7 - Elementos básicos de inferência Bayesiana.

Prof. Vinícius D. Mayrink

EST171 - Estatística Espacial

Sala: 4073 Email: vdm@est.ufmg.br

1^o semestre de 2024

Alguns conceitos básicos da abordagem Bayesiana para a inferência estatística são tratados aqui. O foco é descrever os elementos necessários para a construção de um modelo Bayesiano e as técnicas utilizadas para extrair informação relevante dos dados.

Assuma que $X = (X_1, \ldots, X_n)$ representa uma amostra aleatória contendo observações (dados) relacionadas ao estudo em desenvolvimento.

A função conjunta $p(X|\theta)$ descreve o quão provável é esta amostra para diferentes valores do parâmetro de interesse θ. Esta função é muitas vezes denotada por $L(\theta; X)$ e conhecida como função de verossimilhança.

Desejamos estimar θ. Em uma análise Bayesiana, o pesquisador caracteriza probabilisticamente (subjetivamente) sua incerteza a respeito de θ antes que os dados sejam observados; somente o conhecimento prévio a respeito do problema em questão é levado em conta. Nesta etapa inicial, uma função de probabilidade ou função densidade $p(\theta)$ será escolhida para descrever esta incerteza inicial; a distribuição de probabilidade associada a $p(\theta)$ é denominada distribuição *a priori*.

Muita discussão surgiu a respeito da introdução de informação, diferente daquela trazida pelos dados, no procedimento de inferência. Hoje em dia, a aceitação desta abordagem é muito ampla no meio científico.

Entre os Bayesianos, a grande discussão envolve a especificação de distribuições *a priori* não informativas. Elas são usadas quando nenhum conhecimento prévio está disponível sobre o parâmetro de interesse, determinando que apenas a informação dos dados deve interferir na análise.

Exemplo: Uma moeda (ela pode ser viciada) é lançada 6 vezes. Desejamos avaliar a probabilidade de obtermos "cara" no 7^o lançamento. Considere $H =$ Histórico ou informação disponível. Notação: $X_i = 1$ se o i-ésimo lançamento resulta em "cara" (0 caso contrário).

Primeiro pensamento: $p(X_7 = 1|H_0) = 0.5$.

Informação adicional:
$H_1 = \{X_1 = 1, X_2 = 1, X_3 = 1, X_4 = 1, X_5 = 1, X_6 = 0\}$.

Desejamos obter: $p(X_7 = 1|H_1)$.

Suposição: A ordem de 1's e 0's nos dados não é relevante.
A informação amostral importante é: "1 coroa e 5 caras".

Modelo: $p(X_i|\theta) = \theta^{x_i}(1-\theta)^{1-x_i}$.

Função de verossimilhança:
$p(X_1 = 1, X_2 = 1, X_3 = 1, X_4 = 1, X_5 = 1, X_6 = 0|\theta) = \theta^5(1-\theta)^1$.

Estimativa de máxima verossimilhança: $\hat{\theta} = 5/6 = 0.83$.

$$
\begin{aligned}
p(X_7 = 1|H_1) &= \int_0^1 P(X_7 = 1, \theta|H_1)\ d\theta \\
&= \int_0^1 p(X_7 = 1|\theta, H_1)\ p(\theta|H_1)\ d\theta \\
&\quad \text{por independência teremos...} \\
&= \int_0^1 p(X_7 = 1|\theta)\ p(\theta|H_1)\ d\theta \\
&= \int_0^1 \theta\ p(\theta|H_1)\ d\theta \ = \ E(\theta|H_1)
\end{aligned}
$$

Distribuição *a priori*: $0 \leq \theta \leq 1$, portanto, assuma a distribuição Beta.

$p(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1}(1-\theta)^{b-1}$ para $0 \leq \theta \leq 1$ e $a, b > 0$.

