \sum_{l=L_{burn}+1}^L \mathbb{I}_{K^{(l)}}(\tilde{K}),$$
 onde $\mathbb{I}_{K^{(l)}}(\tilde{K}) = 1$ se na l -ésima iteração $K = \tilde{K}$ e $\mathbb{I}_{K^{(l)}}(\tilde{K}) = 0$, caso contrário;
- (ii) N_{ik} é o número de vezes em que a observação y_i é alocada na componente k nas $L_{\tilde{K}}$ iterações, *i.e.*,
$$N_{ik} = \sum_{l=L_{burn}+1}^L \mathbb{I}_{s_i^{(l)}}(k) \mathbb{I}_{K^{(l)}}(\tilde{K}),$$
 onde $\mathbb{I}_{s_i^{(l)}}(k) = 1$ se $s_i = k$ na l -ésima iteração e $\mathbb{I}_{s_i^{(l)}}(k) = 0$, caso contrário, para $i = 1, \dots, n$ e $k = 1, \dots, \tilde{K}$.

Assim, a estimativa de θ_k é dada pela média dos valores gerados, *i.e.*,

$$\tilde{\theta}_k | \tilde{K} = \frac{1}{L_{\tilde{K}}} \sum_{l=L_{burn}+1}^L \theta_k^{(l)} \mathbb{I}_{K^{(l)}}(\tilde{K}),$$

para $k = 1, \dots, \tilde{K}$.

A estimativa da probabilidade *a posteriori* da observação y_i ser proveniente da componente k é $P_{ik} = N_{ik}/L_{\tilde{K}}$. Se $P_{ik} = \max_{1 \leq k \leq \tilde{K}} (P_{ik})$, então considere que y_i é gerado da componente k , *i.e.*, faça $s_i = k$, para $i = 1, \dots, n$ e $k = 1, \dots, \tilde{K}$.

Capítulo 4

Estudo de simulação e aplicação

Neste capítulo, aplicamos o método proposto, EAM-MCMC, a conjuntos de dados simulados e a um conjunto de dados reais. Para cada conjunto de dados, apresentamos a probabilidade *a posteriori* para o número de componentes K . Além disso, fazemos uma análise empírica da convergência da probabilidade *a posteriori* de K , apresentando a média *ergódica* dos valores de K , o gráfico dos valores de K observados durante as iterações, a estimativa da autocorrelação e a imagem identificada pelo método.

Para a aplicação do método, os hiperparâmetros foram fixados de forma que as distribuições *a priori* sejam não informativas, *i.e.*, apresentem “grandes” variâncias. Para isto, considere (a, b) um intervalo que contenha todos os dados observados e Δ a amplitude deste intervalo, onde a e b são fixados com base na informação de especialistas ou com base nos valores observados. Fixamos α e γ de forma que $E(\sigma_k^2) = c\Delta$, onde c é uma constante. Dessa forma, obtemos $\gamma = (\alpha - 2)c\Delta$. Fixamos $\frac{\alpha}{2} = 2$ (Richardson e Green, 1997) e assumimos $c = 4$. Além disso, fixamos μ_0 como sendo o ponto médio do intervalo, *i.e.*, $\mu_0 = \frac{\Delta}{2}$ e fizemos $\lambda = 10^{-2}$. Assim, obtemos as seguintes distribuições *a priori* para os parâmetros das componentes

$$\mu_k | \sigma_k^2, \lambda, \mu_0 \sim \mathcal{N}\left(\frac{\Delta}{2}, 100\sigma_k^2\right) \quad \text{e} \quad \sigma_k^2 | \alpha, \gamma \sim \mathcal{IG}(2, 4\Delta).$$

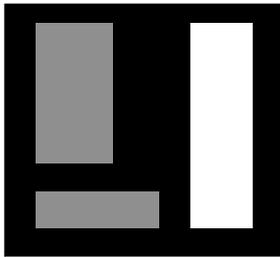
Na próxima seção, apresentamos o valor de Δ e as distribuições *a priori* com os valores dos hiperparâmetros para cada um dos conjuntos de dados simulados. Além disso, para a distribuição *a priori* de K , fixamos $K_m = 10$.

4.1 Dados simulados

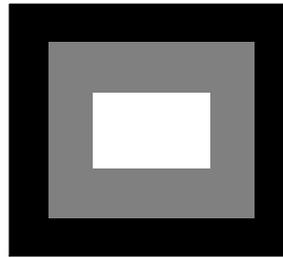
Os dados foram gerados de acordo com o número K_{real} de componentes, quantidade n de *pixels* e intensidade real μ_k dos *pixels* de cada componente, para $k = 1, \dots, K_{real}$, conforme especificados na Tabela 4.1. Assumimos que a variância σ_k^2 do ruído é igual a 1, $\sigma_k^2 = 1$, para todo $k = 1, \dots, K_{real}$. A Figura 4.1 mostra as imagens geradas.

Tabela 4.1: Número de componentes, quantidade de *pixels* e intensidade real

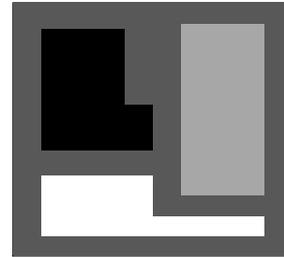
Dados simulados	Número de componentes	Quantidade de <i>pixels</i>	Intensidade real das componentes
A_1	$K_{real} = 3$	$n = 486$	$\mu_1 = -4, \mu_2 = 0, \mu_3 = 3$
A_2	$K_{real} = 3$	$n = 1.000$	$\mu_1 = -4, \mu_2 = 0, \mu_3 = 4$
A_3	$K_{real} = 4$	$n = 1.000$	$\mu_1 = -5, \mu_2 = 0, \mu_3 = 4, \mu_4 = 9$
A_4	$K_{real} = 4$	$n = 1.000$	$\mu_1 = -4, \mu_2 = 0, \mu_3 = 5, \mu_4 = 9$
A_5	$K_{real} = 6$	$n = 1.000$	$\mu_1 = -10, \mu_2 = -5, \mu_3 = 0, \mu_4 = 4, \mu_5 = 9, \mu_6 = 13$
A_6	$K_{real} = 4$	$n = 10.000$	$\mu_1 = -6, \mu_2 = 0, \mu_3 = 7, \mu_4 = 11$



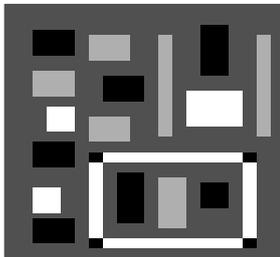
(a) Dados A_1



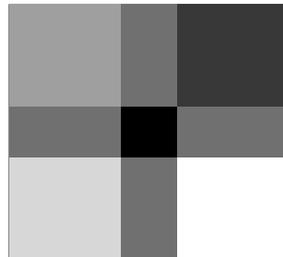
(b) Dados A_2



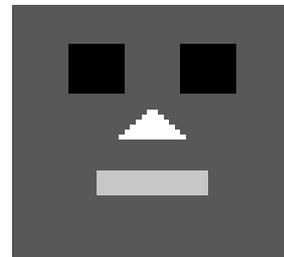
(c) Dados A_3



(d) Dados A_4



(e) Dados A_5



(f) Dados A_6

Figura 4.1: Imagens geradas

Na Figura 4.1(a), temos uma grade \mathcal{G} de dimensão 27×18 dividida em quatro regiões, $M = 4$, e três componentes, $K = 3$. Nas Figuras 4.1(b,c,d,e) temos uma grade \mathcal{G} de dimensão 50×20 divididas em 3, 4, 25 e 9 regiões, respectivamente, e número de componentes igual 3, 4, 4 e 6, respectivamente. Na Figura 4.1(f), temos uma grade \mathcal{G} de dimensão 100×100 dividida em cinco regiões, $M = 5$, e quatro componentes, $K = 4$. A Figura 4.2 mostra os valores gerados de cada componente para cada um dos conjuntos de dados simulados, A_1 a A_6 .

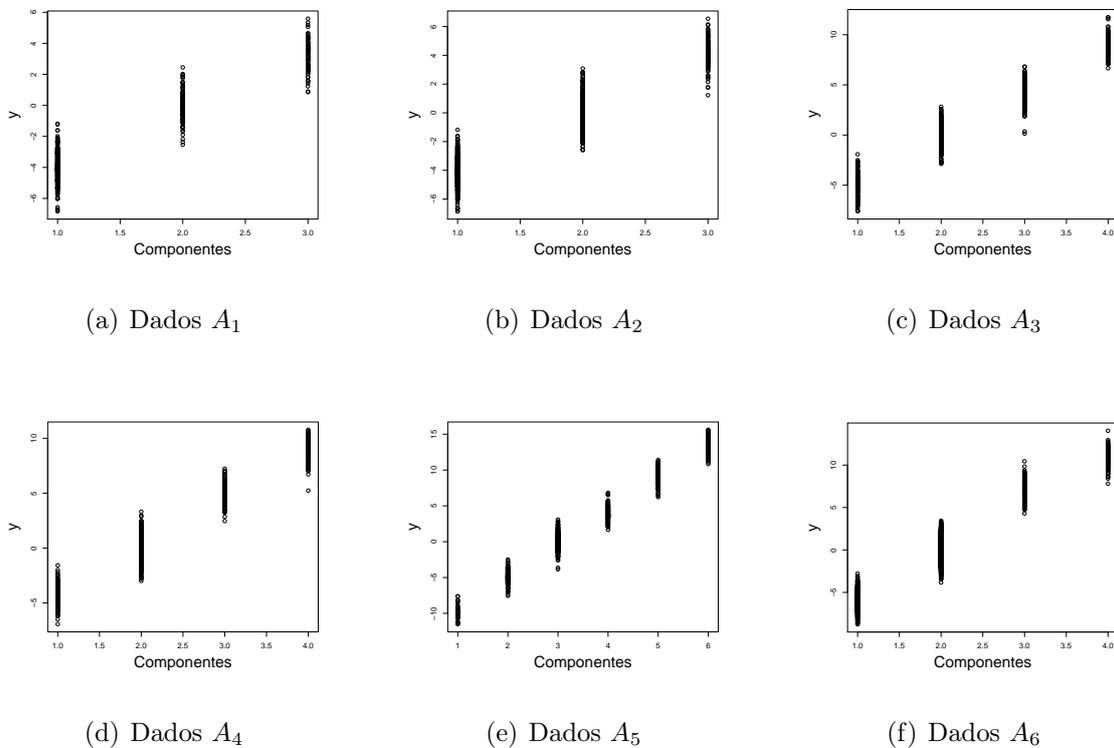


Figura 4.2: Valores gerados de cada componente

A Tabela 4.2 mostra o intervalo de variação dos dados observados, o intervalo (a, b) utilizado, o valor de Δ e as distribuições *a priori* com os valores dos hiperparâmetros para cada um dos conjuntos de dados simulados. As distribuições *a priori* apresentadas nesta tabela possuem “grandes” variâncias, como desejado. Por exemplo, para o conjunto de dados A_1 , gerando 1.000 valores da distribuição $\mathcal{IG}(2; 52)$, utilizando o *software R* fixando uma “semente” de valor 10 (*set.seed(10)*), obtemos uma média igual a 52,1926 e uma variância de 6.377,66.

