Hierchical: between separate & pooled model

separate model: ______

pooled model: ______

Recall:

$$\begin{array}{ccccc} & & \alpha, \beta & & \text{(global)} \\ \theta_1 & \theta_2 & \cdots & \theta_J & \text{(local)} \\ | & | & \cdots & | & \\ y_1 & y_2 & & y_J & \end{array}$$

$$\begin{cases} \theta_j \sim \text{Beta}(\alpha, \beta) \\ y_j \sim \text{Binomial}(n_j, \theta_j) \end{cases}$$

Specification of priors for α, β:
different choices of "flat / non-informative" prior

$[0,1] \to \mathbb{R}^1$ $\quad$ $\alpha+\beta$: total d.o.f.

$\mathbb{R}^+ \to \mathbb{R}$

① On the transformed space $\left(\text{logit}\left(\frac{\alpha}{\alpha+\beta}\right), \log(\alpha+\beta)\right)$ flat

② On . . . . . . $\left(\frac{\alpha}{\alpha+\beta}, (\alpha+\beta)^{-1/2}\right)$ flat

not proper priors: ① check the corresponding posterior distribution is proper. (integrable)

① → improper ... lo...

② → proper posterior

Check. (HINT): needs approximating approaches

Q: what's the corresponding $p(\alpha, \beta)$?

commonly: a parameter θ $\begin{cases} \in (0,1) \\ \in \mathbb{R}^+ \\ \in [a,b] \end{cases}$

transformation →

can transform a certain para' to real domain $\mathbb{R}$

$$\begin{cases} \varphi = \text{logit}(\theta) = \log \frac{\theta}{1-\theta} \in \mathbb{R} \\ \varphi = \log(\theta) \in \mathbb{R} \\ \varphi = \frac{\theta - a}{b - a} \rightarrow \varphi = \text{logit}(\cdots) \end{cases}$$

$\in (0,1)$ $\in \mathbb{R}$

$\sigma^2 \rightarrow \log(\sigma^2)$

$\in \mathbb{R}^+$ $\in \mathbb{R}$

$p(\varphi) \propto 1$ then what's $p(\theta)$?

push the large $\rightarrow$ center

$$p(\theta) = p(\varphi)\left|\frac{\partial \varphi}{\partial \theta}\right| \quad \text{(Jacobian)}$$

positive, and can be multi-dimensional (determinant of the Jacobian)

want $\boxed{p(\theta)\,d\theta}$ VS. $\boxed{p(\varphi)\,d\varphi}$

Now Let $a = \frac{\alpha}{\alpha+\beta}$, $b = (\alpha+\beta)^{-1/2}$

$$\frac{\partial(a,b)}{\partial(\alpha,\beta)} = \begin{pmatrix} \frac{\partial a}{\partial \alpha} & \frac{\partial a}{\partial \beta} \\ \frac{\partial b}{\partial \alpha} & \frac{\partial b}{\partial \beta} \end{pmatrix} = \begin{pmatrix} \frac{\beta}{(\alpha+\beta)^2} & -\frac{\alpha}{(\alpha+\beta)^2} \\ -\frac{1}{2}(\alpha+\beta)^{-3/2} & -\frac{1}{2}(\alpha+\beta)^{-3/2} \end{pmatrix}$$

$$\left|\det\left(\left[\quad\right]\right)\right| = \left|-\frac{1}{2}\frac{\beta}{(\alpha+\beta)^{5/2}} - \frac{1}{2}\frac{\alpha}{(\alpha+\beta)^{5/2}}\right|$$

$$= \left|-\frac{1}{2}(\alpha+\beta)^{-5/2}\right|$$

$$= \frac{1}{2}(\alpha+\beta)^{-5/2}$$

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

$p(\theta) = p(\varphi)\left|\frac{\partial \varphi}{\partial \theta}\right|$

$\downarrow$ α, β $\downarrow$ a, b

↪ let it to be flat $p(\varphi) \propto 1$

$$\boxed{\text{So whole model: } \begin{cases} \theta_j \sim \text{Beta}(\alpha, \beta) \; ; \; p(\alpha, \beta) \propto (\alpha+\beta)^{-5/2} \\ y_j \sim \text{Binomial}(n_j, \theta_j) \end{cases}}$$

Apply Bayes Rules.