Especificação *a priori* vaga: $a = b = 1 \quad \Rightarrow \quad \theta \sim U(0,1)$.

$$\begin{aligned} p(\theta|H_1) &= \frac{p(H_1,\theta)}{p(H_1)} = \frac{p(H_1|\theta)p(\theta)}{p(H_1)} \\ &\propto \theta^5(1-\theta)\ \theta^{a-1}(1-\theta)^{b-1} \\ &\propto \theta^{5+a-1}(1-\theta)^{1+b-1} \end{aligned}$$

este é o núcleo de uma Beta($5 + a$,$1 + b$).

Note que $p(H_1)$ não depende de θ (constante de proporcionalidade).

$$\begin{aligned} p(X_7 = 1|H_1) &= E(\theta|H_1) = \frac{(a+5)}{(a+b+6)} = \\ &= 6/8 = 0.75 \quad \text{se} \quad a = b = 1, \\ &= 7/10 = 0.70 \quad \text{se} \quad a = b = 2. \end{aligned}$$

Teorema de Bayes

Distribuição *a priori*: $p(\theta|H)$.

Função de verossimilhança: $p(X|\theta, H)$.

Distribuição *a posteriori*: $p(\theta|X, H) = \frac{p(\theta, X|H)}{p(X|H)} = \frac{p(X|\theta, H)\ p(\theta|H)}{p(X|H)}$

sendo $p(X|H) = \int_\Theta p(X, \theta|H)\ d\theta$.

Alternativamente, escrevemos $p(\theta|X) \propto p(X|\theta)\ p(\theta)$.

Veja que $p(\theta|X) = k\ p(X|\theta)\ p(\theta)$ com: $1 = \int_\Theta p(\theta|X)\ d\theta = k \int_\Theta p(X|\theta)\ p(\theta)\ d\theta$.
Portanto: $k^{-1} = p(X|H) = \int_\Theta p(X|\theta)\ p(\theta) d\theta = E_\theta[p(X|\theta)]$.
Esta é a distribuição preditiva (ou marginal) de X.

Especificação a priori. Seja θ uma quantidade desconhecida.

Se θ é discreto, uma probabilidade *a priori* para cada possível valor de θ poderá ser avaliada diretamente via p.m.f..

Para um espaço paramétrico contínuo, o conhecimento *a priori* sobre θ pode ser usado para especificar uma p.d.f. *a priori* com certa forma funcional. Exemplo:

- θ apresenta distribuição simétrica em relação à moda;
- sua p.d.f. decai rapidamente ao se distanciar da moda;
- regiões longe da moda apresentam probabilidades irrelevantes.

Estas suposições caracterizam uma Normal com hiperparâmetros especificados de acordo com H.

Se uma probabilidade nula é atribuída a um subconjunto dos possíveis θ, nenhuma informação observada irá mudar a especificação (o que é inadequado).

Regra de Cromwell (Lindley): Devemos sempre associar uma probabilidade *a priori* não nula para cada possível valor de θ, mesmo que alguns deles sejam julgados como muito improváveis.

É possível mostrar que se $(X|\theta) \sim N(\theta, \sigma^2)$ e $\theta \sim N(m, v)$, então $(\theta|X) \sim N(m^*, v^*)$. Portanto, em um modelo Normal, se iniciarmos com uma distribuição *a priori* Normal, teremos uma distribuição *a posteriori* Normal.

Cada nova observação Normal obtida determina mudanças nos parâmetros da distribuição *a posteriori* (que continua sendo Normal).

Definição: Seja $\mathscr{F} = \{p(X|\theta);\ \theta \in \Theta\}$ uma família de distribuições relacionadas à amostra X. A classe $\mathscr{C}$ de distribuições é dita ser uma família conjugada com respeito a $\mathscr{F}$ se para todo $p(X|\theta) \in \mathscr{F}$ e $p(\theta) \in \mathscr{C}$ temos $p(\theta|X) \in \mathscr{C}$.