Tabela 4.2: Valores de Δ e distribuições *a priori*

Dados	Número de componentes	Int. Obs.	(a, b)	Δ	Distribuições <i>a priori</i>
A_1	$K_{real} = 3$	(-6, 8351; 5.5861)	(-7, 6)	13	$\mu_k \cdot \sim \mathcal{N}(7, 5; 100)$ e $\sigma_k^2 \cdot \sim \mathcal{IG}(2; 52)$
A_2	$K_{real} = 3$	(-6, 8567; 6, 5434)	(-7, 7)	14	$\mu_k \cdot \sim \mathcal{N}(7, 0; 100)$ e $\sigma_k^2 \cdot \sim \mathcal{IG}(2; 56)$
A_3	$K_{real} = 4$	(-7, 5974; 11, 7385)	(-8, 12)	20	$\mu_k \cdot \sim \mathcal{N}(10; 100)$ e $\sigma_k^2 \cdot \sim \mathcal{IG}(2; 40)$
A_4	$K_{real} = 4$	(-6, 4870; 11, 3729)	(-7, 12)	19	$\mu_k \cdot \sim \mathcal{N}(9, 5; 100)$ e $\sigma_k^2 \cdot \sim \mathcal{IG}(2; 76)$
A_5	$K_{real} = 6$	(-11, 4953; 15, 6247)	(-12, 16)	28	$\mu_k \cdot \sim \mathcal{N}(14; 100)$ e $\sigma_k^2 \cdot \sim \mathcal{IG}(2; 112)$
A_6	$K_{real} = 4$	(-8, 7539; 14, 0418)	(-9, 15)	24	$\mu_k \cdot \sim \mathcal{N}(12; 100)$ e $\sigma_k^2 \cdot \sim \mathcal{IG}(2; 96)$

Para cada conjunto de dados, simulamos duas cadeias com $L = 55.000$ iterações, desconsiderando as primeiras 5.000 iterações de cada cadeia para eliminar o efeito dos valores iniciais, $L_{burn} = 5.000$. A primeira cadeia foi iniciada com $K = 1$ e a segunda cadeia com $K = 4$. Estes valores iniciais correspondem às imagens apresentadas na Figura 4.3. Para evitar problemas de autocorrelação, das 100.000 iterações provenientes da duas cadeias, consideramos um espaçamento de tamanho 50, obtendo uma amostra efetiva de tamanho 2.000, na qual a inferência *a posteriori* é baseada.



(a) Imagem inicial da cadeia 1

(b) Imagem inicial da cadeia 2

Figura 4.3: Imagens iniciais das cadeias 1 e 2

Para atualizar β na l -ésima iteração, utilizamos o algoritmo Metropolis-Hastings (ver p. 36) com valores candidatos gerados da distribuição gama tais que $\Gamma(a, b)$. Fixamos a e b de forma que $E(\beta^{(l)}) = \beta^{(l-1)}$ e $Var(\beta^{(l)}) = 0,1$. Assim, obtemos $a = 10(\beta^{(l-1)})^2$ e $b = 10\beta^{(l-1)}$, para $l = 2, \dots, L$, com $\beta^{(1)}$ sendo o valor inicial. Fixamos $\beta^{(1)} = 0,5$ para todos os conjuntos de dados.

A Tabela 4.3 apresenta as probabilidades *a posteriori* para o número K de componentes. O máximo *a posteriori* é obtido em $K = K_{real}$, destacado em negrito.

Tabela 4.3: Estimativas para K

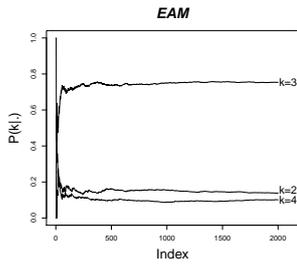
Dados	K_{real}	K	$P(k \cdot)$	Dados	K_{real}	K	$P(K \cdot)$
A_1	3	1	0,002	A_2	3	1	0,002
		2	0,138			2	0,190
		3	0,754			3	0,584
		4	0,101			4	0,204
		≥ 5	0,005			≥ 5	0,020
A_3	4	1	0,001	A_4	4	1	0,003
		2	0,096			2	0,164
		3	0,298			3	0,305
		4	0,505			4	0,351
		5	0,094			5	0,156
A_5	6	≥ 6	0,006	A_6	4	≥ 6	0,021
		≤ 2	0,042			≤ 2	0,072
		3	0,128			3	0,285
		4	0,150			4	0,452
		5	0,224			5	0,152
A_5	6	6	0,410	A_6	4	6	0,034
		7	0,045			7	0,004
		≥ 8	0,001			≥ 8	0,001

Obtidas as estimativas das probabilidades *a posteriori* de K , verificamos a convergência da sequencia de probabilidades de K ao longo das iterações, a capacidade do método de “visitar” diferentes valores de K ao longo das iterações e a estimativa da autocorrelação.

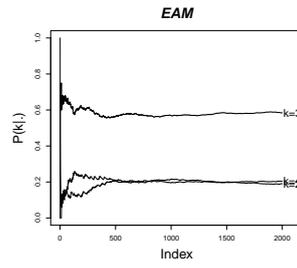
A Figura 4.4 mostra o gráfico de $P(K|\cdot)$ ao longo das iterações. Para manter uma boa visualização no gráfico de $P(K|\cdot)$, apresentamos somente as três maiores probabilidades. Como podemos observar nestas figuras, o número de iterações e o *burn in* utilizados são satisfatórios para se obter estabilidade para $P(K|\cdot)$. Além disso, não detectamos influência dos valores iniciais nas estimativas de K . Iniciando o método com $K = 1$ ou $K = 4$, o método permanece, na maioria das iterações, na vizinhança de K_{real} .

O método também se mostra capaz de “visitar” diferentes valores de K ao longo das iterações, como mostrado na Figura 4.5. Nestas figuras, cada símbolo “o” repre-

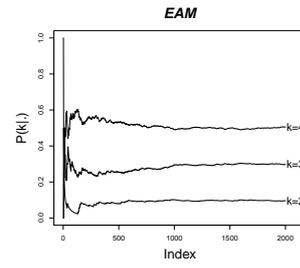
seja um valor de K amostrado.



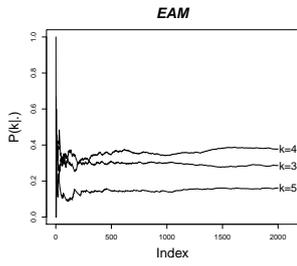
(a) Prob. K, A_1



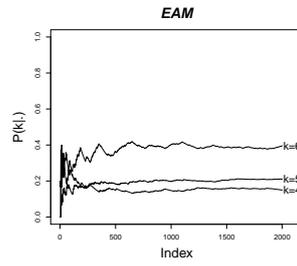
(b) Prob. K, A_2



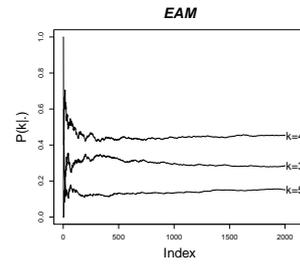
(c) Prob. K, A_3



(d) Prob. K, A_4

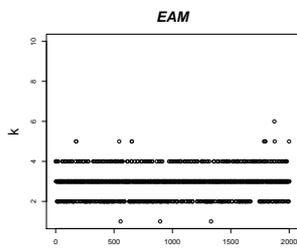


(e) Prob. K, A_5

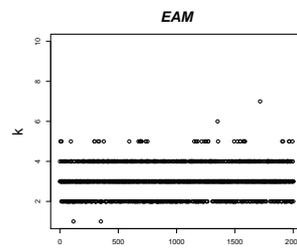


(f) Prob. K, A_6

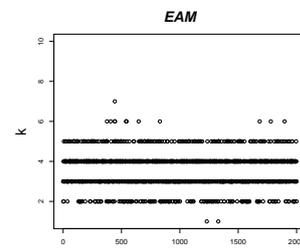
Figura 4.4: Probabilidade de K ao longo das iterações



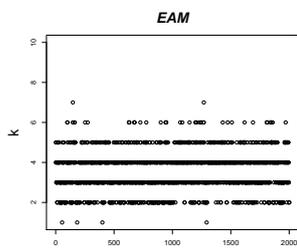
(a) Valores de K, A_1



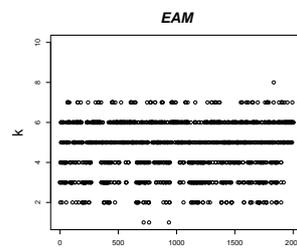
(b) Valores de K, A_2



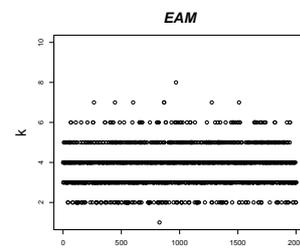
(c) Valores de K, A_3



(d) Valores de K, A_4



(e) Valores de K, A_5



(f) Valores de K, A_6

Figura 4.5: Valores de K amostrados

A Figura 4.6 mostra o gráfico das autocorrelações estimadas. Como as autocorrelações estimadas não são significativas, isto também indica a convergência da cadeia.

