Métodos Computacionais para inferência estatística

Wagner Hugo Bonat

LEG: Laboratório de Estatística e Geoinformação
Universidade Federal do Paraná

30 de julho de 2012

Motivação

- Modelos estocásticos têm sido amplamente utilizados tanto na comunidade científica como no mundo dos negócios em geral.
- Estudos de mercado, predição em séries financeiras, análises de componentes de solo, mapeamento de doenças, entre outros.

Laboratório de Estatística e Geoinformação

Motivação

- Nesta diversidade de aplicações é fácil encontrar situações de relevância prática onde os modelos tradicionais (GLM) não são adequados.
- Em geral por pelo menos uma das seguintes características:

1. efeito não linear de covariáveis,
2. observações correlacionadas no espaço,
3. observações correlacionadas no tempo,
4. heterogeneidade entre unidades não explicada por covariáveis.

Laboratório de Estatística e Geoinformação

Modelos estruturados aditivamente

- Nestas situações a classe dos modelos de regressão estruturados aditivamente têm sido amplamente utilizada.

$$h(E(Y_i|\mu_i)) = \eta_i = \alpha + \sum_{j=1}^{n_f} f^{(j)}(u_{ji}) + \sum_{k=1}^{n_\beta} \beta_k z_{ki} + \epsilon_i \quad (1)$$

Modelos gaussianos latentes são um subconjunto de todos os modelos Bayesianos estruturados aditivamente, onde se supõe uma priori gaussiana para α,$f^{(j)}(.)$, β_k e ϵ_i.

Inferência Bayesiana

- Inferência baseada em simulação, MCMC *Markov Chain Monte Carlo*.
- Desempenho insatisfatório quando aplicado para modelos Gaussianos latentes, basicamente pela alta dependência entre os parâmetros.
- Apesar dos avanços com MCMC ele permanece lento e complicado do ponto de vista do usuário final.

Laboratório de Estatística e Geoinformação

Inferência Bayesiana - Outras abordagens

- *Variational Bayes* (BISHOP, 2006).
- *Expectation-Propagation* (MINKA, 2001).
- A ferramenta mais promissora parece ser *Integrated Nested Laplace approximation*-INLA (RUE et al, 2009).

Laboratório de Estatística e Geoinformação

Modelos Gaussianos Latentes

- Dados observados $\underline{y}$, $y_i|x_i \sim \pi(y_i|x_i, \theta)$
- Campo latente Gaussiano $\underline{x} \sim N(.; Q^{-1}(\underline{\theta}))$
- Hiperparâmetro $\underline{\theta}$
  1. variabilidade
  2. tamanho/força da dependência
  3. parâmetros na verossimilhança

Laboratório de Estatística e Geoinformação

Inferência bayesiana em MGL

- A posteriori pode ser escrita como

$$\pi(\underline{x}, \underline{\theta}|\underline{y}) \propto \pi(\underline{\theta})\pi(\underline{x}|\underline{\theta}) \prod_i \pi(y_i|x_i, \underline{\theta})$$

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

- As marginais posteriori de interesse podem ser escritas como

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

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

Laboratório de Estatística e Geoinformação

Inferência bayesiana aproximada para MLG

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

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

e

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

INLA - Integrated Nested Laplace approximation

- O primeiro passo é usar a seguinte identidade

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

- Importante é notar que

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

Laboratório de Estatística e Geoinformação

INLA - Integrated Nested Laplace approximation

- O núcleo do INLA é aproximar $\pi(\underline{x}|\underline{\theta},\underline{y})$ por $\pi_G(\underline{x}|\underline{\theta},\underline{y})$
- A aproximação usa a moda e a curvatura na moda de $\pi(\underline{x}|\underline{\theta},\underline{y})$.
- Válida em um ponto (moda), substituindo na equação (2) tem-se

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

Laboratório de Estatística e Geoinformação

INLA - Integrated Nested Laplace approximation

- Esta aproximação resolve três problemas no processo de inferência:

  1. Integrar fora a incerteza com respeito a $\underline{\theta}$, quando aproximando $\widetilde{\pi}(x_i|\underline{y})$.
  2. Calcular uma aproximação para a verossimilhança marginal.
  3. Marginais posteriori para os hiperparâmetros $\widetilde{\pi}(\theta_m|\underline{y})$.

