Modelagem Bayesiana e Aplicações

Márcia D'Elia Branco

Universidade de São Paulo
Instituto de Matemática e Estatística
http:www.ime.usp.br/m̃branco

Modelos Hierárquicos

Modelos Hierárquicos

- Embora um modelo sempre seja uma simplificação da realidade, nas últimas décadas modelos mais complexos tem sido considerados.
- Modelos complexos envolvem uma grande quantidade de parâmetros.
- Modelos muito simples são usualmente pouco realistas e podem induzir erros significativos nas conlusões.
- Por outro lado, modelos com muitos parâmetros tem problema de *overfit* e podem ser péssimos para predição.
- Outro problema grave em modelos com muitos parâmetros é da estimabilidade. Quantos parâmetros é possível estimar com meus dados?

Modelos Hierárquicos

- Modelos Bayesianos hierárquicos lidam bem com esse tipo de problema ao acrescentar uma estrutura de dependência entre os parâmetros através de um modelo probabilístico.
- Considere θ um vetor de parâmetros de dimensão p, tal que, condicional a θ as observações y_i sejam independentes.
- Assim

$$f(y \mid \theta) = \prod_{i=1}^{n} f(y_i \mid \theta)$$

- Se $p > n$ é impossível estimar θ sem que suposições adicionais sejam incorporadas ao modelo.
- O modelo Bayesiano completo inclui $h(\theta \mid \Psi)$ em que Ψ é um conjunto de parâmetros de dimensão menor.

Exemplo 1: Modelo binomial/beta hierárquico.

Considere θ a probabilidade de um rato de laboratório ter um determinado tumor.

Em uma amostra de 14 ratos observou-se 4 com o tumor.

Suposições:

$$y \mid \theta \sim Bin(14, \theta)$$
$$\theta \sim Beta(\alpha, \beta)$$

Construindo a distribuição *a priori* usando dados históricos. Em 70 grupos de ratos analisados anteriormente observou-se $\frac{y_j}{n_j}$, $j = 1, \ldots, 70$.

Solução 1: Não hierárquica.

Obteve-se a média dessas proporções e o desvio padrão, resultando em 0.136 e 0.103.

Igualando esses valores aos valores de

$$E[\theta] = \frac{\alpha}{\alpha+\beta}$$

$$DP[\theta] = \sqrt{\frac{\alpha\beta}{[(\alpha+\beta)^2(\alpha+\beta+1)]}}$$

Obtemos $\hat{\alpha} = 1.4$ e $\hat{\beta} = 8.6$

Com $\theta \sim Beta(1.4, 8.6)$ obtemos a distribuição *a posteiori*

$$\theta \mid y \sim Beta(5.4, 18.6).$$

Os valores de médias e desvio padrão *a posteriori* são

$$E[\theta \mid y] = 0.223 \quad , \quad DP[\theta \mid y] = 0.083.$$

Note que a média obtida difere bastante da média dos experimentos anterioris (0.136). Indicando que talvés cada experimento provenha de populações com distintas médias.

Considere agora conjuntamente todos os 71 experimentos, em que cada um deles refere-se a uma população com probabilidade de tumor diferente θ_j .

As limitações da abordagem anterior são as seguintes:

- O uso da mesma distribuição *a priori* obtida anteriormente, implica em usar duas vezes a informação dos dados.
- O uso da estimativa pontual no lugar dos parâmetros α e β ignora a incerteza sobre esses valores.
- Além disso, é rasoável supor que os θ_j estejam conectados de alguma maneira.

Solução 2: Hierárquica.

$$y_i \mid \theta_i \sim Bin(n_i, \theta_i)$$

$$\theta_i \mid \alpha, \beta \sim Beta(\alpha, \beta)$$

$$(\alpha, \beta) \sim h(\alpha, \beta)$$

para $i = 1, 2, \ldots, 71$.