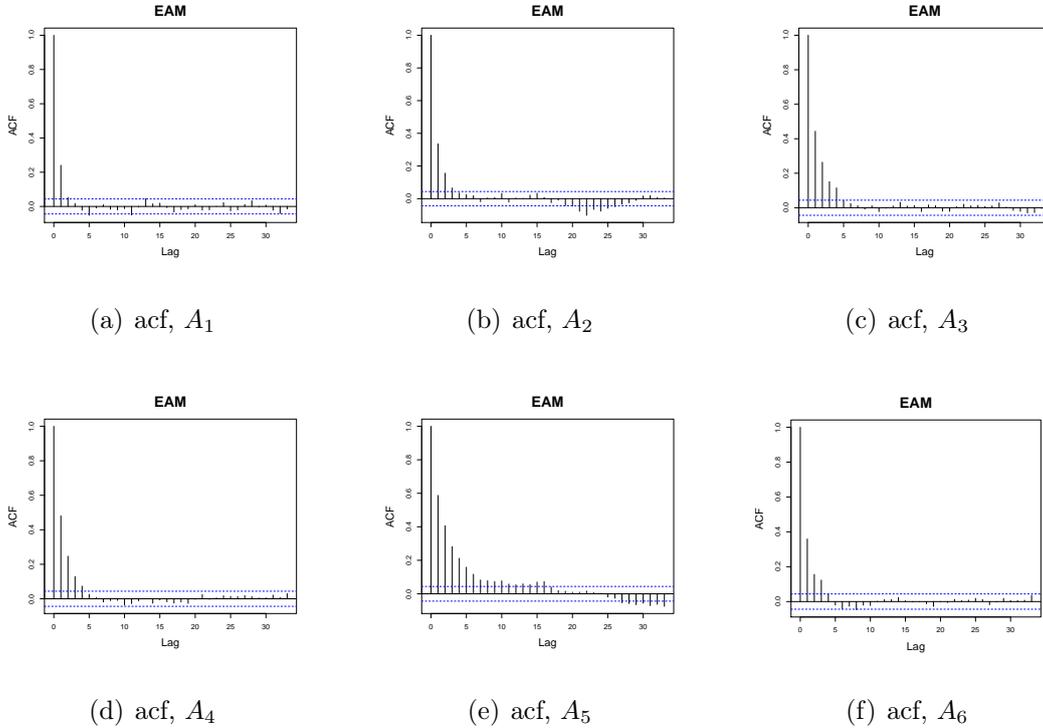


Figura 4.6: Autocorrelação estimada

A Tabela 4.4 mostra a porcentagem de movimentos EA aceitos. Como consideramos distribuições *a priori* não informativas, com grandes variâncias, então consideramos estas porcentagens de aceitação satisfatórias.

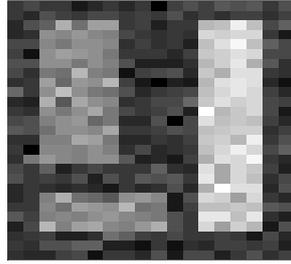
Tabela 4.4: Percentual de movimentos EA aceitos

Movimento	Dados simulados					
	A_1	A_2	A_3	A_4	A_5	A_6
Ejeção	4,430%	6,213%	6,933%	4,690%	7,418%	9,492%
Absorção	4,107%	6,163%	6,783%	4,636%	7,380%	9,356%

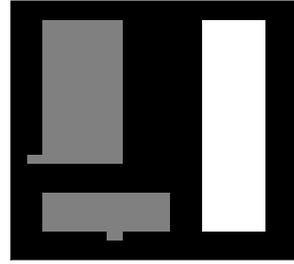
As Figuras 4.7 e 4.8 mostram as imagens geradas sem o ruído (esquerda), imagens geradas com o ruído (centro) e as imagens identificadas pelo método EAM (direita).



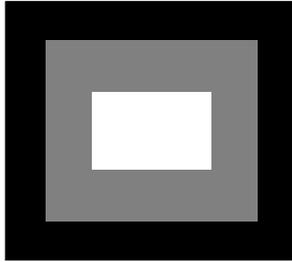
(a) A_1 , Imagem gerada



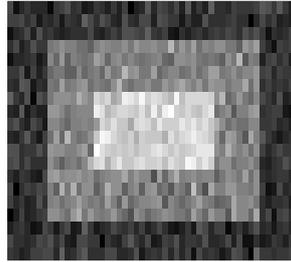
(b) A_1 , Imagem c/ ruído



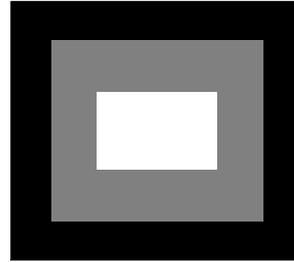
(c) A_1 , Imagem estimada



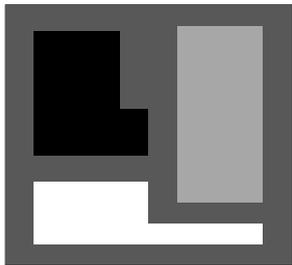
(d) A_2 , Imagem gerada



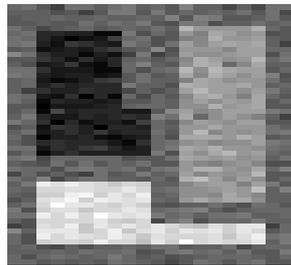
(e) A_2 , Imagem c/ ruído



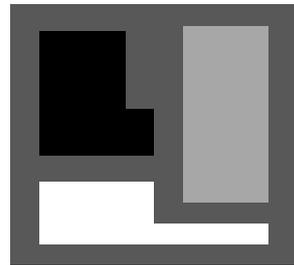
(f) A_2 , Imagem estimada



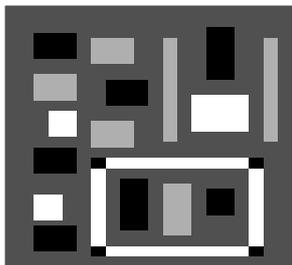
(g) A_3 , Imagem gerada



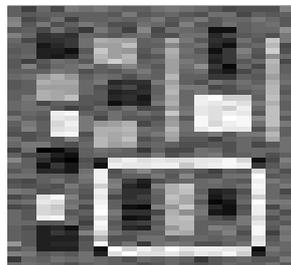
(h) A_3 , Imagem c/ ruído



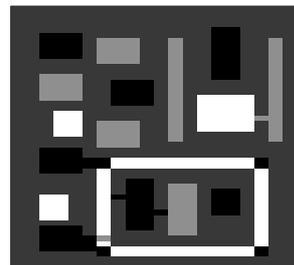
(i) A_3 , Imagem estimada



(j) A_4 , Imagem gerada



(k) A_4 , Imagem c/ ruído



(l) A_4 , Imagem estimada

Figura 4.7: Imagens estimadas pelo EAM-MCMC, conjuntos de dados A_1 a A_4

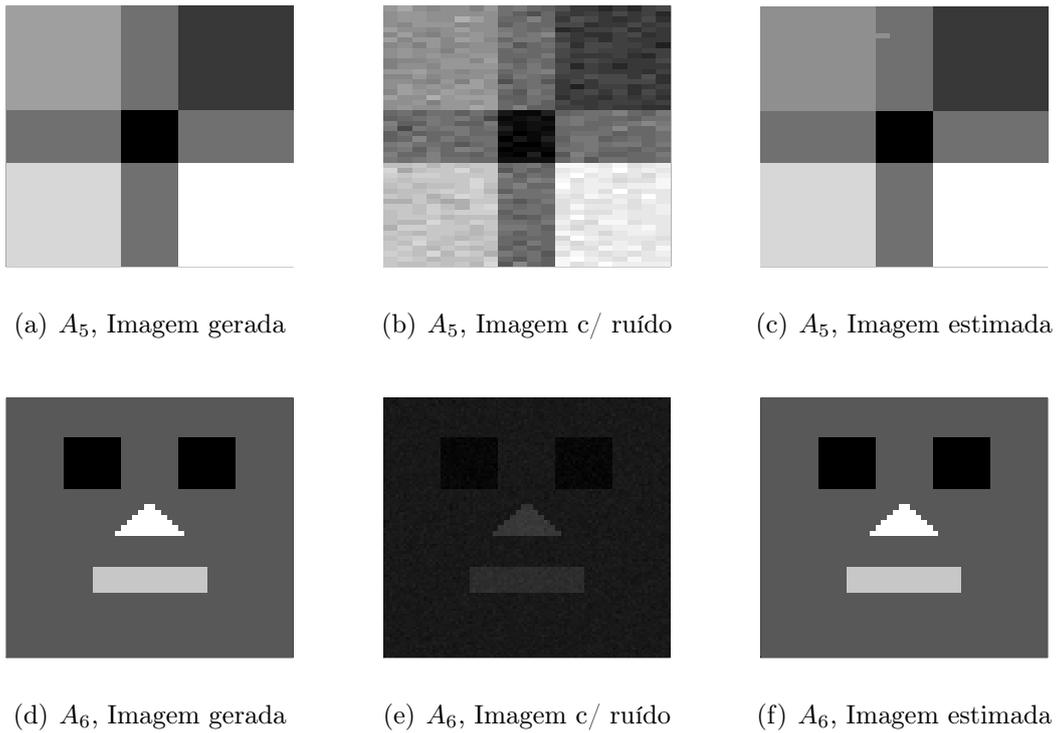


Figura 4.8: Imagens estimadas pelo EAM-MCMC, conjuntos de dados A_5 e A_6

Como podemos notar, o método apresenta bons resultados. As imagens nas Figuras 4.7(d,g) e 4.8(d) são perfeitamente identificadas. As outras imagens são satisfatoriamente estimadas, alterando a cor original de apenas alguns *pixels*.