(Conditioned on θ_j, y_j is independent with α, β)

$$p(\underbrace{\alpha, \beta, \theta_{1:J}}_{\text{unknown}} \mid y_{1:J}) \propto p(\alpha, \beta, \theta_{1:J})\, p(y_{1:J} \mid \cancel{\alpha}, \beta, \theta_{1:J})$$

joint

$$\propto p(\alpha, \beta)\, \underbrace{p(\theta_{1:J} \mid \alpha, \beta)\, p(y_{1:J} \mid \theta_{1:J})}_{\text{Conditional independence}}$$

$$\propto (\alpha+\beta)^{-5/2} \prod_{j=1}^{J} p(\theta_j \mid \alpha, \beta) \prod_{j=1}^{J} p(y_j \mid \theta_j)$$

Beta dis'

$$\propto (\alpha+\beta)^{-5/2} \prod_{j=1}^{J} \left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta_j^{\alpha-1} (1-\theta_j)^{\beta-1} \right] \prod_{j=1}^{J} \theta_j^{y_j} (1-\theta_j)^{n_j - y_j}$$

$$p(\theta_j \mid \alpha, \beta, y_j) \propto p(\alpha, \beta, \theta_{1:J} \mid y_{1:J}) \rightarrow$$

$\boxed{\text{conditional probability distribution}}$

$$\propto \theta_j^{\alpha+y_j-1} (1-\theta_j)^{\beta+n_j-y_j-1}$$

$$\Rightarrow \theta_j \mid \alpha, \beta, y_j \sim \text{Beta}(\alpha+y_j, \beta+n_j-y_j)$$

marginal distribution:

$$p(\alpha, \beta \mid y_{1:J}) = \int p(\alpha, \beta, \theta_{1:J} \mid y_{1:J})\, d\theta_{1:J}$$

$$\propto (\alpha+\beta)^{-5/2} \prod_{j=1}^{J} \left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \underline{\frac{\Gamma(\alpha+y_j)\Gamma(\beta+n_j-y_j)}{\Gamma(\alpha+\beta+n_j)}} \right]$$

Why hierarchical × use data twice:

① inference: $p(\alpha, \beta, \theta_{1:J} \mid y_{1:J})$ (all quantites)

(quantity of interest)

$$\mathbb{E}[\theta_j \mid y_{1:J}] = \int \theta_j\, p(\alpha, \beta, \theta_{1:J} \mid y_{1:J})\, d\alpha\, d\beta\, d\theta_{1:J\setminus j}$$

自底向上

73 dimensions is too large! → simpler!

② Computation: ← hierarchical model reduces

sampling from $p(\alpha, \beta, \theta_{1:J} \mid y_{1:J})$

(condition) based on tree diagram

$$= \boxed{p(\alpha, \beta \mid y_{1:J})} \prod_{j=1}^{J} p(\theta_j \mid \alpha, \beta, y_j)$$

((conditionali)

independence)

e.g. $\longrightarrow (\alpha+\beta)^{-5/2} \prod_j [\cdots]$ Beta($\alpha+y_j$, $\beta+n_j-y_j$)

Draw samples for α, β

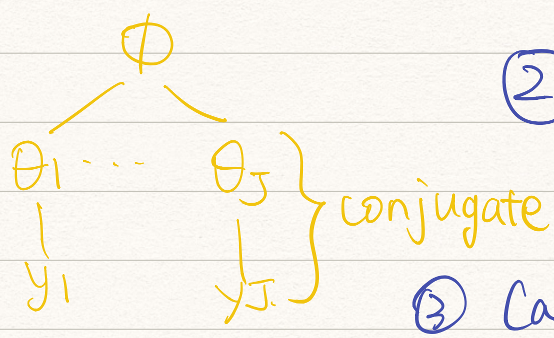
for each pair of (α, β)

sample θ_j from $p(\theta_j | \alpha, \beta, y_j)$

Bayes analysis for <u>conjugate</u> hierarchical models

(θ_j and y_j gives conjugacy)

① $p(\theta, \phi | y) \propto p(y|\theta)\, p(\theta|\phi)\, p(\phi)$



② derive $p(\theta | \phi, y)$ <u>analytically</u>

③ Calculate $p(\phi | y)$

$$= \int p(\phi, \theta | y)\, d\theta$$

如求 $p(\alpha, \beta \mid y, ij)$ 时可用这个方法 $\longrightarrow$

$$= \frac{p(\theta, \phi, y)}{p(\theta \mid \phi, y)}$$

Computation: $\phi \sim p(\phi \mid y)$,

$\theta \sim p(\theta \mid \phi, y)$