(α, β) são denominados hiperparâmetros e $h(\alpha, \beta)$ hiperpriori.

A distribuição a posteriori conjunta:

$$h(\theta, \alpha, \beta \mid y) \propto f(y \mid \alpha, \beta, \theta) h(\theta \mid \alpha, \beta) h(\alpha, \beta) \propto$$

$$h(\alpha, \beta) \prod_{i=1}^{71} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta_i^{y_i+\alpha-1} (1-\theta_i)^{n_i-y_i+\beta-1}.$$

A distribuição a posteriori de θ condicional a (α, β) é o produto de Betas. Cada componente dado por

$$\theta_i \mid \alpha, \beta, y \sim Beta(\alpha + y_i, \beta + n_i - y_i).$$

A distribuição a posteriori marginal de (α, β) é

$$\int_{\Theta} h(\theta, \alpha, \beta \mid y) d\theta \propto$$

$$h(\alpha, \beta) \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_{\Theta} \prod_{i=1}^{71} \theta_i^{y_i+\alpha-1}(1-\theta_i)^{n_i-y_i+\beta-1} d\theta =$$

$$h(\alpha, \beta) \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \prod_{i=1}^{71} \frac{\Gamma(\alpha+y_i)+\Gamma(\beta+n_i-y_i)}{\Gamma(\alpha+\beta+n_i)}.$$

Não tem forma analitica conhecida mas pode ser avaliada numericamente fixando um conjunto de valores para (α, β).

A escolha da hiperpriori: $h(\alpha, \beta)$. Vamos considerar uma priori não informativa. Qual?

- O uso da regra de Jeffreys neste caso é bastante complicado devido a forma da verossimilhança $f(y \mid \alpha, \beta)$
- O uso da Uniforme imprópria $h(\alpha, \beta) \propto 1$ também não é viável pelo fato de resultar em uma posteriori imprópria (exercício!).
- Uma solução proposta na literatura é o uso de uniformes independentes na seguinte reparametrização:

$$\frac{\alpha}{\alpha+\beta} \quad \text{e} \quad (\alpha+\beta)^{-0.5}.$$

- Usando a regra de transformação de variáveis obtemos

$$h(\alpha, \beta) \propto (\alpha+\beta)^{-5/2}.$$

A última priori garante uma distribuição a posteriori própria. Podemos simular da distribuição a posteriori conjunta usando o seguinte esquema.

1. Simular $(\alpha, \beta)^j$ da posteriori marginal.
2. Simular y_i da $Beta(\alpha^j + y_i, \beta^j + n_i - y_i)$ para $i = 1, \ldots, 71$.
3. Repetir para $j = 1, 2, \ldots, M$.

Note que não estamos usando GS e sim simulação de MC usual.

Alternativamente, pode-se usar o OpenBUGS.

Esquema geral do MH Bayesiano

Modelo com três níveis de hierarquia:

$$y \mid \theta \sim f(y \mid \theta)$$

$$\theta \mid \Psi \sim h(\theta \mid \Psi)$$

$$\Psi \sim h(\Psi)$$

É posssível estabelecer mais níveis propondo priori para Ψ e assim sucessivamente.

Exemplo 2: Estimação de pequenas áreas.

y_i é o número observado de casos na região i e
E_i o número esperado de casos por região, o qual depende do tamanho da população exposta em cada região (valores conhecidos).