A Tabela 4.5 mostra a estimativa e o intervalo de 95% de credibilidade para o parâmetro β . Na terceira coluna desta tabela também é apresentada a porcentagem de valores de β aceitos nas $L - L_{burn}$ iterações.

Tabela 4.5: Estimativa e intervalo de credibilidade de 95% para β

Dados	Estimativa	Int. Cred. 95%	% aceitação
A_1	1,214	(0,983; 1,773)	11,873%
A_2	1,356	(1,047; 2,039)	12,713%
A_3	1,455	(1,143; 2,197)	12,763%
A_4	1,180	(0,909; 1,669)	10,936%
A_5	1,719	(1,329; 2,494)	10,854%
A_6	1,801	(1,500; 2,000)	3,292%

Como a densidade geradora de candidatos para β é dada pela própria distribuição *a priori*, consideramos esta taxa de aceitação satisfatória. Esta porcentagem de

aceitação pode ser aumentada se houver a disponibilidade de informação de especialistas para fixar os valores dos hiperparâmetros a e b .

4.2 Aplicação

Nesta seção, aplicamos o EAM a um conjunto de dados reais. Este conjunto de dados é proveniente de uma imagem de ressonância magnética do cérebro de um adulto na presença de um tumor. A imagem foi obtida no website <http://neurosurgery.ufl.edu/patient-care/diseases-conditions/adult-brain-tumors/>.

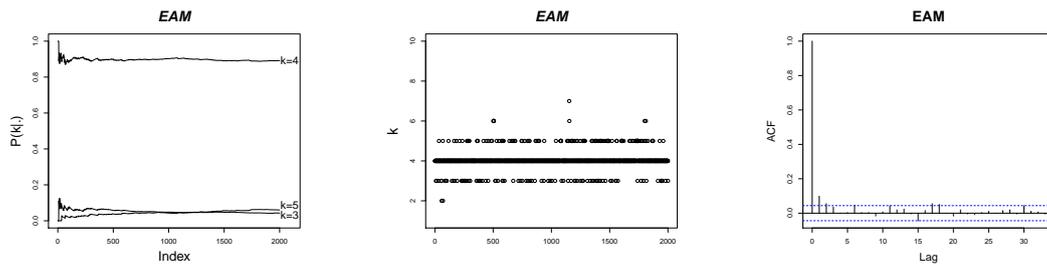
Esta imagem está definida sobre uma grade \mathcal{G} de dimensão 179×210 . Ou seja, a imagem é composta por $n = 37.590$ *pixels*. Para obter os níveis de intensidade de tons de cinza dos n *pixels*, utilizamos o pacote *jpeg* e o comando *readJPEG* do *software R*.

Obtida a matriz com as intensidades dos tons de cinza da imagem, aplicamos o EAM, utilizando $L = 55.000$ iterações e $L_{burn} = 5.000$. Porém, aqui utilizamos apenas uma cadeia, iniciada com $K = 1$. Além disso, para evitar problemas de autocorrelação, também consideramos um espaçamento de tamanho 25, obtendo uma amostra efetiva de tamanho 2.000 sobre a qual a inferência *a posteriori* é baseada.

As estimativas das probabilidades *a posteriori* para $K = 1, 2, 3, 4, 5$ e $K \geq 6$ são dadas por 0, 0,001, 0,044, 0,892, 0,060 e 0,004, respectivamente. O máximo *a posteriori* é dado em $K = 4$ com $P(K = 4|\cdot) = 0,892$.

A Figura 4.9 apresenta a performance do EAM. Como podemos observar, o método apresenta estabilidade para $P(K|\cdot)$ ao longo das iterações e a amostra de K é não correlacionada, indicando a convergência do método. Além disso, o método é capaz de “visitar” valores K de 2 até 7 ao longo das iterações.

A Figura 4.10 mostra a imagem original e a imagem identificada pelo método EAM. Como podemos notar, a figura identificada pelo EAM ressalta as cores preto, cinza escuro, cinza claro e branco da figura original. Estes resultados podem contribuir para a identificação de regiões passíveis de melhores estudos e análises clínicas.

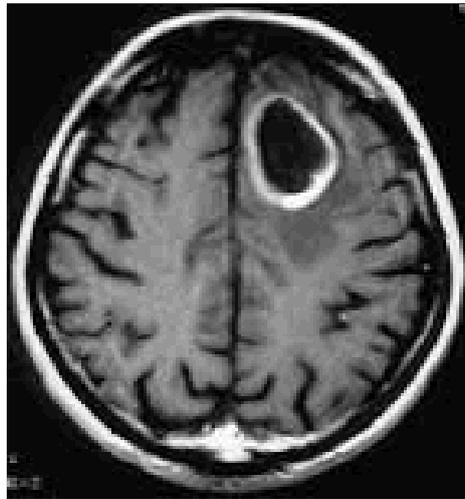


(a) $P(K|\cdot)$

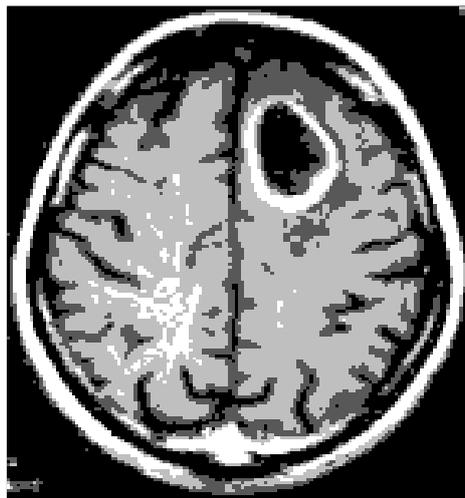
(b) Amostra de K

(c) acf.

Figura 4.9: Performance do EAM



(a) Imagem real



(b) Imagem por EAM

Figura 4.10: Imagem real e imagem identificada pelo método EAM

Capítulo 5

Considerações finais

Nesta pesquisa, desenvolvemos um novo algoritmo MCMC para segmentação de imagens com número de componentes desconhecido. Denominamos o algoritmo proposto de ejeção-absorção metropolizado (EAM). O termo “metropolizado” vem do fato de utilizarmos uma parametrização aumentada, como proposto por Carlin e Chib (1995), Godsill (2001) e Dellaportas *et al.* (2002), para desenvolver o algoritmo em um espaço paramétrico de dimensão fixa, e assim obter uma probabilidade de aceitação que é dada pela probabilidade de aceitação Metropolis-Hastings. Isto torna o algoritmo eficiente, pois podemos utilizar a distribuição condicional dos parâmetros como distribuição geradora de candidatos, e assim obter uma probabilidade de aceitação que depende apenas dos dados associados à(s) componente(s) selecionada(s) para uma ejeção (E) ou absorção (A).

Um outro diferencial do algoritmo proposto é que quando é “criada” uma nova componente, através de um movimento E ou A, este movimento determina uma nova partição na grade \mathcal{G} . Tal fato é devido à forma como implementamos os movimentos EA, baseando-se nos dados observados, ao invés dos parâmetros através da especificação de uma função de transição. Além disso, o método proposto é computacionalmente simples de ser implementado.

Verificamos a performance do algoritmo proposto utilizando seis conjuntos de dados simulados e um conjunto de dados reais. Os resultados obtidos mostram uma boa performance do algoritmo na estimação de K e reconstrução da imagem.

Na aplicação ao conjunto de dados reais, o algoritmo EAM identificou quatro componentes. Destacamos na reconstrução desta figura pelo EAM, a pequena região do lado superior direito, de cor predominantemente preta envolta por um anel de cor branca. Esta é uma das características de tumores cerebrais, cuja presença foi ressaltada pelo método.

Dessa forma, o método proposto pode contribuir com pesquisas, especialmente, na área médica, para a identificação de regiões passíveis de melhores estudos e análises clínicas. Além disso, os resultados obtidos pelo método podem ser utilizados para “criar” uma base de dados de imagens analisadas, destacando as regiões com anatomias normal e anormal. Os resultados poderão ser utilizados em outros estudos para identificação de anormalidades através de comparação.

Com o aumento do uso de computadores para obter informação de imagens, um prático e eficiente algoritmo computacional se torna necessário. O algoritmo EAM, resultado desta pesquisa, é um passo nesta direção. Os resultados obtidos com a aplicação do método às imagens simuladas e à imagem real mostram que o algoritmo pode ser utilizado como uma alternativa eficaz para a segmentação de imagens. O código fonte usado nas análises dos conjunto de dados foi desenvolvido em *software R* (*the Comprehensive R Archive Network*, <http://cran.r-project.org>) e está disponível mediante solicitação por e-mail ao autor.

Como proposta de pesquisa futura, destacamos o interesse em (i) flexibilizar o modelo bayesiano hierárquico proposto, assumindo o processo Dirichlet como distribuição *a priori* para os parâmetros das componentes; (ii) considerar o modelo desenvolvido e o modelo de misturas de processos Dirichlet com distribuições *a priori* e distribuição base sendo não conjugadas e (iii) estudar formas alternativas de marcação das componentes de modo a permitir a estimação das quantidades de interesse na presença do *label switching*.

Apêndice

A₁ - Integração termodinâmica

Integração termodinâmica vem do termo em inglês *thermodynamic integration* (TI), proposta por Ogata (1989) e é um método de integração numérica para funções com altas dimensões, baseado no algoritmo Metropolis-Hastings. O método reduz a integral de uma função, com dimensão alta, em uma integral unidimensional do valor esperado de uma função com relação a um parâmetro de escala definido em um intervalo $[a, b]$.

