

22S:138
Introduction to Empirical Bayes

Lecture 22
 Nov. 16, 2009

Kate Cowles
 374 SH, 335-0727
 kcowles@stat.uiowa.edu

Compound sampling framework

- observed data conditionally independent given parameters

$$Y_i | \theta_i \sim f_i(y_i | \theta_i), \quad i = 1, \dots, n$$

- family of prior distributions indexed by low-dimensional parameter η

$$\theta_i \sim g(\theta_i | \eta)$$

Empirical Bayes

- Bayesian analysis requires specifying fixed values for parameters of highest-stage priors
 - these values must come from source other than the current dataset
- goal of empirical Bayes analysis is to fit hierarchical models *without* introducing information external to the current dataset
- EB approach
 - estimates final-stage parameters using current data
 - then proceeds as though prior were known
 - requires adjustment to posterior standard deviations and credible sets

Parametric EB (PEB) point estimation

- if η were known (fully Bayes)

$$p(\theta_i | y_i, \eta) = \frac{f_i(y_i | \theta_i) g(\theta_i | \eta)}{m_i(y_i | \eta)}$$

- i.e. posterior for θ_i depends on data only through y_i
- m_i is marginal likelihood of y_i
- PEB used marginal distribution of *all* the data to estimate η
 - $m(\mathbf{y} | \eta)$
 - use maximum likelihood or method of moments to get estimate $\hat{\eta}$
- plug $\hat{\eta}$ into above expression to get *estimated* posterior $p(\theta_i | y_i, \hat{\eta})$
 - use estimated posterior for all inference
 - e.g. point estimate of posterior mean
 - this point estimate depends on *all* the data through $\hat{\eta} = \hat{\eta}(\mathbf{y})$

Example: Normal/Normal models

- two-stage model

$$Y_i|\theta_i \sim N(\theta_i, \sigma^2), \quad i = 1, \dots, n$$

$$\theta_i|\mu \sim N(\mu, \tau^2), \quad i = 1, \dots, n$$

- first assume both τ^2 and σ^2 known

– calculations we have seen in GCSR show that marginally Y_i s are i.i.d.

* (i.e. with θ_i s integrated out)

* $Y_i|\mu \sim N(\mu, \sigma^2 + \tau^2)$

– so the marginal likelihood of all the Y_i s is

$$m(\mathbf{y}|\mu) = \frac{1}{[2\pi(\sigma^2 + \tau^2)]^{n/2}}$$

$$\exp\left[-\frac{1}{2(\sigma^2 + \tau^2)} \sum_{i=1}^n (y_i - \mu)^2\right]$$

– EB analysis requires estimation of μ

– marginal MLE of μ

$$\hat{\mu} = \bar{y}$$

- now suppose τ^2 , as well as μ , is unknown

– estimates of both μ and τ^2 are needed

– marginal MLEs

* $\hat{\mu} = \bar{y}$

* $\hat{\tau}^2 = (s^2 - \sigma^2)^+ = \max\{0, (s^2 - \sigma^2)\}$

where $s^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$

• the variation in the data over and above that expected if all the θ_i s were equal

* MMLE for B

$$\hat{B} = \frac{\sigma^2}{\sigma^2 + \hat{\tau}^2} = \frac{\sigma^2}{\sigma^2 + (s^2 - \sigma^2)^+}$$

* PEB estimates of θ_i

$$\hat{\theta}_i^{\mu, \tau^2} = \bar{y} + (1 - \hat{B})(y_i - \bar{y})$$

* amount of shrinkage is controlled by the estimated heterogeneity in the data

– estimated posterior distribution of θ_i

$$p(\theta_i|y_i, \hat{\mu}) = N(B\hat{\mu} + (1-B)y_i, (1-B)\sigma^2)$$

where

$$B = \frac{\sigma^2}{\sigma^2 + \tau^2}$$

exactly the same as fully Bayesian posterior for this case *except* that known prior mean μ is replaced by sample mean computed from all the data