- Integração numérica sobre um domínio multidimensional.

INLA - Estratégia geral

1. **Selecione um conjunto de** $\Theta = (\underline{\theta}_1, \ldots, \underline{\theta}_k)$
2. Para $k = 1$ até K faça
3. Calcule $\widetilde{\pi}(\underline{\theta}_k | \underline{y})$
4. **Calcule** $\widetilde{\pi}(x_i | \underline{\theta}_k, \underline{y})$ **como uma função de** x_i
5. Fim para
6. **Calcule** $\widetilde{\pi}(x_i | \underline{y}) = \sum_k \widetilde{\pi}(x_i | \underline{\theta}_k, \underline{y}) \widetilde{\pi}(\underline{\theta}_k | \underline{y}) \Delta_k$

Laboratório de Estatística e Geoinformação

INLA - Estratégia geral

- Para este algoritmo funcionar precisamos saber como obter duas quantidades

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

Laboratório de Estatística e Geoinformação

Explorando $\widetilde{\pi}(\underline{\theta}|\underline{y})$

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 Σ e defina $\underline{\theta}$ através de z com

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

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

Explorando $\widetilde{\pi}(\underline{\theta}|\underline{y})$

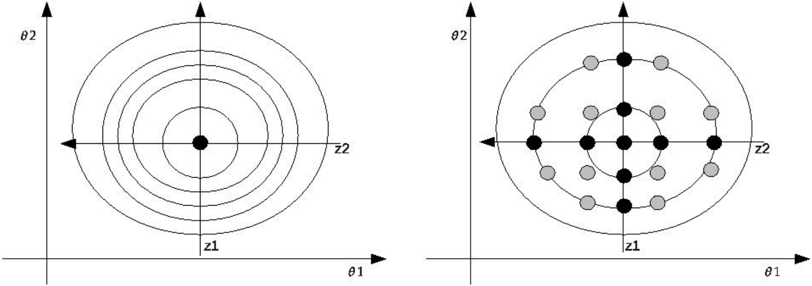

Figura 1: Ilustração da exploração da marginal posteriori para θ.

Laboratório de Estatística e Geoinformação

Aproximando $\pi(x_i|\underline{\theta}, \underline{y})$

- Rue et. al (2009) fazem três propostas
  1. Aproximação gaussiana $\widetilde{\pi}_G(x_i|\underline{\theta}, \underline{y})$ já explicada
  2. Aproximação de Laplace
  3. Aproximação de Laplace Simplificada

Aproximando $\pi(x_i|\underline{\theta}, \underline{y})$

$$\widetilde{\pi}_{LA}(x_i|\underline{\theta}, \underline{y}) \propto \left.\frac{\pi(\underline{y}|\underline{\theta}, \underline{x})\pi(\underline{x}|\underline{\theta})\pi(\underline{\theta})}{\widetilde{\pi}_{GG}(\underline{x}_{-i}|x_i, \underline{\theta}, \underline{y})}\right|_{\underline{x}_{-i}=\underline{x}^*_{-i}(x_i,\underline{\theta})} \tag{3}$$

- Muito cara computacionalmente
- Duas mudanças são propostas em Rue et. al (2009)
  1. $$\underline{x}^*_{-i} \approx E_{\widetilde{\pi}_G}(\underline{x}_{-i}|x_i) \tag{4}$$
  2. Somente alguns x_j próximos a x_i devem impactar na marginal de x_i.

Aproximando $\pi(x_i|\underline{\theta}, \underline{y})$

- A esperança condicional na equação (4) implica que

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

para algum $a_{ij}(\underline{\theta})$ quando $j \neq i$. Uma regra simples é

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

Laboratório de Estatística e Geoinformação

Aproximando $\pi(x_i|\underline{\theta}, \underline{y})$

- A expressão 3 ainda precisa ser calculada para diferentes valores de x_i.
- Para selecionar estes pontos usamos $\widetilde{\pi}_G(x_i|\underline{\theta}, \underline{y})$.

