

## 0.1 Inferência bayesiana aproximada em modelos gaussianos latentes

Para ilustrar as principais idéias da abordagem INLA para inferência bayesiana aproximada em modelos gaussianos latentes, considere o seguinte modelo. Seja  $y_i$  o número de casos de uma doença em uma área  $i$  para um determinado período de tempo. Assuma que as contagens  $y_i$  são observações condicionalmente independentes com distribuição Poisson, ou seja,

$$y_i \sim P(\exp(\eta_i))$$

onde  $\eta_i$  denota o log-risco relativo para a área  $i$ . Podemos assumir que o log-risco relativo seja decomposto como

$$\underline{\eta} = \mu \mathbf{1} + \underline{u} + \underline{v} \quad (1)$$

onde  $\mu$  é o intercepto,  $\underline{u}$  é um componente espacialmente estruturado, e  $\underline{v}$  é um componente não estruturado.

Para priori temos as seguintes suposições

1.  $\mu \sim N(0, \tau_\mu)$  em geral  $\tau_\mu = 0$
2.  $\underline{v} \sim N(0, \tau_v K_v)$  onde  $K_v = I$
3.  $\underline{u} \sim N(0, \tau_u K_u)$  onde  $K_u = n_i$  para  $i = j$ ,  $-1$  para  $i \sim j$  e 0 caso contrário.

Neste casos temos que  $\underline{\theta} = (\tau_u, \tau_v)^T$  denotam os parâmetros de precisão do modelo. A posteriori é dada por

$$\begin{aligned} \pi(\mu, \underline{u}, \underline{v}, \underline{\theta} | \underline{y}) \propto & \exp\left(-\frac{\tau_\mu}{2} \mu^2\right) \tau_v^{\frac{n}{2}} \left(-\frac{\tau_v}{2} \sum_i v_i^2\right) \tau_u^{\frac{(n-1)}{2}} \exp\left(-\frac{\tau_u}{2} \sum_{i,j} (u_i - u_j)^2\right) \\ & \exp\left(\sum_i y_i (\mu + u_i + v_i) - \exp(\mu + u_i + v_i)\right) \pi(\underline{\theta}) \end{aligned}$$

Para  $\pi(\underline{\theta})$  escolhemos priori's  $G(1, 0.01)$  independentes. Note que  $\underline{x} = (\mu, \underline{u}^T, \underline{v}^T)^T$  condicionado em  $\underline{\theta}$  é um CAMG. Além disso, verifica-se que  $y_i$  depende da soma dos três componentes de  $\underline{x}$ , devido a 1. Fazendo uma reparametrização para usar  $\underline{\eta}$  ao invés de  $\underline{v}$  temos

$$\underline{\eta} | \underline{u}, \underline{\theta}, \mu \sim N(\mu \mathbf{1} + \underline{u}, \tau_v I)$$

Com essa reparametrização  $\underline{x} = (\mu, \underline{u}^T, \underline{\eta}^T)^T$  e a posteriori pode ser escrita como

$$\pi(\underline{x}, \underline{\theta} | \underline{y}) \propto \tau_v^{\frac{n}{2}} \tau_u^{\frac{(n-1)}{2}} \exp\left(-\frac{1}{2} \underline{x}^T Q \underline{x}\right) \exp\left(\sum_i y_i \eta_i - \exp(\eta_i)\right) \pi(\underline{\theta})$$

onde

$$Q = \begin{pmatrix} \tau_m u + n \tau_v & \tau_v 1^T & -\tau_v 1^T \\ \tau_v 1 & \tau_u K_u + \tau_v I & -\tau_v I \\ -\tau_v 1 & -\tau_v I & \tau_v I \end{pmatrix}$$

Até aqui montamos a estrutura do modelo e escrevemos a posteriori conjunta de  $\underline{x}$  e  $\underline{\theta}$  de uma forma conveniente. Nesta matriz o termo espacialmente estruturado  $\underline{u}$  tem uma restrição de soma 0, ou seja,  $1^T \underline{u} = 0$ . A proposta inicial para a inferência através de MCMC seria usar um algoritmo *one-block* e gerar uma *proposal* conjunta usando

$$\begin{aligned} \tau_u^* &\sim q(\tau_u^* | \tau_u) \\ \tau_v^* &\sim q(\tau_v^* | \tau_v) \\ \underline{x}^* &\sim \pi(\underline{x} | \underline{\theta}^*, \underline{y}) \end{aligned}$$

e aceitar/rejeitar  $(\underline{\theta}^*, \underline{x}^*)$  conjuntamente, dentro de um algoritmo do tipo *Metropolis Hastings*.