Utilizamos o método TI para estimar a constante de normalização $g(\beta, K)$ da distribuição de Gibbs, dada em (2.1), *i.e.*,

$$g(\beta, K) = \sum_{\mathbf{s} \in \Omega_{\mathbf{s}}} \exp \left\{ \beta \sum_{i \sim j} \mathbb{I}_{s_i}(s_j) \right\},$$

onde β é o parâmetro de escala definido no intervalo $[a, b]$. Em ambos os lados da Equação (2.1) aplicamos o logaritmo e, em seguida, diferenciamos em relação a β , temos

$$\begin{aligned} \frac{d}{d\beta} \log [g(\beta, K)] &= \frac{d}{d\beta} \log \left[\sum_{\mathbf{s} \in \Omega_{\mathbf{s}}} \exp \left\{ \beta \sum_{i \sim j} \mathbb{I}_{s_i}(s_j) \right\} \right] \\ &= \frac{1}{\sum_{\mathbf{s} \in \Omega_{\mathbf{s}}} \exp \left\{ \beta \sum_{i \sim j} \mathbb{I}_{s_i}(s_j) \right\}} \frac{d}{d\beta} \sum_{\mathbf{s} \in \Omega_{\mathbf{s}}} \exp \left\{ \beta \sum_{i \sim j} \mathbb{I}_{s_i}(s_j) \right\} \\ &= g^{-1}(\beta, K) \sum_{\mathbf{s} \in \Omega_{\mathbf{s}}} \left[\frac{d}{d\beta} \exp \left\{ \beta \sum_{i \sim j} \mathbb{I}_{s_i}(s_j) \right\} \right] \end{aligned}$$

$$\begin{aligned}
&= g^{-1}(\beta, K) \sum_{\mathbf{s} \in \Omega_{\mathbf{s}}} \left[\exp \left\{ \beta \sum_{i \sim j} \mathbb{I}_{s_i}(s_j) \right\} \sum_{i \sim j} \mathbb{I}_{s_i}(s_j) \right] \\
&= \sum_{\mathbf{s} \in \Omega_{\mathbf{s}}} \left[g^{-1}(\beta, K) \exp \left\{ \beta \sum_{i \sim j} \mathbb{I}_{s_i}(s_j) \right\} \sum_{i \sim j} \mathbb{I}_{s_i}(s_j) \right] \\
&= \sum_{\mathbf{s} \in \Omega_{\mathbf{s}}} \left[\pi(\mathbf{S} = \mathbf{s} | \beta, K) \sum_{i \sim j} \mathbb{I}_{s_i}(s_j) \right] \\
&= \mathbb{E}_{\mathbf{s} | \beta, K} \left[\sum_{i \sim j} \mathbb{I}_{s_i}(s_j) \right].
\end{aligned}$$

Concluimos,

$$\frac{d}{d\beta} \log [g(\beta, K)] = \mathbb{E}_{\mathbf{s} | \beta, K} \left[\sum_{i \sim j} \mathbb{I}_{s_i}(s_j) \right]. \quad (5.1)$$

Integrando os dois lados da Equação (5.1) em relação a β , temos

$$\log [g(\beta, K)] = n \log(K) + \int_a^b \mathbb{E}_{\mathbf{s} | \beta, K} \left[\sum_{i \sim j} \mathbb{I}_{s_i}(s_j) \right] d\beta, \quad (5.2)$$

onde $n \log(K)$ é o valor de $\log [g(0, K)]$.

Dessa forma, dado o valor de K , estimamos o valor do $\log [g(\beta, K)]$, calculando duas integrais em (5.2), uma interna (ou esperança) e outra externa.

Para estimar a integral interna, utilizamos um algoritmo MCMC, gerando N configurações \mathbf{s} através do algoritmo Swendsen-Wang, conforme descrito na Seção 2.3.3. Então, aproximamos o valor esperado $\mathbb{E}_{\mathbf{s} | \beta, K}$ pela média dos N valores $\sum_{i \sim j} \mathbb{I}_{s_i}(s_j)$.

Para estimar a integral externa, utilizamos métodos numéricos, como por exemplo, o método do trapézio. Para isto, consideramos uma grade com T valores de β , $0 \leq \beta_1, \dots, \beta_K \leq \beta$. Para cada β_t , $t = 1, \dots, T$, realizamos N simulações do grafo através do algoritmo Swendsen-Wang e estimamos conforme descrito no parágrafo acima. Então, aproximamos a integral externa pelo método dos trapézios.

A_2 - Densidade conjunta de (y, u) no algoritmo *slice sampling*.

Considere uma variável aleatória Y com densidade $\pi(y)$ proporcional a uma função $f(y)$. Então, considerando a constante normalizadora η , $\eta = \int f(y)dy$, temos

$$\pi(y) = \frac{f(y)}{\eta}. \quad (5.3)$$

Seja U uma variável auxiliar com densidade uniforme no intervalo $(0, f(y))$. Então, dado y , a densidade condicional de u é dada por

$$\pi(u|y) = \frac{1}{f(y)}. \quad (5.4)$$

Considerando as equações (5.3) e (5.4), temos

$$\pi(y, u) = \pi(u|y)\pi(y) = \frac{1}{f(y)} \frac{f(y)}{\eta} = \frac{1}{\eta}.$$

Então, a densidade conjunta de (Y, U) é dada por

$$\pi(y, u) = \begin{cases} 1/\eta, & \text{se } 0 < u < f(y) \\ 0, & \text{caso contrário} \end{cases},$$

onde $\eta = \int f(y) dy$.

Observe que se calcularmos a densidade marginal de Y , voltamos à distribuição em (5.3). De fato,

$$\pi(y) = \int_0^{f(y)} \pi(y, u) du = \int_0^{f(y)} \frac{1}{\eta} du = \frac{f(y)}{\eta}.$$

A₃ - Cálculo do conjunto \mathbb{S} do Exemplo 3.

No Exemplo 3 do Capítulo 2, aplicamos o algoritmo *slice sampler* para gerar valores de uma distribuição Gama truncada no intervalo (a, b) . O algoritmo consiste em dois passos, gerar $\mathbf{u}|y$ e gerar $y|\mathbf{u}$.

Iniciamos o algoritmo selecionando aleatoriamente $y^{(1)} \in (a, b)$. Para a l -ésima iteração, $l = 2, \dots, L$, geramos o valor $u_k^{(l)}$ da distribuição uniforme $\mathcal{U}_k(0, f_k(y^{(l-1)}))$, $k = 1, 2$, e em seguida geramos o valor $y^{(l)}$ da distribuição condicional $Y|u_1^{(l)}, u_2^{(l)} \sim \mathcal{U}(\mathbb{S}^{(l)})$, onde $\mathbb{S}^{(l)} = \{y : f_1(y) \geq u_1^{(l)}, f_2(y) \geq u_2^{(l)}, a < y < b\}$.

Dado $\mathbf{u}^{(l)} = (u_1^{(l)}, u_2^{(l)})$, a densidade de $Y^{(l)}|\mathbf{u}^{(l)}$ é dada pela distribuição uniforme $\mathcal{U}(\mathbb{S}^{(l)})$. Então

$$\begin{aligned} \pi(y|\mathbf{u}) &\propto \mathbb{I}_{\{f_1(y) \geq u_1^{(l)}, f_2(y) \geq u_2^{(l)}, a \leq y \leq b\}}(y) \\ &= \mathbb{I}_{\{y^{\alpha-1} \geq u_1^{(l)}, \exp\{-\beta y\} \geq u_2^{(l)}, a \leq y \leq b\}}(y) \\ &= \mathbb{I}_{\left\{y \geq (u_1^{(l)})^{\frac{1}{\alpha-1}}, -\beta y \geq \log(u_2^{(l)}), a \leq y \leq b\right\}}(y), \end{aligned}$$

onde $u_1^{(l)} \geq 0$ e $u_2^{(l)} > 0$. Logo,

$$\pi(y|\mathbf{u}) \propto \mathbb{I}_{\left\{y \geq (u_1^{(l)})^{\frac{1}{\alpha-1}}, y \leq -\frac{1}{\beta} \log(u_2^{(l)}), a \leq y \leq b\right\}}(y),$$

onde $u_1^{(l)} \geq 0$, $u_2^{(l)} > 0$, $\alpha > 1$ e $\beta > 0$. Observe que

$$y \geq (u_1^{(l)})^{\frac{1}{\alpha-1}} \quad \text{e} \quad y \geq a.$$

Por outro lado,

$$y \leq -\frac{1}{\beta} \log(u_2^{(l)}) \quad \text{e} \quad y \leq b.$$

Assim, temos $\mathbb{S}^{(l)} = (a', b')$, para $a', b' \in \mathbb{R}$, $a' < b'$, $a' = \max\left\{a, (u_1^{(l)})^{\frac{1}{\alpha-1}}\right\}$ e $b' = \min\left\{b, -\frac{1}{\beta} \log(u_2^{(l)})\right\}$. Ou seja, a distribuição condicional de $Y|\mathbf{u}^{(l)}$ é dada pela distribuição uniforme no intervalo (a', b') , *i.e.*,

$$Y|\mathbf{u}^{(l)} \sim \mathcal{U}(a', b').$$

A₄ - Cálculo do conjunto \mathbb{S} do Exemplo 4.

