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 2

Modelo Hierárquico Normal

Primeiro nível

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

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

Segundo nível

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

Obtivemos a seguinte distribuição a posteriori condicional aos hiperparâmetros:

$$h(\theta \mid \mu, \tau, y) = \prod_{j=1}^{J} \phi(\theta_j; A_j, V_j)$$

em que $\phi()$ representa a fdp da normal e

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

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

Reescrevemos

$$V_j = \tau^2 T_j \quad \text{e} \quad A_j = (1 - T_j)\bar{y}_j + T_j\mu,$$

com $T_j = \frac{\sigma_j^2}{\sigma_j^2 + \tau^2}$.

Escolha das (hiper)prioris.

1. Não informativas impróprias:

Se $h(\mu) \propto 1$ então a distribuição a posteriori condicional a τ é

$$\mu \mid \tau, y \sim N(\hat{\mu}, V_\mu)$$

com

$$V_\mu = \left[\sum_{j=1}^{J} \frac{1}{\sigma_j^2 + \tau^2} \right]^{-1} \qquad \hat{\mu} = V_\mu \sum_{j=1}^{J} \frac{\bar{y}_j}{\sigma_j^2 + \tau^2}$$

Usualmente considera-se uma uniforme na transformação $log(\tau)$. Mas neste problema não é adequada pois a posteriori resulta imprópria.

Alternativamente, considera-se uma priori uniforme imprópria em τ. (Exercício: mostrar que a posteriori é própria!).

A distribuição a posteriori para τ.

Note que

$$h(\tau \mid y) = \frac{h(\mu, \tau \mid y)}{h(\mu \mid \tau, y)}$$

Lembramos que obtivemos na última aula

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

Como $h(\mu, \tau) \propto 1$, resulta

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

Como essa distribuição é unidimensional não é difícil simular. Podemos usar o algoritmo dada na função *sim.x ()* no apêndice B do livro Turkman et al..

Esquema de simulação da distribuição a posteriori

1. Simular τ^t da marginal $h(\tau \mid y)$.
2. Simular μ^t da $N(\hat{\mu}^t, V_\mu^t)$
3. Simular θ_j^t da $N(A_j^t, V_j^t)$ independente para $j = 1, \ldots, J$.
4. Fazer $t = 1, 2, \ldots, M$

Não há necessidade de controlar a convergência, pois é um esquema de simulação direta que não envolve Cadeias de Markov.

2. Prioris Vagas (pouco informativas e próprias)

$$\mu \sim N(0, 10^k) \quad , \quad \gamma = \frac{1}{\tau^2} \sim Gama(\epsilon, \epsilon)$$

Escolhendo k grande e ϵ pequeno.

$$E[\gamma] = 1 \qquad Var[\gamma] = \frac{1}{\epsilon}$$

O modelo pode ser escrito no OpenBUGS ou JAGS e a inferência obtida via MCMC.

3. Algumas alternativas de prioris para τ

3.1. *Half-Cauchy* : $\tau = \frac{d_1}{\sqrt{d_2}}$ com

$$d_1 \sim N(0, D) I_{(0,\infty)} \quad \text{e} \quad d_2 \sim X_1^2.$$

Exercício: Mostrar que τ tem distribuição de Cauchy restrita aos valores positivos.

3.2 . Uniforme nos pesos (*shrinkage*) .

Considera uma Uniforme no intervalo $(0, 1)$ para os pesos $T = \frac{\sigma^2}{\sigma^2+\tau^2}$.

Vamos obter a distribuição de τ^2.

A transformação inversa é dada por $\tau^2 = \frac{\sigma^2}{T} - \sigma^2$.
Então

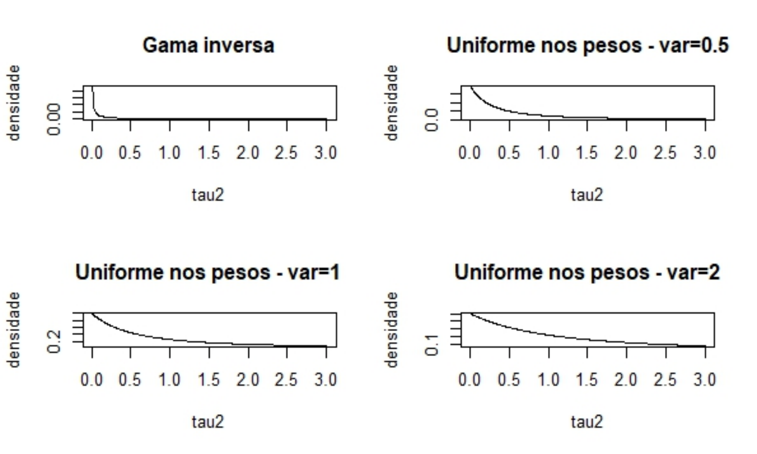
$$P(\tau^2 \leq y) = P(\frac{\sigma^2}{T} - \sigma^2 \leq y) = P(T \geq \frac{\sigma^2}{\sigma^2 + y}) =$$

$$1 - \frac{\sigma^2}{\sigma^2 + y} = \frac{y}{\sigma^2 + y}.$$

Derivando obtemos a densidade que resulta em

$$h(\tau^2 \mid \sigma^2) = \frac{\sigma^2}{(\sigma^2 + \tau^2)^2} \quad , \tau^2 > 0.$$

Prioris para τ^2



3.3. Proposta de Gustafson et al. (2006).

$$h(\tau \mid \sigma^2) = \frac{a}{\sigma^2(1 + \tau^2/\sigma^2)^{a+1}}$$

- Para grandes valores de a evita estimativas muito altas para τ. Priori conservativa.
- Para $a = 1$ ela é igual a Uniforme nos pesos.

