MLG

Curso de Modelos Lineares Generalizado - DEST/UFMG
Marcos Oliveira Prates

17 de outubro de 2018

Conceitos basicos

- Esquema de inferência:
  - Obter informação sobre um parâmetro de interesse desconhecido $\theta = (\theta_1, \ldots, \theta_k)$
  - Usar a informação contida nas respostas **Y**
  - Incluir informação a priori
- Portanto temos o seguinte esquema estocástico:
  - $\pi(\theta)$ = distribuição de probabilidade a priori
  - $\pi(\mathbf{y}|\theta)$ = verossimilhança

Conceitos basicos

- Assim através do Teorema de Bayes obtemos

$$\begin{aligned} \pi(\theta|\mathbf{y}) &= \frac{\pi(\mathbf{y},\theta)}{\pi(\mathbf{y})} = \frac{\pi(\mathbf{y}|\theta)\pi(\theta)}{\pi(\mathbf{y})} \\ &\propto \pi(\mathbf{y}|\theta)\pi(\theta) \end{aligned}$$

- Infelizmente não é possível fazer inferência sem a constante de normalização
- Apenas em casos muito simples, conseguimos obter essa constante
- **Desafio como amostrar então da distribuição a posteriori de θ?**

- Veremos agora um exemplo desse tipo de método: **Amostrador de Gibbs ou Gibbs Sampling**.
- Nesse método amostramos um parâmetro, θ_i, de cada vez.
- Amostramos da distribuição condicional de um parâmetro quando os demais, $(\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, , \theta_k)$, são mantidos fixos.

- Seja (X, Y) um vetor aleatório.
- Vamos dedotar sua distribuição conjunta por

$$f_{X,Y}(x, y)\,.$$

- Como gerar amostras desse vetor?
- Podemos gerar X e Y independentemente, um ignorando o outro? Não.
- Suponha que sejamos capazes de encontrar as distribuições condicionais

$$f_{X|Y}(x|y) \quad f_{Y|X}(y|x)\,.$$

- Seja (X, Y) um vetor aleatório.
- Vamos dedotar sua distribuição conjunta por

$$f_{X,Y}(x, y)\,.$$

- Como gerar amostras desse vetor?
- Podemos gerar X e Y independentemente, um ignorando o outro? Não.
- Suponha que sejamos capazes de encontrar as distribuições condicionais

$$f_{X|Y}(x|y) \quad f_{Y|X}(y|x)\,.$$

- Especifique valores iniciais

$$X^{(0)} = x_0 \quad Y^{(0)} = y_0 \,.$$

- Para i de 1 até B faça:
    - $X^{(i)} \sim f_{X|Y}(x^{(i-1)}|y^{(i-1)})$
    - $Y^{(i)} \sim f_{Y|X}(y^{(i-1)}|x^{(i)})$
- A partir de um B grande:
    - o vetor gerado $(x^{(i)}, y^{(i)})$ segue a distribuição $f_{X,Y}(x, y)$.
- Vejamos agora como aplicar o algoritmo para amostrar de distribuições a *posteriori*.

- Suponha que queremos gerar uma amostra da distribuição a *posteriori* de um vetor de parâmetro θ

$$\pi(\theta|\mathbf{y}).$$

- Iniciamos o algoritmo com um conjunto de valores iniciais arbitrários para o vetor

$$\theta^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \ldots, \theta_k^{(0)}).$$

- Atualizamos cada um dos parâmetros $\theta_1, \theta_2, \ldots, \theta_k$ de uma vez.
- Essa atualizações são feitas através das distribuições condicionais

$$\pi(\theta_i|\theta_{-i}, \mathbf{y}).$$

onde

$$\theta_{-i} = (\theta_1, \theta_2, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_k).$$

Passo 1

- Geramos

$$\theta_1^{(1)} \sim \pi(\theta_1|\theta_2^{(0)},\ldots,\theta_k^{(0)},\mathbf{y})$$

$$\theta_2^{(1)} \sim \pi(\theta_1|\theta_1^{(1)},\ldots,\theta_k^{(0)},\mathbf{y})$$

$$\vdots$$

$$\theta_k^{(1)} \sim \pi(\theta_1|\theta_1^{(1)},\ldots,\theta_{k-1}^{(1)},\mathbf{y})$$