Podemos dizer que a classe de distribuições Normais é uma família conjugada com relação à classe de distribuições amostrais Normais.

As principais famílias conjugadas

Distribuição Binomial: A família de distribuições Beta é conjugada com o modelo Bernoulli ou Binomial.

Distribuição $N(\mu, \sigma^2)$ com variância conhecida: A família Normal é conjugada com o modelo Normal. Para uma amostra de tamanho n temos:

$$L(\mu; X) \propto \exp\left\{-\frac{1}{2\sigma^2}\left[-2\mu\sum_{i=1}^{n} X_i + n\mu^2\right]\right\}.$$

Assumindo $\mu \sim N(m, v)$ teremos $(\mu|X) \sim N(m^*, v^*)$ sendo:

$$v^* = \left(\frac{n}{\sigma^2} + \frac{1}{v}\right)^{-1} \quad \text{e} \quad m^* = v^*\left(\frac{\sum_{i=1}^{n} X_i}{\sigma^2} + \frac{m}{v}\right).$$

Distribuição Poisson:

Suponha que $X = \{X_1, \ldots, X_n\}$ é uma amostra aleatória da Poisson(θ).

A verossimilhança será:

$$p(X|\theta) = \prod_{i=1}^{n} p(X_i|\theta) = \prod_{i=1}^{n} \frac{\theta^{X_i} \exp\{-\theta\}}{X_i!} \propto \theta^{\sum_{i=1}^{n} X_i} \exp\{-n\theta\}$$

A família de distribuições Gama é conjugada com o modelo Poisson.
Assuma $\theta \sim Ga(a, b)$, então $(\theta|X) \sim Ga(a^*, b^*)$, sendo:

$$a^* = a + \sum_{i=1}^{n} X_i \quad \text{e} \quad b^* = b + n.$$

Distribuição Exponencial:

Suponha que $X = \{X_1, \ldots, X_n\}$ é uma amostra aleatória da Exp(θ). A verossimilhança será:

$$p(X|\theta) = \prod_{i=1}^{n} p(X_i|\theta) = \prod_{i=1}^{n} \theta \exp\{-\theta X_i\} \propto \theta^n \exp\left\{-\theta \sum_{i=1}^{n} X_i\right\}.$$

A família de distribuições Gama é conjugada com o modelo Exponencial. Assuma $\theta \sim Ga(a, b)$, então $(\theta|X) \sim Ga(a^*, b^*)$, sendo:

$$a^* = a + n \quad \text{e} \quad b^* = b + \sum_{i=1}^{n} X_i.$$

Distribuição Multinomial:

Suponha que $X = \{X_1, \ldots, X_q\}$ e $\theta = (\theta_1, \ldots, \theta_q)$ sejam, respectivamente, o número de casos observados e as probabilidades associadas a cada uma de q categorias; assuma tamanho amostral n. Considere $\sum_{i=1}^{q} X_i = n$ e $\sum_{i=1}^{q} \theta_i = 1$. Dizemos que X tem distribuição multinomial com parâmetros n e $(\theta_1, \ldots, \theta_q)$.

$$L(\theta; X) = p(X|\theta) = \frac{n!}{\prod_{i=1}^{q} X_i!} \prod_{i=1}^{q} \theta_i^{X_i} \ \propto \ \prod_{i=1}^{q} \theta_i^{X_i}.$$

Este núcleo é o mesmo da distribuição Dirichlet.

Se θ segue a distribuição Dirichlet com parâmetro $a = (a_1, \ldots, a_q)'$, sendo $a_i > 0$ e $\sum_{i=1}^{q} a_i = \alpha$, então sua p.d.f. será:

$$p(\theta) = \frac{\Gamma(\alpha)}{\prod_{i=1}^{q} \Gamma(a_i)} \prod_{i=1}^{q} \theta_i^{a_i - 1}.$$

A família Dirichlet com parâmetros $a = (a_1, \dots, a_q)$, notação $\text{Dir}(a)$, é conjugada com o modelo multinomial.