No Exemplo 4 do Capítulo 2 aplicamos o algoritmo *slice sampler* para gerar valores de uma distribuição que é proporcional ao produto de três funções. Iniciamos o algoritmo selecionando aleatoriamente $y^{(1)} \in \mathbb{R}$. Para a l -ésima iteração, $l = 1, \dots, L$, geramos $u_k^{(l)}$ da distribuição uniforme $\mathcal{U}_k(0, f_k(y^{(l-1)}))$, com $k = 1, 2, 3$, e em seguida geramos $y^{(l)}$ da distribuição condicional $Y|u_1^{(l)}, u_2^{(l)}, u_3^{(l)} \sim \mathcal{U}(\mathbb{S}^{(l)})$, onde

$$\mathbb{S}^{(l)} = \left\{ y : f_1(y) \geq u_1^{(l)}, f_2(y) \geq u_2^{(l)}, f_3(y) \geq u_3^{(l)} \right\}.$$

Como existem as inversas das funções $f_1(\cdot)$, $f_2(\cdot)$ e $f_3(\cdot)$, então o conjunto $\mathbb{S}^{(l)}$ pode ser reescrito em função da variável auxiliar $u_1^{(l)}$, $u_2^{(l)}$ e $u_3^{(l)}$. Dado $\mathbf{u}^{(l)} = (u_1^{(l)}, u_2^{(l)}, u_3^{(l)})$, a densidade de $Y^{(l)}|\mathbf{u}^{(l)}$ é dada pela distribuição uniforme $\mathcal{U}(\mathbb{S}^{(l)})$,

$$\begin{aligned} \pi(y|\mathbf{u}) &\propto \mathbb{I}_{\left\{1+\sin^2(3y) \geq u_1^{(l)}, 1+\cos^4(5y) \geq u_2^{(l)}, \exp\left\{-\frac{y^2}{2}\right\} \geq u_3^{(l)}\right\}}(y) \\ &= \mathbb{I}_{\left\{\sin^2(3y) \geq u_1^{(l)} - 1, \cos^4(5y) \geq u_2^{(l)} - 1, -\frac{y^2}{2} \geq \log(u_3^{(l)})\right\}}(y) \\ &= \mathbb{I}_{\left\{|\sin(3y)| \geq (u_1^{(l)} - 1)^{\frac{1}{2}}, |\cos(5y)| \geq (u_2^{(l)} - 1)^{\frac{1}{4}}, y^2 \leq -2 \log(u_3^{(l)})\right\}}(y) \\ &= \mathbb{I}_{\left\{|\sin(3y)| \geq (u_1^{(l)} - 1)^{\frac{1}{2}}, |\cos(5y)| \geq (u_2^{(l)} - 1)^{\frac{1}{4}}, |y| \leq (-2 \log(u_3^{(l)}))^{\frac{1}{2}}\right\}}(y), \end{aligned}$$

onde $u_1^{(l)} \geq 1$, $u_2^{(l)} \geq 1$ e $0 \leq u_3^{(l)} \leq 1$. Portanto, a distribuição condicional de $Y|\mathbf{u}^{(l)}$ é dada pela distribuição uniforme no intervalo

$$\left\{ y : |\sin(3y)| \geq (u_1^{(l)} - 1)^{\frac{1}{2}} \right\} \cap \left\{ y : |\cos(5y)| \geq (u_2^{(l)} - 1)^{\frac{1}{4}} \right\} \cap \left\{ y : |y| \leq (-2 \log(u_3^{(l)}))^{\frac{1}{2}} \right\}.$$

Para maiores detalhes, ver Robert e Casela (2004).

A₅ - Distribuições condicionais

Neste apêndice, apresentamos as distribuições condicionais dadas em (3.9) a (3.12).

A distribuição *a posteriori* conjunta dos parâmetros, $(\boldsymbol{\theta}_K, \mathbf{S}, \beta, K)$ é dada por

$$\begin{aligned} \pi(\boldsymbol{\theta}_K, \mathbf{S}, \beta, K | \mathbf{y}) &\propto P(\mathbf{y} | \mathbf{s}, \boldsymbol{\theta}_K, K) \pi(\mathbf{S} = \mathbf{s} | \beta, K) \pi(\boldsymbol{\theta}_K | K) \pi(\beta) \pi(K) \\ &= \left[\prod_{k=1}^K \left(\prod_{y_i \in D_k} f(y_i | \theta_k) \right) \right] g^{-1}(\beta, K) \exp \left\{ \beta \sum_{i \sim j} \mathbb{I}_{s_i}(s_j) \right\} \\ &\times \prod_{k=1}^K (\sigma_k^2)^{-\frac{1}{2}} \exp \left\{ -\frac{\lambda}{2\sigma_k^2} (\mu_k - \mu_0)^2 \right\} (\sigma_k^2)^{-\left(\frac{\alpha}{2}+1\right)} \exp \left\{ -\frac{\gamma}{2\sigma_k^2} \right\} \\ &\times \beta^{a-1} \exp \{-b\beta\}, \end{aligned}$$

onde

$$\begin{aligned} \prod_{k=1}^K \left(\prod_{y_i \in D_k} f(y_i | \theta_k) \right) &\propto \prod_{k=1}^K \left(\prod_{y_i \in D_k} (\sigma_k^2)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2\sigma_k^2} (y_i - \mu_k)^2 \right\} \right) \\ &= \prod_{k=1}^K \left((\sigma_k^2)^{-\frac{n_k}{2}} \exp \left\{ -\frac{1}{2\sigma_k^2} \sum_{y_i \in D_k} (y_i - \mu_k)^2 \right\} \right). \end{aligned}$$

Dados $(\mathbf{y}, \mathbf{s}, K)$ e os hiperparâmetros $(\alpha, \gamma, \lambda, \mu_0)$, a distribuição condicional de $\theta_k = (\mu_k, \sigma_k^2)$, é dada por

$$\begin{aligned} \pi(\theta_k | \mathbf{y}, \mathbf{s}, K) &\propto (\sigma_k^2)^{-\frac{n_k}{2}} \exp \left\{ -\frac{1}{2\sigma_k^2} \sum_{y_i \in D_k} (y_i - \mu_k)^2 \right\} \\ &\times (\sigma_k^2)^{-\frac{1}{2}} \exp \left\{ -\frac{\lambda}{2\sigma_k^2} (\mu_k - \mu_0)^2 \right\} (\sigma_k^2)^{-\left(\frac{\alpha}{2}+1\right)} \exp \left\{ -\frac{\gamma}{2\sigma_k^2} \right\} \\ &\propto (\sigma_k^2)^{-\left(\frac{n_k+\alpha+1}{2}+1\right)} \exp \left\{ -\frac{1}{2\sigma_k^2} \sum_{y_i \in D_k} (y_i - \mu_k)^2 - \frac{\lambda}{2\sigma_k^2} (\mu_k - \mu_0)^2 \right\} \exp \left\{ -\frac{\gamma}{2\sigma_k^2} \right\} \\ &\propto (\sigma_k^2)^{-\left(\frac{n_k+\alpha+1}{2}+1\right)} \exp \left\{ -\frac{1}{2\sigma_k^2} \left[\sum_{y_i \in D_k} (y_i^2 - 2y_i\mu_k + \mu_k^2) + \lambda (\mu_k^2 - 2\mu_k\mu_0 + \mu_0^2) \right] \right\} \\ &\quad \times \exp \left\{ -\frac{\gamma}{2\sigma_k^2} \right\} \end{aligned}$$

$$\begin{aligned}
& \propto (\sigma_k^2)^{-\left(\frac{n_k+\alpha+1}{2}+1\right)} \exp \left\{ -\frac{1}{2\sigma_k^2} \left[\left(\sum_{y_i \in D_k} y_i^2 \right) - 2\mu_k \left(\sum_{y_i \in D_k} y_i \right) + n_k\mu_k^2 + \lambda\mu_k^2 \right. \right. \\
& \qquad \qquad \qquad \left. \left. - 2\mu_k\lambda\mu_0 + \lambda\mu_0^2 \right] \right\} \exp \left\{ -\frac{\gamma}{2\sigma_k^2} \right\} \\
& \propto (\sigma_k^2)^{-\left(\frac{n_k+\alpha+1}{2}+1\right)} \exp \left\{ -\frac{1}{2\sigma_k^2} \left[\mu_k^2(n_k + \lambda) - 2\mu_k \left(\sum_{y_i \in D_k} y_i + \lambda\mu_0 \right) \right] \right\} \\
& \qquad \qquad \qquad \times \exp \left\{ -\frac{1}{2\sigma_k^2} \left[\gamma + \left(\sum_{y_i \in D_k} y_i^2 \right) + \lambda\mu_0^2 \right] \right\} \\
& \propto (\sigma_k^2)^{-\left(\frac{n_k+\alpha+1}{2}+1\right)} \exp \left\{ -\frac{n_k + \lambda}{2\sigma_k^2} \left[\mu_k^2 - 2\mu_k \left(\frac{\sum_{y_i \in D_k} y_i + \lambda\mu_0}{n_k + \lambda} \right) + \left(\frac{\sum_{y_i \in D_k} y_i + \lambda\mu_0}{n_k + \lambda} \right)^2 \right] \right\} \\
& \qquad \qquad \qquad \times \exp \left\{ -\frac{1}{2\sigma_k^2} \left[\gamma + \left(\sum_{y_i \in D_k} y_i^2 \right) + \lambda\mu_0^2 - \frac{\left(\sum_{y_i \in D_k} y_i + \lambda\mu_0 \right)^2}{n_k + \lambda} \right] \right\} \\
& \propto (\sigma_k^2)^{-\left(\frac{n_k+\alpha+1}{2}+1\right)} \exp \left\{ -\frac{1}{2\left(\frac{\sigma_k^2}{n_k+\lambda}\right)} \left[\mu_k^2 - \left(\frac{\sum_{y_i \in D_k} y_i + \lambda\mu_0}{n_k + \lambda} \right) \right]^2 \right\} \\
& \qquad \qquad \qquad \times \exp \left\{ -\frac{1}{2\sigma_k^2} \left[\gamma + \left(\sum_{y_i \in D_k} y_i^2 \right) + \lambda\mu_0^2 - \frac{\left(\sum_{y_i \in D_k} y_i + \lambda\mu_0 \right)^2}{n_k + \lambda} \right] \right\}.
\end{aligned}$$

Portanto,

$$\sigma_k^2 | \alpha, \gamma, \lambda, \mathbf{s}, K, \mathbf{y} \sim \mathcal{IG} \left(\frac{\alpha + n_k + 1}{2}, \frac{\gamma}{2} + \frac{\sum_{y_i \in D_k} y_i^2 + \lambda\mu_0^2}{2} - \frac{\left(\sum_{y_i \in D_k} y_i + \lambda\mu_0 \right)^2}{2(n_k + \lambda)} \right) e$$

$$\mu_k | \mu_0, \sigma_k^2, \lambda, K, \mathbf{y} \sim \mathcal{N} \left(\frac{\sum_{y_i \in D_k} y_i + \lambda\mu_0}{n_k + \lambda}, \frac{\sigma_k^2}{n_k + \lambda} \right).$$

A distribuição condicional de β dado (\mathbf{s}, a, b, K) é dada por

$$\begin{aligned}\pi(\beta|\mathbf{s}, a, b, K) &\propto \pi(\mathbf{s}|\beta, K)\pi(\beta|a, b) \\ &\propto g^{-1}(\beta, K) \exp \left\{ \beta \sum_{i \sim j} \mathbb{I}_{s_i}(s_j) \right\} \beta^{a-1} \exp \{-b\beta\}.\end{aligned}$$

A distribuição condicional de \mathbf{S} dado $(\mathbf{y}, \boldsymbol{\theta}_K, \beta, K)$ é dada por

$$\begin{aligned}\pi(\mathbf{S} = \mathbf{s}|\mathbf{y}, \boldsymbol{\theta}_K, \beta, K) &\propto \pi(\mathbf{S} = \mathbf{s}|\beta, K)P(\mathbf{y}|\mathbf{s}, \boldsymbol{\theta}_K, K) \\ &\propto \exp \left\{ \beta \sum_{i \sim j} \mathbb{I}_{s_i}(s_j) \right\} \prod_{k=1}^K \left((\sigma_k^2)^{-\frac{n_k}{2}} \exp \left\{ -\frac{1}{2\sigma_k^2} \sum_{y_i \in D_k} (y_i - \mu_k)^2 \right\} \right).\end{aligned}$$

A_6 - Constante normalizadora em (3.17)

Nesta seção, apresentamos os cálculos para obtenção de $\mathbf{I}(\cdot)$ dada em (3.17).