Exemplo: Livro Gelman et al. (2014), pag.119.

Interesse: avaliar o efeito de um programa de treinamento para melhorar a performace em um determinado teste.

Foram analisadas $J = 8$ escolas. Para cada uma delas foi obtida uma medida resumo das diferenças de desempenho dos alunos "treinados"e "não treinados", denotada por y_j , $j = 1, \ldots, 8$.

Além disso, as variâncias amostrais de cada grupo são consideradas muito próximas das populacionais, devido ao fato dos tamanhos de amostras serem grandes. Os dados são apresentados na tabela a seguir.

Escola	y_j	σ_j
A	28	15
B	8	10
C	-3	16
D	7	11
E	-1	9
F	1	11
G	18	10
H	12	18

- Olhando para as estimativas pontuais pode parecer que algumas escolas tiveram um desempenho destacado em relação às outras.
- No entanto, ao analisarmos os desvio padrões (valores muito altos) notamos que não há diferença significativa entre elas. Os intervalos de credibilidades para cada um dos θ_j se interseccionam.
- Podemos então pensar que todas escolas provêm de uma mesma população e considerar um único θ. Neste caso, a estimativa pontual é 7.7 com um desvio padrão de 4.1.
- As conclusões obtidas com os dois modelos são bastante distintas. Por exemplo, considerando a análise individual temos que $P(\theta_1 > 28 \mid y) = 0.5$. Já na análise conjunta, $P(\theta_1 > 28 \mid y) \approx 0$.

Foi ajustado o modelo com hiperpriori uniforme em (μ, τ) .
Na tabela a seguir temos as estimativas dos θ_j para uma amostra simulada de tamanho $M = 200$.

Escola	2.5 %	25 %	mediana	75 %	97.5 %
A	-2	7	10	16	31
B	-5	3	8	12	23
C	- 11	2	7	11	19
D	-7	4	8	11	21
E	-9	1	5	10	18
F	-7	2	6	10	28
G	-1	7	10	15	26
H	-6	3	8	13	33

Modelo Hierárquico Normal

- Os intervalos de credibilidade novamente se interseccionam e a ordem é mantida. A média geral também não difere muito do estudo anterior.
- No entanto, os valores de estimativas pontuais diferem bastante da análise individual anterior. A variabilidade entre as estimativas pontuais é bem menor. A observação da distribuição a posteriori de τ permite inferir sobre essa variabilidade.
- Comparando com a análise conjunta (único θ), notamos que os intervalos individuais para cada θ_j posssuem uma amplitude maior do que o intervalo único na análise global.
- Outra diferença $P(\theta_1 > 28) \approx 0.10$ o que difere de ambas análises anterioris.
- Para mais detalhes ver Gelman et al. (2014).

1. **Modelo t-Student**

Mantemos a suposição de normalidade no nível 1 e flexibilizamos a suposição no nível 2.

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

Considerarando a representação hierárquica da distribuição t-Student, como mistura de normais, temos

$$y_j \mid \theta_j \sim N(\theta_j, \sigma_j^2)$$

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

$$\lambda_j \mid \nu \sim Gama(\nu/2, \nu/2)$$

- O modelo t-Student acomoda melhor observações discrepantes.
- Os novos parâmetros, λ_j, introduzidos no modelo podem ser usados para análise de *outliers*. Valores altos de λ_j estão associados a valores discrepantes.
- O modelo normal resulta quando $\nu \to \infty$.
- Um novo problema surge, escolher a distribuição a priori para os graus de liberdades ν .

2. Modelos assimétricos

O modelo skew-normal também pode ser estabelecido de forma hierárquica a partir da normal.

$$y_j \mid \theta_j \sim N(\theta_j, \sigma_j^2)$$

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

$$w_j \sim N(0,1) I_{(0,\infty)}$$

- O novo parâmetro $\lambda \in R$ é um parâmetro de forma.
- Se $\lambda = 0$ o modelo é simétrico, voltamos ao modelo normal.
- Se $\lambda > 0 [\lambda < 0]$ resulta numa distribuição assimétrica positiva [negativa] .

Correção da pag. 62 do livro Congdon (2014) :

Se $\frac{1}{\tau^2} \sim Gama(\epsilon, \epsilon)$ então, a distribuição a priori de θ_j condicional a μ tem distribução $t - Student$ com $\nu = 2\epsilon$ graus de liberdades.

Exercício !!