A distribuição *a posteriori* será então:

$$p(\theta|X) \propto \left[\prod_{i=1}^{q} \theta_i^{X_i}\right] \left[\prod_{i=1}^{q} \theta_i^{a_i - 1}\right] = \prod_{i=1}^{q} {\theta_i}^{X_i + a_i - 1}$$

a qual é uma Dirichlet com parâmetro $a^* = (a_1 + X_1, \dots, a_q + X_q)$ e denotada por $(\theta|X) \sim \text{Dir}(a^*)$.

A constante de proporcionalidade é dada por: $\frac{\Gamma(\alpha+n)}{\prod_{i=1}^{q} \Gamma(a_i+X_i)}$.

Esta análise conjugada generaliza a análise para amostras da Bernoulli com distribuição *a priori* Beta para a probabilidade de sucesso.

Distribuição $N(\mu, \sigma^2)$ com média conhecida e variância desconhecida:

Seja $X = (X_1, \ldots, X_p)$ uma amostra aleatória da $N(\mu, \sigma^2)$, sendo μ conhecido e $\phi = 1/\sigma^2$ a precisão desconhecida.

A maneira de parametrizar uma distribuição é escolhida de acordo com o interesse do pesquisador. Uma escolha de parametrização pode facilitar os cálculos, a interpretação e a implementação computacional de um problema.

Função de verossimilhança:

$$L(\phi; X) = p(X|\mu, \phi) \propto \phi^{n/2} \exp\left\{-\frac{\phi}{2}\left[\sum_{i=1}^{n} X_i^2 - 2\mu \sum_{i=1}^{n} X_i + n\mu^2\right]\right\}.$$

A distribuição *a priori* conjugada deve ter um núcleo parecido com a expressão de $L(\phi; X)$. Podemos considerar $\phi \sim Ga(a, b)$, então

$$\begin{aligned} p(\phi|X) &\propto L(\phi; X)\, p(\phi) \\ &\propto \phi^{n/2+a-1} \exp\left\{ -\phi \left[b + \frac{1}{2} \left(\sum_{i=1}^{n} X_i^2 - 2\mu \sum_{i=1}^{n} X_i + n\mu^2 \right) \right] \right\}. \end{aligned}$$

A última expressão é o núcleo da $Ga(a^*, b^*)$ sendo $a^* = a + n/2$ e $b^* = b + \frac{1}{2} \left(\sum_{i=1}^{n} X_i^2 - 2\mu \sum_{i=1}^{n} X_i + n\mu^2 \right)$.

Distribuição $N(\mu, \sigma^2)$ com média e variância desconhecidas:

Assuma que $X_1, \ldots, X_n$ é amostra aleatória da $N(\mu, \sigma^2)$, sendo $\phi = 1/\sigma^2$ a precisão. Neste caso, precisamos especificar uma distribuição *a priori* conjunta para (μ, ϕ). Isso será feito em dois estágios:

- $(\mu|\phi) \sim N[m, v/\phi]$;
- $\phi \sim Ga(a, b)$;

sendo $(m \in \mathbb{R}, v > 0, a > 0, b > 0)$ especificados de acordo com a informação inicial H.

Função de verossimilhança:

$$L(\mu, \phi; X) \propto \phi^{n/2} \exp\left\{-\frac{\phi}{2}\left[\sum_{i=1}^{n} X_i^2 - 2\mu \sum_{i=1}^{n} X_i + n\mu^2\right]\right\}.$$