Passo 2

- Geramos

$$\theta_1^{(2)} \sim \pi(\theta_1|\theta_2^{(1)},\ldots,\theta_k^{(1)},\mathbf{y})$$

$$\theta_2^{(2)} \sim \pi(\theta_1|\theta_1^{(2)},\ldots,\theta_k^{(1)},\mathbf{y})$$

$$\vdots$$

$$\theta_k^{(2)} \sim \pi(\theta_1|\theta_1^{(2)},\ldots,\theta_{k-1}^{(2)},\mathbf{y})$$

Passo j

- Geramos

$$
\begin{aligned}
\theta_1^{(j)} &\sim \pi(\theta_1|\theta_2^{(j-1)},\ldots,\theta_k^{(j-1)},\mathbf{y})\\
\theta_2^{(j)} &\sim \pi(\theta_1|\theta_1^{(j)},\ldots,\theta_k^{(j-1)},\mathbf{y})\\
&\vdots\\
\theta_k^{(j)} &\sim \pi(\theta_1|\theta_1^{(j)},\ldots,\theta_{k-1}^{(j)},\mathbf{y})
\end{aligned}
$$

Exemplo 1:

- Suponha que temos n observações
  $Y_i \sim \text{Poisson}(\lambda), \quad i = 1, \ldots, n$
- Vamos definir $\lambda \sim \Gamma(\alpha, \beta)$
- Queremos fazer inferência sobre o parâmetro λ
- Portanto, precisamos calcular $\pi(\lambda|\mathbf{y})$.

- Vejamos como fica a distribuição a *posteriori* de λ (fazer no quadro)

$$\begin{aligned}\pi(\lambda|\boldsymbol{y}) &\propto \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{y_i}}{y_i!} e^{-\lambda\beta}\lambda^{\alpha-1} \\ &\propto \lambda^{\sum_{i=1}^{n} y_i+\alpha-1} e^{-\lambda(\beta+n)}\end{aligned}$$

- Portanto,

$$\lambda \sim \Gamma(\alpha + n\bar{y}, \beta + n)$$

- Como conhecemos a posteriori de λ podemos amostrar da distribuição

Exemplo 2:

- Suponha que temos N observações
  $Y_i \sim \text{Bin}(n_i, p), \quad i = 1, \ldots, N$
- Vamos definir $p \sim \text{Beta}(\alpha, \beta)$
- Queremos fazer inferência sobre o parâmetro p
- Portanto, precisamos calcular $\pi(p|\mathbf{y})$.

- Vejamos como fica a distribuição a *posteriori* de λ (fazer no quadro)

$$\begin{aligned}\pi(p|\boldsymbol{y}) &\propto \prod_{i=1}^{n} p^{y_i} - (1-p)^{n_i - y_i} p^{\alpha-1}(1-p)^{\beta-1} \\ &\propto p^{\sum_{i=1}^{n} y_i + \alpha - 1}(1-p)^{\sum_{i=1}^{N}(n_i - y_i) + \beta - 1}\end{aligned}$$

- Portanto,

$$p \sim \text{Beta}(\alpha + n \times \hat{y}, \beta + \sum_{i=1}^{N}(n_i - y_i))$$

- Como conhecemos a posteriori de p podemos amostrar da distribuição

Metropolis Hasting

- Até agora vimos como amostrar sequencialmente quando as condicionais completas possuem forma fechada.
- O que fazer quando elas não possuem? Nesse caso utilizamos um passo de Metropolis

Metropolis Hasting

- O passo de Metropolis consiste no seguinte:
  - Proponha uma distribuição conhecida $q(\theta_i)$ para gerar um candidato $\theta_i^{(t)*}$
  - Para testar se $\theta_i^{(t)*}$ é um bom candidato calcule

$$\text{ratio} = \frac{\pi(\theta_i^{(t)*}|(\theta_1^{(t+1)},\dots,\theta_{i-1}^{(t+1)},\theta_{i+1}^{(t)},\dots,\theta_k^{(t)}),\boldsymbol{y})}{\pi(\theta_i^{(t)}|(\theta_1^{(t+1)},\dots,\theta_{i-1}^{(t+1)},\theta_{i+1}^{(t)},\dots,\theta_k^{(t)}),\boldsymbol{y})}\,\frac{q(\theta_i^{(t)*})}{q(\theta_i^{(t)})}$$

  - Agora faça: Se $u < \min(1, \text{ratio})\theta_i^{(t+1)} = \theta_i^{(t)*}$, se não $\theta_i^{(t+1)} = \theta_i^{(t)}$, onde $U \sim \text{Unif}(0, 1)$

Exemplo

- Suponha que temos N observações $Y_i \sim \text{Bin}(n_i, p_i), \quad i = 1, \dots, N$ e $\text{logit}(p_i) = \beta_0 + \beta_1 x_i$
- Escolha uma priori para β_0 e β_1, e.g., $\beta_j \sim N(0, \sigma_j^2)\, j = 0, 1$.
- Queremos fazer inferência sobre o parâmetro $\beta = (\beta_0, \beta_1)$
- Portanto, precisamos calcular $\pi(\beta|\mathbf{y})$.