Dado \mathbf{s} , temos $D_k = \{y_i : s_i = k\}$. A probabilidade conjunta de \mathbf{y} dado \mathbf{s} e K é dada por

$$\begin{aligned} P(\mathbf{y}|\mathbf{s}, K) &= \int \prod_{k=1}^K \left(\prod_{y_i \in D_k} f(y_i|\theta_k) \right) \pi(\theta_k) d\boldsymbol{\theta}_K \\ &= \prod_{k=1}^K \int \left(\prod_{y_i \in D_k} f(y_i|\theta_k) \right) \pi(\theta_k) d\boldsymbol{\theta}_K. \end{aligned}$$

Denotando cada uma das K integrais por $\mathbf{I}(D_k)$, *i.e.*,

$$\mathbf{I}(D_k) = \int \left(\prod_{y_i \in D_k} f(y_i|\theta_k) \right) \pi(\theta_k) d\boldsymbol{\theta}_K,$$

para $k = 1, \dots, K$, temos

$$\begin{aligned} \mathbf{I}(D_k) &= \int_{\sigma_k^2} \int_{\mu_k} \frac{1}{(2\pi\sigma_k^2)^{\frac{n_k}{2}}} \exp \left\{ -\frac{1}{2\sigma_k^2} \sum_{y_i \in D_k} (y_i - \mu_k)^2 \right\} \frac{\lambda^{\frac{1}{2}}}{(2\pi\sigma_k^2)^{\frac{1}{2}}} \\ &\quad \exp \left\{ -\frac{\lambda}{2\sigma_k^2} (\mu_k - \mu_0)^2 \right\} \frac{\alpha_2^{\alpha_1}}{\Gamma(\alpha_1)} (\sigma_k^2)^{(\alpha_1-1)} \exp \{-\alpha_2\sigma_k^2\} d\mu_k d\sigma_k^2 \\ &= \int_{\sigma_k^2} \frac{1}{(2\pi\sigma_k^2)^{\frac{n_k}{2}}} \frac{\lambda^{\frac{1}{2}}}{(2\pi\sigma_k^2)^{\frac{1}{2}}} \frac{\alpha_2^{\alpha_1}}{\Gamma(\alpha_1)} (\sigma_k^2)^{(\alpha_1-1)} \exp \{-\alpha_2\sigma_k^2\} \\ &\quad \int_{\mu_k} \exp \left\{ -\frac{1}{2\sigma_k^2} \sum_{y_i \in D_k} (y_i - \mu_k)^2 \right\} \exp \left\{ -\frac{\lambda}{2\sigma_k^2} (\mu_k - \mu_0)^2 \right\} d\mu_k d\sigma_k^2. \end{aligned}$$

Para facilitar a notação, chamamos de I_1 a integral em relação a μ_k e a calculamos separadamente. Assim,

$$\begin{aligned} I_1 &= \int_{\mu_k} \exp \left\{ -\frac{1}{2\sigma_k^2} \sum_{y_i \in D_k} (y_i - \mu_k)^2 \right\} \exp \left\{ -\frac{\lambda}{2\sigma_k^2} (\mu_k - \mu_0)^2 \right\} d\mu_k \\ &= \int_{\mu_k} \exp \left\{ -\frac{1}{2\sigma_k^2} \left[\sum_{y_i \in D_k} (y_i - \mu_k)^2 + \lambda(\mu_k - \mu_0)^2 \right] \right\} d\mu_k \end{aligned}$$

$$\begin{aligned}
&= \int_{\mu_k} \exp \left\{ -\frac{1}{2\sigma_k^2} \left[\sum_{y_i \in D_k} y_i^2 - 2\mu_k \sum_{y_i \in D_k} y_i + n_k \mu_k^2 + \lambda \mu_k^2 - 2\lambda \mu_k \mu_0 + \lambda \mu_0^2 \right] \right\} d\mu_k \\
&= \int_{\mu_k} \exp \left\{ -\frac{1}{2\sigma_k^2} \left[(n_k + \lambda) \mu_k^2 - 2\mu_k \left(\sum_{y_i \in D_k} y_i + \lambda \mu_0 \right) + \sum_{y_i \in D_k} y_i^2 + \lambda \mu_0^2 \right] \right\} d\mu_k \\
&= \int_{\mu_k} \exp \left\{ -\frac{n_k + \lambda}{2\sigma_k^2} \left[\mu_k^2 - 2\mu_k \frac{\sum_{y_i \in D_k} y_i + \lambda \mu_0}{n_k + \lambda} + \left(\frac{\sum_{y_i \in D_k} y_i + \lambda \mu_0}{n_k + \lambda} \right)^2 \right] \right\} \\
&\quad \exp \left\{ -\frac{n_k + \lambda}{2\sigma_k^2} \left[\frac{\sum_{y_i \in D_k} y_i^2 + \lambda \mu_0^2}{n_k + \lambda} + \left(\frac{\sum_{y_i \in D_k} y_i + \lambda \mu_0}{n_k + \lambda} \right)^2 \right] \right\} d\mu_k \\
&= \int_{\mu_k} \exp \left\{ -\frac{1}{2 \left(\frac{\sigma_k^2}{n_k + \lambda} \right)} \left(\mu_k - \frac{\sum_{y_i \in D_k} y_i + \lambda \mu_0}{n_k + \lambda} \right)^2 \right\} \\
&\quad \exp \left\{ -\frac{1}{2\sigma_k^2} \left[\sum_{y_i \in D_k} y_i^2 + \lambda \mu_0^2 + \frac{\left(\sum_{y_i \in D_k} y_i + \lambda \mu_0 \right)^2}{n_k + \lambda} \right] \right\} d\mu_k \\
&= \int_{\mu_k} \frac{1}{\left(2\pi \frac{\sigma_k^2}{n_k + \lambda} \right)^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2 \left(\frac{\sigma_k^2}{n_k + \lambda} \right)} \left(\mu_k - \frac{\sum_{y_i \in D_k} y_i + \lambda \mu_0}{n_k + \lambda} \right)^2 \right\} d\mu_k \\
&\quad \left(2\pi \frac{\sigma_k^2}{n_k + \lambda} \right)^{\frac{1}{2}} \exp \left\{ -\frac{1}{2\sigma_k^2} \left[\sum_{y_i \in D_k} y_i^2 + \lambda \mu_0^2 + \frac{\left(\sum_{y_i \in D_k} y_i + \lambda \mu_0 \right)^2}{n_k + \lambda} \right] \right\} \\
&= \left(\frac{2\pi \sigma_k^2}{n_k + \lambda} \right)^{\frac{1}{2}} \exp \left\{ -\frac{1}{2\sigma_k^2} \left[\sum_{y_i \in D_k} y_i^2 + \lambda \mu_0^2 + \frac{\left(\sum_{y_i \in D_k} y_i + \lambda \mu_0 \right)^2}{n_k + \lambda} \right] \right\}.
\end{aligned}$$

Portanto, $I_1 = \left(\frac{2\pi\sigma_k^2}{n_k + \lambda}\right)^{\frac{1}{2}} \exp \left\{ -\frac{1}{2\sigma_k^2} \left[\sum_{y_i \in D_k} y_i^2 + \lambda\mu_0^2 + \frac{\left(\sum_{y_i \in D_k} y_i + \lambda\mu_0\right)^2}{n_k + \lambda} \right] \right\}$.