Distribuição *a posteriori*:

$$
\begin{aligned}
p(\mu,\phi|X) &\propto p(X|\mu,\phi)\; p(\mu,\phi) \\
&\propto p(X|\mu,\phi)\; p(\mu|\phi)\; p(\phi) \\
&\propto \phi^{n/2} \exp\left\{-\frac{\phi}{2}\left[\sum_{i=1}^{n} X_i^2 - 2\mu\sum_{i=1}^{n} X_i + n\mu^2\right]\right\} \times \\
&\quad \times\; v^{-1/2}\, \phi^{1/2}\; \exp\left\{-\frac{\phi}{2v}[\mu^2 - 2\mu m + m^2]\right\} \times \\
&\quad \times\; \phi^{a-1} \exp\{-b\phi\} \\
&\quad \text{após alguns cálculos...} \\
&\propto \phi^{A-1} \exp\{-\phi B\} \quad W^{-1/2} \exp\left\{-\frac{1}{2W}(\mu - M)^2\right\}
\end{aligned}
$$

Na última expressão temos os núcleos:
$(\mu|\phi, X) \sim N(M, W)$ e $(\phi|X) \sim Ga(A, B)$ sendo:

$$M = \left(n + \frac{1}{v}\right)^{-1} \left[\frac{m}{v} + \sum_{i=1}^{n} X_i\right], \quad W = \frac{v}{(nv+1)\phi},$$
$$B = b + \frac{m^2}{2v} + \frac{\sum_{i=1}^{n} X_i^2}{2} - \frac{1}{2}\left(n + \frac{1}{v}\right) M^2 \quad \text{e} \quad A = a + \frac{n}{2}.$$

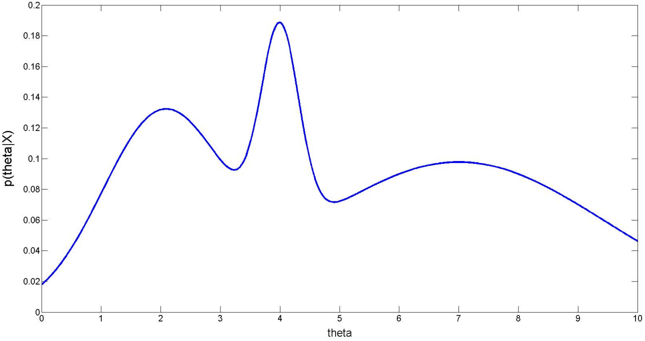
Conclusão: $(\mu|\phi, X) \sim N(M, W)$ e $(\phi|X) \sim Ga(A, B)$.

Neste caso, a distribuição *a priori* e *a posteriori* são denominadas "Normal-Gama".

A família de distribuições Normal-Gama é conjugada ao modelo Normal com média e a variância desconhecidas.

Estimação

Considere a p.d.f. *a posteriori* abaixo ilustrando algumas dificuldades que podemos enfrentar no procedimento de inferência Bayesiana. Apesar das dificuldades, temos aqui tudo que precisamos para uma descrição probabilística sobre θ.



O gráfico da p.d.f. *a posteriori* é a melhor descrição que temos no processo de inferência Bayesiana. Entretanto, algumas vezes é útil sumarizar ainda mais a informação *a posteriori*.

Estimação pontual: A forma mais simples de sumarização é a estimação pontual, onde buscamos determinar um único valor representando a quantidade desconhecida θ. Denote por $\hat{\theta}$ o estimador pontual de θ; devemos considerar estimadores que representam uma medida de centralidade. Três escolhas familiares são:

- Média *a posteriori*: $\hat{\theta} = E(\theta|X)$;
- Mediana *a posteriori*: $\hat{\theta}$ é tal que $\int_{-\infty}^{\hat{\theta}} p(\theta|X)d\theta = 0.5$;
- Moda *a posteriori*: $\hat{\theta}$ é tal que $p(\hat{\theta}|X) = \sup_{\theta} p(\theta|X)$.

Estimação intervalar: Seja θ uma quantidade desconhecida definida em Θ. Uma região $\mathscr{C} \subset \Theta$ é uma região de 100(1-α)% de credibilidade para θ se $P(\theta \in \mathscr{C}|X) \geq 1 - \alpha$. Neste caso, $1 - \alpha$ é o "nível de credibilidade". Se θ é um escalar, a região $\mathscr{C}$ poderá ser um intervalo do tipo $[c_1, c_2]$.