A abordagem INLA trabalha diferente sem usar algoritmo de simulação. Basicamente desejamos obter as distribuições a posteriori, para o campo latente  $\underline{x}$  e para os parâmetros de precisão  $\underline{\theta}$ . O INLA trabalha usando que as marginais a posteriori de interesse podem ser escritas como,

$$\pi(x_i | \underline{y}) = \int \pi(x_i | \underline{\theta}, \underline{y}) \pi(\underline{\theta} | \underline{y}) d\underline{\theta}$$

e

$$\pi(\theta_j | \underline{y}) = \int \pi(\underline{\theta} | \underline{y}) d\underline{\theta}_{-j}$$

O fato chave da abordagem INLA é construir aproximações aninhadas para cada um dos componentes.

$$\tilde{\pi}(x_i | \underline{y}) = \int \tilde{\pi}(x_i | \underline{\theta}, \underline{y}) \tilde{\pi}(\underline{\theta} | \underline{y}) d\underline{\theta}$$

e

$$\tilde{\pi}(\theta_j | \underline{y}) = \int \tilde{\pi}(\underline{\theta} | \underline{y}) d\underline{\theta}_{-j}$$

O primeiro passo é usar a seguinte identidade

$$\pi(\underline{\theta}|\underline{y}) = \frac{\pi(\underline{y}|\underline{x}, \underline{\theta})\pi(\underline{x}|\underline{\theta})\pi(\underline{\theta})}{\pi(\underline{y})\pi(\underline{x}|\underline{\theta}, \underline{y})} \propto \frac{\pi(\underline{y}|\underline{x}, \underline{\theta})\pi(\underline{x}|\underline{\theta})\pi(\underline{\theta})}{\pi(\underline{x}|\underline{\theta}, \underline{y})} \quad (2)$$

O fato mais importante na equação é a parte inferior da equação que tem a seguinte forma,

$$\pi(\underline{x}|\underline{\theta}, \underline{y}) \propto \exp\left(-\frac{1}{2}\underline{x}^T \underline{Q}\underline{x} + \sum_i \log \pi(y_i|x_i, \underline{\theta})\right)$$

O núcleo da abordagem INLA é usar uma aproximação Gaussiana para a distribuição  $\pi(\underline{x}|\underline{\theta}, \underline{\theta})$ . Esta aproximação é denotada por  $\pi_G(\underline{x}|\underline{\theta}, \underline{\theta})$ , e é obtida obtendo a moda e a curvatura na moda de  $\pi(\underline{x}|\underline{\theta}, \underline{\theta})$ .

Para um dado  $\underline{\theta}$  sabemos construir uma aproximação gaussiana para a distribuição  $\pi(\underline{x}|\underline{\theta}, \underline{\theta})$ . Sendo assim, podemos usá-la em 2 para aproximar  $\pi(\underline{\theta}|\underline{y})$ . Note que a aproximação só é válida em um ponto

$$\tilde{\pi}(\underline{\theta}|\underline{y}) \propto \left. \frac{\pi(\underline{y}|\underline{x}, \underline{\theta})\pi(\underline{x}|\underline{\theta})\pi(\underline{\theta})}{\tilde{\pi}_G(\underline{x}|\underline{\theta}, \underline{y})} \right|_{\underline{x}=\underline{x}^*(\underline{\theta})}$$

Sendo assim, basta avaliar  $\tilde{\pi}(\underline{\theta}|\underline{y})$  para uma grade de  $\underline{\theta}$  que obteremos uma aproximação para  $\pi(\underline{\theta}|\underline{y})$ . No caso do exemplo temos  $\underline{\theta} = (\tau_u, \tau_v)$ , portanto  $\pi(\underline{\theta}|\underline{y})$  será bidimensional.

A  $\tilde{\pi}(\underline{\theta}|\underline{y})$  é usada para resolver três problemas no processo de inferência.