Substituindo I_1 em $\mathbf{I}(D_k)$, temos

$$\mathbf{I}(D_k) = \int_{\sigma_k^2} (2\pi\sigma_k^2)^{-\frac{n_k}{2}} \frac{\lambda^{\frac{1}{2}}}{(2\pi\sigma_k^2)^{\frac{1}{2}}} \frac{\alpha_2^{\alpha_1}}{\Gamma(\alpha_1)} (\sigma_k^2)^{(\alpha_1-1)} \exp \left\{ -\alpha_2\sigma_k^2 \right\} \left(\frac{2\pi\sigma_k^2}{n_k + \lambda}\right)^{\frac{1}{2}} \exp \left\{ -\frac{1}{2\sigma_k^2} \left[\sum_{y_i \in D_k} y_i^2 + \lambda\mu_0^2 + \frac{\left(\sum_{y_i \in D_k} y_i + \lambda\mu_0\right)^2}{n_k + \lambda} \right] \right\} d\sigma_k^2$$

$$= \left(\frac{\lambda}{n_k + \lambda}\right)^{\frac{1}{2}} \frac{\alpha_2^{\alpha_1}}{\Gamma(\alpha_1)} \left(\frac{1}{2\pi}\right)^{\frac{n_k}{2}} \int_{\sigma_k^2} \left(\frac{1}{\sigma_k^2}\right)^{\frac{n_k}{2} + \alpha_1 - 1} \exp \left\{ -\frac{\alpha_2}{2\sigma_k^2} \left[1 + \frac{\sum_{y_i \in D_k} y_i^2 + \lambda\mu_0^2}{2\alpha_2} + \frac{\left(\sum_{y_i \in D_k} y_i + \lambda\mu_0\right)^2}{2\alpha_2(n_k + \lambda)} \right] \right\} d\sigma_k^2$$

$$= \left(\frac{\lambda}{n_k + \lambda}\right)^{\frac{1}{2}} \frac{\Gamma(\alpha_1 + \frac{n_k}{2})}{\Gamma(\alpha_1)} \left(\frac{1}{2\pi\alpha_2}\right)^{\frac{n_k}{2}} \left[1 + \frac{\sum_{y_i \in D_k} y_i^2 + \lambda\mu_0^2}{2\alpha_2} + \frac{\left(\sum_{y_i \in D_k} y_i + \lambda\mu_0\right)^2}{2\alpha_2(n_k + \lambda)} \right]^{(\alpha_1 + \frac{n_k}{2})} \int_{\sigma_k^2} \frac{\left[1 + \frac{\sum_{y_i \in D_k} y_i^2 + \lambda\mu_0^2}{2\alpha_2} + \frac{\left(\sum_{y_i \in D_k} y_i + \lambda\mu_0\right)^2}{2\alpha_2(n_k + \lambda)} \right]^{(\alpha_1 + \frac{n_k}{2})}}{\Gamma(\alpha_1 + \frac{n_k}{2})} \left(\frac{\alpha_2}{\sigma_k^2}\right)^{\frac{n_k}{2} + \alpha_1 - 1} \exp \left\{ -\frac{\alpha_2}{2\sigma_k^2} \left[1 + \frac{\sum_{y_i \in D_k} y_i^2 + \lambda\mu_0^2}{2\alpha_2} + \frac{\left(\sum_{y_i \in D_k} y_i + \lambda\mu_0\right)^2}{2\alpha_2(n_k + \lambda)} \right] \right\} d\sigma_k^2$$

$$= \left(\frac{\lambda}{n_k + \lambda} \right)^{\frac{1}{2}} \frac{\Gamma(\alpha_1 + \frac{n_k}{2})}{\Gamma(\alpha_1)} \left(\frac{1}{2\pi\alpha_2} \right)^{\frac{n_k}{2}} \left[1 + \frac{\sum_{y_i \in D_k} y_i^2 + \lambda\mu_0^2}{2\alpha_2} + \frac{\left(\sum_{y_i \in D_k} y_i + \lambda\mu_0 \right)^2}{2\alpha_2(n_k + \lambda)} \right]^{(\alpha_1 + \frac{n_k}{2})} .$$

Portanto,

$$I(D_k) = \lambda^* \Gamma^* \gamma^* \left[1 + \frac{\sum_{y_i \in D_k} y_i^2 + \lambda\mu_0^2}{2\alpha_2} + \frac{\left(\sum_{y_i \in D_k} y_i + \lambda\mu_0 \right)^2}{2\alpha_2(n_k + \lambda)} \right]^{(\alpha_1 + \frac{n_k}{2})} ,$$

onde $\lambda^* = \left(\frac{\lambda}{n_k + \lambda} \right)^{\frac{1}{2}}$, $\Gamma^* = \frac{\Gamma(\frac{\alpha + n_k}{2})}{\Gamma(\frac{\alpha}{2})}$ e $\gamma^* = \left(\frac{1}{\gamma\pi} \right)^{\frac{n_k}{2}}$.

Referências Bibliográficas

- Askari, G., Xu, A., Li, Y. & Alavipanah, S. K. (2013). Automatic determination of number of homogenous regions using RJMCMC. *Indian Society of Remote Sensing*, **41** (3), 509-521.
- Barker, S. A., Kokaram, A. C. & Rayner, P. J. W. (1997). Unsupervised image segmentation using Markov random field models. *Society of Photographic Instrumentation Engineers's 43rd Annual Meeting*, 19-24.
- Bhattacharya, S. (2008). Gibbs sampling based Bayesian analysis of mixture with unknown number of components. *Sankhyā*, **70**, 133-155.
- Carlin, B. P. & Chib, S. (1995). Bayesian model choice via Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society B*, **57** (3), 473-484.
- Celeux, G., Hurn, M., and Robert, C. P. (2000). Computational and Inferential difficulties with mixture posterior distributions. *Journal of the American Statistical Association*, **95**, 957-970.
- Chalmond, B. (1989). An iterative Gibbsian technique for reconstruction of M-ary images. *Pattern Recognition*, **22**, 747-761.
- Chib, S. & Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician*, **49** (4), 327-335.
- Dellaportas, P., Forster, J. J., and Ntzoufras, I. (2002). On Bayesian model and variable selection using MCMC. *Statistics and Computing*, **12**, 27-36.

- Dobra, A., Jones, B., Hans, C., Nevins, J. & West, M. (2004). Sparse graphical models for exploring gene expression data. *Journal of Multivariate Analysis*, **90**, 196-212.
- Dryden, I. L., Scarr, M. R. & Taylor, C. C. (2003). Bayesian texture segmentation of weed and crop images using reversible jump Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society C*, **52**, 31-50.
- Edwards, R. G. & Sokal, A. D. (1988). Generalization of the Fortuin-Kasteleyn-Swendsen-Wang representation and Monte Carlo algorithm. *Physical Review D*, **38** (6), 2009-2012.
- Feng, D. & Tierney, L. (2011). mritc: A Package for MRI Tissue Classification. *Journal of Statistical Software*, **44**, 1-20.
- Ferreira, A. R. (2009). Bayesian mixture models of variable dimension for image segmentation. *Computer Methods and Programs in Biomedicine*, **94** (1), 1-14.
- Frühwirth-Schnatter, S. (2001). Markov Chain Monte Carlo estimation of classical and dynamic switching and mixture models. *Journal of the American Statistical Association*, **96**, 194-209.
- Gelman, A. & Meng, X. (1998). Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Statistical Science*, **13** (2), 163-185.
- Godsill, S. J. (2001). On the Relationship between Markov Chain Monte Carlo Methods for Model Uncertainty. *Journal of Computational and Graphical Statistics*, **10**, 230-428.
- Gonzalez, R. C. & Woods, R. (1992). *Digital Image Processing*. Addison Weley.
- Green, P. J. & Richardson, S. (2002). Hidden Markov models and disease mapping. *Journal of the American Statistical Association*, **97**, 1055-1070.
- Hans, C., Dobra, A. & West, M. (2007). Shotgun stochastic search in regression with many predictors. *Journal of the American Statistical Association*, **102**, 507-516.

- Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57** (1), 97-109.
- Hermes, L, Zöllner, T. & Buhmann, J. M. (2002). Parametric distributional clustering for image segmentation. In *Lecture notes in computer science: Computer vision-ECCV 02*, **2352**, 577-591.
- Ising, E. (1925). Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik*, **31**, 253-258.
- Jasra, A., Holmes, C. C., and Stephens, D. A. (2005). Markov Chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling. *Statistical Science*, **20**, 50-67.
- Johnson, T. D. & Piert, M. (2009). A Bayesian analysis of dual autoradiographic images. *Computational Statistics and Data Analysis*, **53**, 4570-4583.
- Jones, B., Carvalho, C., Dobra, A., Hans, C., Carter, C. & West, M. (2005). Experiments in Stochastic Computation for High-dimensional Graphical Models. *Statistical Science*, **20**, 388-400.
- López-Rubio, E. (2010). Restoration of images corrupted by Gaussian and uniform impulsive noise. *Pattern Recognition*, **43**, 1835-1846.
- Neal, R. M. (2003). Slice sampling. *The Annals of Statistics*, **31** (3), 705-767.
- Ogata, Y. (1989). A Monte Carlo method for high dimensional integration. *Numerische Mathematik*, **55**, 137-157.
- Potts, R. B. (1952). Some generalized order-disorder transformations. *Proceedings of the Cambridge Philosophical Society*, **48**, 106-109.
- Puzicha, J., Hofman, T. & Buhmann, J. M. (1999). Histogram clustering for unsupervised segmentation and image retrieval. *Pattern Recognition Letters*, **20**, 899-909.

- Richardson, S. & Green, P. J. (1997). On Bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society B*, **59** (4), 731-792.
- Robert, C. P. & Casella, G. (2004). Monte Carlo statistical methods. *New York: Springer*.
- Samson, C., Blanc-Féraud, L., Aubert, G. & Zerubia, J. (2000). A variational model for image classification and retoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22**(5), 460-472.
- Shi, J. & Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22** (8), 888-905.
- Stephens, M. (2000). Bayesian Analysis of mixture models with an unknown number of components-an alternative to reversible jump method. *The Annals of Statistics*, **34**, 187-220.
- Stephens, M. (2000). Dealing with label switching in mixture models. *Journal of the Royal Statistical Society B*, **62**, 795-809.
- Swendsen, R. H. & Wang, J. S. (1987). Nonuniversal critical dynamics in Monte Carlo simulations. *Physical Review Letters*, **58**, 86-88.
- Unnikrishnan, R., Pantofaru, C. & Hebert, M. (2007). Toward objective evaluation of image segmentation algorithms. *IEEE Transactions on pattern analysis and machine intelligence*, **29** (6), 929-944.
- Won, C. S. & Derin, H. (1992). Unsupervised segmentation of noisy and textured images using Markov random fields. *Graphical Models and Image Processing*, **54**, 308-328.
- Zhang, X.; Johnson, T. D. & Little, R. J. A. (2007). A Bayesian image analysis of the change in tumor/brain contrast uptake induced by radiation via RJMCMC. *The University of Michigan Department of Biostatistics Working Paper Series*, **75**.

Zhang, X.; Johnson, T. D.; Little, R. J. A. & Cao, Y. (2010). Bayesian image analysis of radiation induced change in tumor vascular permeability. *Bayesian Analysis*, **5** (1), 189-212.