- Para isso vamos utilizar o amostrador de Gibbs com passo de Metropolis pois a posteriori não possui forma fechada.

$$\pi(\beta|\mathbf{y}) \propto \prod_{i=1} \left(\frac{e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}} \right)^{y_i} \left(\frac{1}{1+e^{\beta_0+\beta_1 x_i}} \right)^{n_i-y_i} \prod_{j=1}^{2} e^{-\frac{\beta_j^2}{2\sigma_j^2}}$$

- Portanto, a posteriori não possui forma fechada.
- Assim

$$\pi(\beta_0|\beta_1,\mathbf{y}) \propto \prod_{i=1} \left(\frac{e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}} \right)^{y_i} \left(\frac{1}{1+e^{\beta_0+\beta_1 x_i}} \right)^{n_i-y_i} e^{-\frac{\beta_0^2}{2\sigma_0^2}} \text{ e}$$

$$\pi(\beta_1|\beta_0,\mathbf{y}) \propto \prod_{i=1} \left(\frac{e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}} \right)^{y_i} \left(\frac{1}{1+e^{\beta_0+\beta_1 x_i}} \right)^{n_i-y_i} e^{-\frac{\beta_1^2}{2\sigma_1^2}}$$

- Como não conhecemos a constante de normalização não é possível amostrar diretamente da condicional completa.
- Vamos assim, utilizar o passo de Metropolis propondo $q(\beta_j^{(t)*}) = N(\beta_j(t), \sigma_j^2)$
- Dessa forma temos
  - Amostre $\beta_0^{(t+1)*}$
  - Calcule o ratio $= \frac{\pi(\beta_0^{(t+1)*}|\beta_1^{(t)}\boldsymbol{y})}{\pi(\beta_0^{(t)}|\beta_1^{(t)}\boldsymbol{y})} \frac{q(\beta_0^{(t)}}{q(\beta_0^{(t+1)*}}$
  - Faça $\beta_0^{(t+1)} = \beta_0^{(t+1)*}$ se $u <$ ratio e $\beta_0^{(t+1)} = \beta_0^{(t)}$ caso contrario

-
    - Amostre $\beta_1^{(t+1)*}$
    - Calcule o ratio $= \frac{\pi(\beta_1^{(t+1)*}|\beta_0^{(t+1)}\boldsymbol{y})}{\pi(\beta_1^{(t)}|\beta_0^{(t+1)}\boldsymbol{y})} \frac{q(\beta_1^{(t)}}{q(\beta_1^{(t+1)*}}$
    - Faça $\beta_1^{(t+1)} = \beta_1^{(t+1)*}$ se $u <$ ratio e $\beta_1^{(t+1)} = \beta_1^{(t)}$ caso contrario
- Repita o procedimento B vezes.

Maneiras automáticas

- Atualmente existe uma variedades de programas capazes de fazer inferência Bayesiana de forma automática.
- Por exemplo
  - os pacotes *ars*, *MCMCpack*, *mcmc*, *INLA* , *NIMBLE*, *STAN* para o *R*
  - OpenBUGS, JAGS, MultiBUGS
  - SAS

- A aplicação desses software (pacotes) varia. O OpenBUGS oferece uma ferramenta bem flexível para representação hierárquica de modelos Bayesianos
- Para chamar o OpenBUGS de dentro do *R* alguns pacotes foram desenvolvidos: *rbugs*, *R2OpenBUGS* e *BRugs*
- O funcionamento do OpenBUGS depende de 3 aquivos bases
    1. Arquivo com o modelo a ser ajustado
    2. Arquivo com os dados
    3. Arquivo com os valores iniciais da cadeia de Markov

- A sintaxe do OpenBUGS é equivalente a sintaxe do *R*. Portanto, um exemplo de codigo para o modelo é:

```
model {
    for (i in 1 : N) {
      theta[i] ~ dgamma(alpha, beta)
      lambda[i] <- theta[i] * t[i]
      x[i] ~ dpois(lambda[i])
    }
    alpha ~ dexp(1)
    beta ~ dgamma(0.1, 1.0)
}
```

- O codigo pumps.R ensina como chamar o OpenBUGS de dentro do *R* através do *rbugs*