Modelo hierárquico:

$$y_i \mid \lambda_i \sim Poisson(\lambda_i E_i)$$

$$\lambda_i \mid \alpha, \beta \sim Gama(\alpha, \beta)$$

$$\alpha \sim Gama(a_0, b_0)\ ,\ \ \beta \sim Gama(c_0, d_0)$$

A distribuição a posteriori condicional:

$$h(\lambda \mid \alpha, \beta, y) \propto \prod_{i=1}^{n} \lambda_i^{y_i+\alpha-1} e^{-(\beta+E_i)\lambda_i}.$$

Resulta em um produto de distribbuições gamas onde cada componente é dada por

$$\lambda_i \mid \alpha, \beta, y \sim Gama(y_i + \alpha, \beta + E_i).$$

Assim

$$E[\lambda_i \mid \alpha, \beta, y] = \frac{y_i + \alpha}{\beta + E_i}.$$

A distribuição a posteriori marginal de (α, β):

$$h(\alpha, \beta \mid y) = \int_{\Theta} h(\alpha, \beta, \lambda \mid y) d\lambda \propto$$

$$\alpha^{a_0 - 1} e^{-b_0 \alpha} \beta^{c_0 - 1} e^{-d_0 \beta} \int_{\Theta} \prod_{i=1}^{n} \lambda_i^{y_i + \alpha - 1} e^{-(E_i + \beta)\lambda_i} d\lambda =$$

$$\alpha^{a_0 - 1} e^{-b_0 \alpha} \beta^{c_0 - 1} e^{-d_0 \beta} \prod_{i=1}^{n} \frac{(y_i + \alpha - 1)!}{(E_i + \beta)^{y_i + \alpha}}$$

- Usando um algoritmo para simular da posteriori marginal de (α, β), podemos proceder como no exemplo anterior.
- Alternativamente, pode ser usar o OpenBUGS.
- Ver aplicação no exemplo 2.3 (pag.57) do livro *Applied Bayesian Modelling* do P.Congdon.

Exemplo 3: Modelo de efeito aleatório normal/normal.

Considere y_{ij} uma medição da unidade amostral i do grupo j.
$j = 1, 2, \ldots, J$ e $i = 1, 2, \ldots, n_j$.

$$y_{ij} \mid \theta_j \sim N(\theta_j, \sigma^2),$$

com σ^2 conhecido.

Sabemos que

$$\bar{y}_j \mid \theta_j \sim N(\theta_j, \sigma_j^2)$$

com $\bar{y}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} y_{ij}$ estatística suficiente para θ_j e $\sigma_j^2 = \frac{\sigma^2}{n_j}$.

Segundo nível da hierarquia.

$$\theta_j \mid \mu, \tau^2 \sim N(\mu, \tau^2)$$

Note que

$$y_{ij} = \mu + b_j + \epsilon_{ij} \quad , \quad b_j \sim N(0, \tau^2) \quad , \epsilon_{ij} \sim N(0, \sigma^2).$$

A distribuição a posteriori conjunta:

$$h(\theta, \mu, \tau \mid y) \propto h(\mu, \tau) \prod_{j=1}^{J} \phi(\theta_j; \mu, \tau^2) \phi(\bar{y}_j; \theta_j, \sigma_j^2).$$

em que $\phi()$ representa a fdp da Normal.

A distribuição a posteriori condicional aos hiperparâmetros:

Será um produto de normais, com cada componente dada por:

$$\theta_j \mid \mu, \tau, y \sim N(A_j, V_j),$$

em que:

$$V_j = \left[\frac{1}{\sigma_j^2} + \frac{1}{\tau^2}\right]^{-1} \qquad A_j = V_j \left[\frac{\bar{y}_j}{\sigma_j^2} + \frac{\mu}{\tau^2}\right]$$

Note que o valor esperado de θ_j é uma combinaçao linear entre a média do grupo j e a média global μ.

A distribuição a posteriori marginal dos hiperparametros:

$$h(\mu, \tau \mid y) \propto h(\mu, \tau) f(y \mid \mu, \tau)$$

Para cada j, usando o resultado da preditiva de uma normal com variância conhecida, temos que

$$\bar{y}_j \mid \mu, \tau \sim N(\mu, \sigma_j^2 + \tau^2)$$

Assim

$$f(y \mid \mu, \tau) = \prod_{j=1}^{J} \phi(\bar{y}_j; \mu, \sigma_j^2 + \tau^2)$$

Completar!!!