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

Para representar a densidade $\widetilde{\pi}_{LA}(x_i|\underline{\theta}, \underline{y})$, usa-se

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

Laboratório de Estatística e Geoinformação

INLA - Estratégia geral

1. Selecione um conjunto de $\Theta = (\underline{\theta}_1, \ldots, \underline{\theta}_k)$
2. Para $k = 1$ até K faça
3. Calcule $\widetilde{\pi}(\underline{\theta}_k|\underline{y})$
4. Calcule $\widetilde{\pi}(x_i|\underline{\theta}_k, \underline{y})$ como uma função de x_i
5. Fim para
6. Calcule $\widetilde{\pi}(x_i|\underline{y}) = \sum_k \widetilde{\pi}(x_i|\underline{\theta}_k, \underline{y})\widetilde{\pi}(\underline{\theta}_k|\underline{y})\Delta_k$

Laboratório de Estatística e Geoinformação

Exemplos computacionais com INLA

- Modelo de regressão Poisson

```
> dados.poisson <- read.table("../../POISSON.txt", header=TRUE)
> head(dados.poisson)

   y        cov
1  6 0.00000000
2 10 0.05050505
3  7 0.10101010
4 11 0.15151515
5 13 0.20202020
6  4 0.25252525
```

- Ajustando o modelo via INLA

```
> require(INLA)
> mod.poisson <- inla(y ~cov, family = "poisson", data = dados.poisson)
```

Exemplos computacionais com INLA

- Resumo do ajuste

```
> summary(mod.poisson)

Call:
"inla(formula = y ~ cov, family = \"poisson\", data = dados.poisson)"

Time used:
 Pre-processing    Running inla Post-processing           Total
     0.06723642      0.01298118      0.01254129      0.09275889

Fixed effects:
                 mean         sd 0.025quant  0.5quant 0.975quant          kld
(Intercept) 1.9737071 0.05108772  1.8727900 1.9739712  2.0731452 7.391718e-07
cov         0.5085499 0.01381904  0.4815489 0.5085135  0.5357629 1.190952e-09

The model has no random effects

The model has no hyperparameters

Expected number of effective parameters(std dev): 2.054(0.00)
Number of equivalent replicates : 48.68

Marginal Likelihood:  -297.05
Warning: Interpret the marginal likelihood with care if the prior model is improper.
```

Laboratório de Estatística e Geoinformação

Exemplos computacionais com INLA

- Modelo Beta Longitudinal

```
> dados.beta.long <- read.table("../../dadosBeta.txt", header=TRUE)
> i <- 1:150
> j <- 151:300
```

- Ajustando o modelo via INLA

```
> mod.beta.long <- inla(y ~cov + f(i, model="iid2d", n =300) +
+                       f(j, cov, copy="i"), family = "beta", data = dados.beta.long)
```

Exemplos computacionais com INLA

- Resumo do ajuste

```
> summary(mod.beta.long)

Call:
c("inla(formula = y ~ cov + f(i, model = \"iid2d\", n = 300) + f(j, ",  "    cov, copy = \"i\"), family =

Time used:
 Pre-processing    Running inla Post-processing           Total
     0.14096737      2.23673844      0.08089542      2.45860124

Fixed effects:
                 mean        sd 0.025quant  0.5quant 0.975quant        kld
(Intercept) 0.6760767 0.1169750  0.4472828 0.6757476  0.9067829 0.01866807
cov         0.8821051 0.2144836  0.4616026 0.8819220  1.3037362 0.05348653

Random effects:
Name          Model                  Max KLD
i   IID2D model
j   Copy

Model hyperparameters:
                                             mean    sd      0.025quant
precision parameter for the beta observations 10.3400  2.0234  7.0131
Precision for i (component 1)                 7.1439  2.9851  2.7291
Precision for i (component 2)                 3.8736  2.3724  0.8290
Rho1:2 for i                                 -0.2753  0.2660 -0.7579
                                             0.5quant 0.975quant
precision parameter for the beta observations 10.1147  14.9318
Precision for i (component 1)                 6.6963  14.2107
Precision for i (component 2)                 3.3884   9.7418
Rho1:2 for i                                 -0.2789   0.2448

Expected number of effective parameters(std dev): 52.78(16.02)
Number of equivalent replicates : 2.842

Marginal Likelihood:  -2754.65
Warning: Interpret the marginal likelihood with care if the prior model is improper.
```