1. Integrar fora a incerteza com respeito a  $\underline{\theta}$  quando calculando a aproximação para a marginal posteriori do campo latente  $\tilde{\pi}(x_i|\underline{y})$ .
2. É usada para calcular uma aproximação para a verossimilhança marginal.
3. Finalmente para o cálculo de marginais posteriori para os hiperparâmetros  $\tilde{\pi}(\theta_m|\underline{y})$

Todos estes procedimentos envolvem integração numerica sobre um dominio multidimensional. Assim precisamos estar aptos a localizar a área de mais alta densidade em  $\tilde{\pi}(\underline{\theta}|\underline{y})$  e calcular a integral consistindo em uma grade de  $d$  pontos. Feito isso somos capazes de calcular  $\tilde{\pi}(\underline{\theta}|\underline{y})$  e possíveis marginais  $\tilde{\pi}(\theta_m|\underline{y})$ . Agora falta construir uma aproximação para  $\pi(x_i|\underline{y})$ . A estratégia geral é apresentada no seguinte algoritmo

1. Selecione um conjunto de  $\Theta = (\underline{\theta}_1, \dots, \underline{\theta}_k)$
2. Para  $k = 1$  até  $K$  faça

3. Calcule  $\tilde{\pi}(\underline{\theta}_k|y)$
4. Calcule  $\tilde{\pi}(x_i|\underline{\theta}_k, y)$  como uma função de  $x_i$
5. Fim para
6. Calcule  $\tilde{\pi}(x_i|y) = \sum_k \tilde{\pi}(x_i|\underline{\theta}_k, y) \tilde{\pi}(\underline{\theta}_k|y) \delta_k$

Para este algoritmo funcionar precisamos saber de duas coisas

1. Como selecionar um conjunto (possivelmente pequeno) de pontos  $\Theta = (\underline{\theta}_1, \dots, \underline{\theta}_k)$ .
2. Como construir uma boa aproximação para  $\pi(x_i|\underline{\theta}_k, y)$

O primeiro problema pode facilmente ser resolvido localizando uma grade sobre a área de mais alta densidade e calcular  $\tilde{\pi}(\underline{\theta}|y)$  em cada ponto desta grade. Em termos gerais imagine o exemplo dado anteriormente,  $\tilde{\pi}(\underline{\theta}|y)$  é bidimensional, suponha que tenha a seguinte forma:

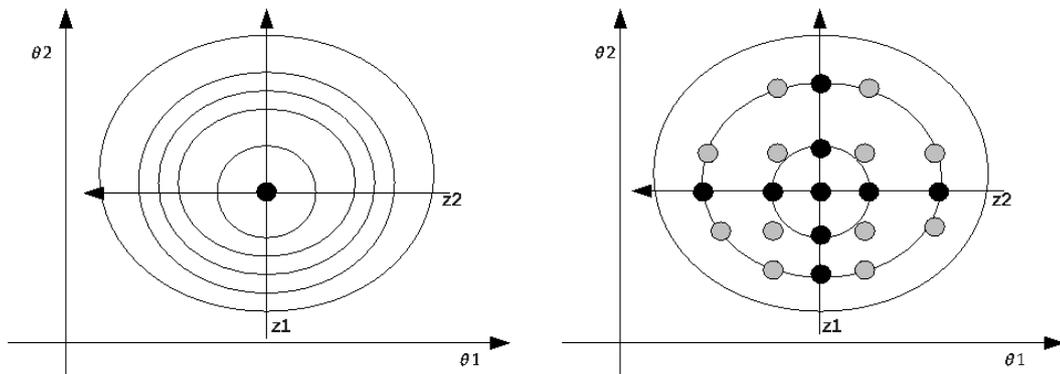


Figura 1: Ilustração da exploração da marginal posteriori para  $\theta$ .

O procedimento de exploração pode ser resumido como

1. Encontre seu ponto de máximo de  $\underline{\theta}^*$ .
2. Na moda calcule a Hessiana  $H > 0$ . Seja  $\Sigma = H^{-1}$ . Para facilitar a exploração use variáveis padronizadas  $z$  ao invés de  $\underline{\theta}$ . Seja  $\Sigma = V\delta V^T$  a decomposição em autovalores e autovetores de  $\Sigma$  e defina  $\underline{\theta}$  através de  $z$  com

$$\underline{\theta}(z) = \underline{\theta}^* + V\delta^{1/2}z$$

3. Explore a  $\log \tilde{\pi}(\underline{\theta}, \underline{y})$  usando a z-reparametrização

O próximo passo é aproximar  $\pi(x_i | \underline{\theta}, \underline{y})$ . Existem três propostas para esta aproximação.