Exemplo: Seja $X_1, \ldots, X_n$ uma amostra aleatória tal que $X_i \sim N(\mu, \sigma^2)$ com variância σ^2 conhecida. Considere a distribuição *a priori*: $\mu \sim N[m, v]$ sendo m e v especificados pelo pesquisador.

Função de verossimilhança:

$$
\begin{aligned}
p(X|\mu) &= \prod_{i=1}^{n} (2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{1}{2\sigma^2}[X_i^2 - 2X_i\mu + \mu^2]\right\} \\
&= (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n} X_i^2 - 2\mu\sum_{i=1}^{n} X_i + n\mu^2\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\frac{n}{\sigma^2}\mu^2 - 2\mu\frac{\sum_{i=1}^{n} X_i}{\sigma^2}\right]\right\}
\end{aligned}
$$

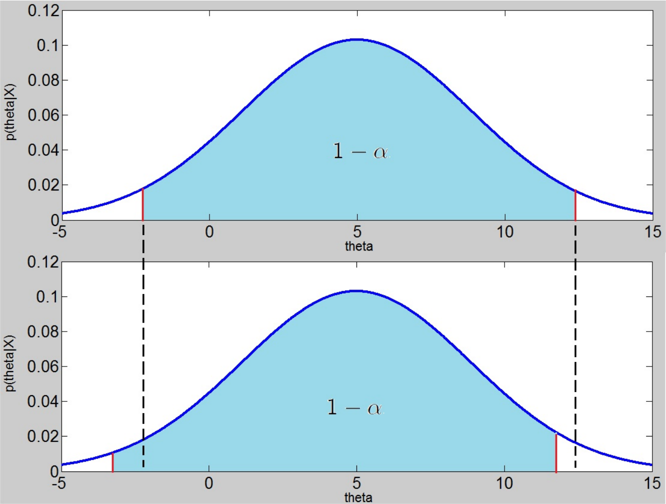
A distribuição *a posteriori* será dada por $p(\mu|X) \propto p(X|\mu)\; p(\mu)$:

$$\begin{aligned} p(\mu|X) &\propto \exp\left\{-\frac{1}{2}\left[\frac{n}{\sigma^2}\mu^2 - 2\mu\frac{\sum_{i=1}^n X_i}{\sigma^2}\right]\right\}\; \exp\left\{-\frac{1}{2}\left[\frac{1}{v}\mu^2 - 2\mu\frac{m}{v}\right]\right\}. \\ &\propto \exp\left\{-\frac{1}{2}\left(\frac{n}{\sigma^2}+\frac{1}{v}\right)\left[\mu^2 - 2\mu\left(\frac{n}{\sigma^2}+\frac{1}{v}\right)^{-1}\left(\frac{\sum_{i=1}^n X_i}{\sigma^2}+\frac{m}{v}\right)\right]\right\} \end{aligned}$$

A última expressão é o núcleo da $N(m^*, v^*)$ com
$v^* = \left(\frac{n}{\sigma^2}+\frac{1}{v}\right)^{-1}$ e $m^* = v^*\left(\frac{\sum_{i=1}^n X_i}{\sigma^2}+\frac{m}{v}\right)$.

Dado X, podemos escrever: $\dfrac{\mu - m^*}{\sqrt{v^*}} \sim N(0,1)$.

Muitos intervalos com 100(1-α)% de credibilidade podem ser construídos para μ.



O primeiro gráfico mostra um intervalo Bayesiano de credibilidade 100(1-α)% HPD (*Highest Posterior Density*) para o caso Normal. O segundo gráfico também mostra um intervalo de credibilidade 100(1-α)%, mas ele não é HPD. O intervalo HPD terá amplitude menor.

Computação Bayesiana - Gibbs Sampling

O Teorema de Bayes fornece $p(\theta|X) = \frac{p(X|\theta)\, p(\theta)}{p(X)}$ sendo $p(X)$ uma constante normalizadora que precisa ser determinada para que a expressão fechada da distribuição *a posteriori* seja conhecida.

Em muitos casos, a constante normalizadora não pode ser obtida (cálculos complicados). Conheceremos apenas o núcleo da distribuição alvo: $p(\theta|X) \propto p(X|\theta)\, p(\theta)$. Para superar este obstáculo, utilizaremos métodos computacionais que permitem amostrar indiretamente valores da distribuição *a posteriori*.

O *Gibbs Sampling* foi a primeira classe de esquemas largamente empregada para simulação estocástica via cadeia de *Markov*.

Origem: Discutido pela primeira vez em Geman e Geman (1984). Entretanto, Gelfand e Smith (1990) mostraram que o Gibbs poderia ser usado, em muitos casos, para amostrar da distribuição *a posteriori*.

Desejamos amostrar da distribuição de interesse $p(\theta|X)$, sendo $\theta = (\theta_1, \theta_2, \ldots, \theta_d)'$. A amostragem direta não é possível.

O Gibbs Sampling é baseado em gerações sucessivas a partir de distribuições condicionais completas $p(\theta_i|\theta_{-i})$ sendo $\theta_{-i} = (\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_d)'$.

O algoritmo:

1. Inicializar o contador de iterações ($j = 1$) e escolher os valores iniciais $\theta^{(0)} = (\theta_1^{(0)}, ..., \theta_d^{(0)})'$;
2. Obter o novo valor $\theta^{(j)} = (\theta_1^{(j)}, \ldots, \theta_d^{(j)})'$ a partir de gerações sucessivas considerando as distribuições condicionais completas:

$$\begin{aligned}
\theta_1^{(j)} &\sim p(\theta_1 \mid \theta_2^{(j-1)}, \ldots, \theta_d^{(j-1)}), \\
\theta_2^{(j)} &\sim p(\theta_2 \mid \theta_1^{(j)}, \theta_3^{(j-1)}, \ldots, \theta_d^{(j-1)}), \\
&\vdots \\
\theta_d^{(j)} &\sim p(\theta_d \mid \theta_1^{(j)}, \ldots, \theta_{d-1}^{(j)});
\end{aligned}$$

3. Faça $j = j + 1$ e retorne ao passo 2 até obter a amostra desejada.

Conforme j cresce a cadeia aproxima-se de uma condição de equilíbrio. Após algumas iterações (período de aquecimento - *burn in*) ela irá convergir e neste momento o valor $\theta^{(j)}$ será tratado como uma observação da distribuição $p(\theta|X)$.

A amostra gerada não é exatamente aleatória; haverá dependência entre os valores da cadeia. Mesmo que n seja grande, se a autocorrelação da cadeia é alta, o tamanho efetivo da amostra será menor que n. Estratégia para inferência: formar a amostra final com valores selecionados a cada L iterações ($L = lag$).

O *Gibbs Sampling* não é o único método da classe *Markov Chain Monte Carlo* (MCMC). Outro método MCMC popular é o *Metropolis-Hastings*.

Se alguma condicional completa é desconhecida, podemos utilizar outros métodos (*Metropolis-Hastings*, *Slice Sampling*, ARS - *Adaptive Rejaction Sampling*, etc). como um passo dentro do *Gibbs Sampling*. Esta estratégia é usada pelo *OpenBugs* (BUGS = *Bayesian inference Using Gibbs Sampling*).

Autocorrelação: Seja $z_1, \ldots, z_n$ uma série de observações. A autocorrelação estimada de ordem j é dada por: $\hat{\rho}_j = \left[\sum_{i=j+1}^{n} (z_i - \bar{z})\,(z_{i-j} - \bar{z})\right] / \sum_{i=1}^{n} (z_i - \bar{z})^2$.