- A aproximação Gaussiana  $\pi_G(x | \theta_k, \underline{y})$  já explicada.
- A aproximação de Laplace.
- A aproximação de Laplace Simplificada.

Neste momento vou explicar a de Laplace por ser a mais acurada. A forma natural de melhorar a aproximação Gaussiana é calcular a aproximação de Laplace.

$$\tilde{\pi}_{LA}(x_i | \underline{\theta}, \underline{y}) \propto \frac{\pi(y | \underline{\theta}, x) \pi(x | \underline{\theta}) \pi(\underline{\theta})}{\tilde{\pi}_{GG}(\underline{x}_{-i} | x_i, \underline{\theta}, \underline{y})} \Bigg|_{\underline{x}_{-i} = \underline{x}_{-i}^*(x_i, \underline{\theta})} \quad (3)$$

O problema com esta aproximação é que  $\tilde{\pi}_{GG}$  precisa ser recalculada para cada valor de  $x_i$  e  $\underline{\theta}$ . Isto é muito caro computacionalmente. A primeira modificação a ser feita consiste em evitar o passo de otimização no cálculo de  $\tilde{\pi}_{GG}(\underline{x}_{-i} | x_i, \underline{\theta}, \underline{y})$  aproximando a configuração modal por

$$\underline{x}_{-i}^* \approx E_{\tilde{\pi}_G}(\underline{x}_{-i} | x_i) \quad (4)$$

O lado direito é calculado sob a densidade condicional derivada vindo da aproximação gaussiana  $\tilde{\pi}_G(x | \underline{\theta}, \underline{y})$ . A segunda mudança materializa a seguinte intuição: Somente aqueles  $x_j$  que estão próximos de  $x_i$  podem ter impacto na marginal de  $x_i$ . Se a dependência entre  $x_j$  e  $x_i$  decai com o aumento da distância entre  $i$  e  $j$ , somente aqueles  $x_j$ 's em uma região de interesse acerca de  $i$ ,  $R_i(\underline{\theta})$ , determinam a marginal de  $x_i$ .

A esperança condicional em 4 implica que

$$\frac{E_{\tilde{\pi}_G}(x_j | x_i) - \mu_j(\underline{\theta})}{\sigma_j(\underline{\theta})} = a_{ij}(\underline{\theta}) \frac{x_i - \mu_i(\underline{\theta})}{\sigma_i(\underline{\theta})} \quad (5)$$

para algum  $a_{ij}(\underline{\theta})$  quando  $j \neq i$ . Assim, um critério para construir o conjunto  $R_i(\underline{\theta})$  é

$$R_i(\underline{\theta}) = (j : |a_{ij}(\underline{\theta})| > 0.001)$$

O importante é a economia computacional quando calculamos o denominador de 3, onde agora só é necessário fatorar uma matriz esparsa de dimensão  $R_i(\underline{\theta}) \times R_i(\underline{\theta})$ . A expressão 3 simplificada como explicado acima, precisa ser calculada para diferentes valores de  $x_i$  para

encontrar a densidade. Para selecionar estes pontos, usa-se a média e a variância da aproximação gaussiana, e escolhemos diferentes valores para as variáveis padronizadas

$$x_i^{(s)} = \frac{x_i - \mu_i(\underline{\theta})}{\sigma_i(\underline{\theta})}$$

Para a escolha das abscissas recorre-se a quadratura de Gauss-Hermite. Para representar a densidade  $\tilde{\pi}_{LA}(x_i|\underline{\theta}, \underline{y})$ , usa-se

$$\tilde{\pi}_{LA}(x_i|\underline{\theta}, \underline{y}) \propto N(x_i; \mu_i(\underline{\theta}), \sigma_i^2(\underline{\theta})) \times \exp(\text{cubic spline}(x_i)) \quad (6)$$

O spline cúbico é ajustado para a diferença da log-densidade de  $\tilde{\pi}_{LA}(x_i|\underline{\theta}, \underline{y})$  e  $\tilde{\pi}_G(x_i|\underline{\theta}, \underline{y})$  nas abscissas selecionadas, e então a densidade é normalizada usando integração por quadratura.