– PEB point estimate of θ_i

$$\hat{\theta}_i^{\mu} = B\bar{y} + (1 - B)y_i$$

$$= \bar{y} + (1 - B)(y_i - \bar{y})$$

– inference for single component borrows information from data on all components

– shrinkage estimator

Example: the dyes data

- recall 2-stage model

$$Y_i|\theta_i \sim N(\theta_i, \sigma^2), \quad i = 1, \dots, n$$

$$\theta_i|\mu \sim N(\mu, \tau^2), \quad i = 1, \dots, n$$

- consider individual batch means as the y_i s, $i = 1, \dots, 6$

1505 1528 1564 1498 1600 1470

- suppose

– σ_{indiv}^2 for individual observations was known to be 2500 gm²

– 5 observations per batch, so variance of these batch means is known to be

$$\sigma^2 = \frac{\sigma_{indiv}^2}{5} = 500$$

- suppose τ^2 was known to be 1600
- MMLE of μ

$$\hat{\mu} = \bar{y} = 1527.5$$

- $B = \frac{\sigma^2}{\sigma^2 + \tau^2} = \frac{500}{500 + 1600} = 0.238$
- then EB point estimates of θ_i s

$$\theta_1^\mu = (1527.5) + (1 - 0.238)(1505 - 1527.5) = 1510.4$$

$$\theta_2^\mu = (1527.5) + (1 - 0.238)(1528 - 1527.5) = 1527.9$$

$$\theta_3^\mu = (1527.5) + (1 - 0.238)(1564 - 1527.5) = 1555.3$$

$$\theta_4^\mu = (1527.5) + (1 - 0.238)(1498 - 1527.5) = 1505.0$$

$$\theta_5^\mu = (1527.5) + (1 - 0.238)(1600 - 1527.5) = 1582.7$$

$$\theta_6^\mu = (1527.5) + (1 - 0.238)(1470 - 1527.5) = 1483.7$$

Comments

- EB estimates: compromise between
 - pooling all data ($\hat{B} = 1$)
 - using only data from i th observation or group to estimate i th parameter ($\hat{B} = 0$)
- one difficulty with PEB approach: choosing how to estimate hyperparameters
- contrast with fully Bayesian approach
 - would add another level to hierarchy
 - * add prior on μ and τ^2
 - replaces estimation with integration
 - avoids problem of selecting estimation method
 - automatically propagates to the posterior distribution the uncertainty induced by estimating μ and τ^2
 - however, requires selection of hyperpriors

Example continued

- Suppose $\sigma^2 = 500$ is still known
- but τ^2 is unknown
- $s^2 = \frac{1}{6} \sum_{i=1}^6 (y_i - \bar{y})^2 = 1878.6$
- then $\hat{\tau}^2 = (s^2 - \sigma^2)^+ = 1878.6 - 500 = 1378.6$
- MMLE $\hat{B} = \frac{500}{500 + 1378.6} = 0.266$
- then EB point estimates of θ_i s

$$\theta_1^{\mu, \tau^2} = (1527.5) + (1 - 0.266)(1505 - 1527.5) = 1521.5$$

$$\theta_2^{\mu, \tau^2} = (1527.5) + (1 - 0.266)(1528 - 1527.5) = 1527.6$$

$$\theta_3^{\mu, \tau^2} = (1527.5) + (1 - 0.266)(1564 - 1527.5) = 1537.2$$

$$\theta_4^{\mu, \tau^2} = (1527.5) + (1 - 0.266)(1498 - 1527.5) = 1519.7$$

$$\theta_5^{\mu, \tau^2} = (1527.5) + (1 - 0.266)(1600 - 1527.5) = 1546.8$$

$$\theta_6^{\mu, \tau^2} = (1527.5) + (1 - 0.266)(1470 - 1527.5) = 1512.2$$

“Naive” EB interval estimation

- given estimated posterior $p(\theta_i | y_i, \hat{\eta})$, use like any other posterior distribution to obtain HPD or equal tail credible set for θ_i
- from elementary math stat

$$Var(\theta_i | \mathbf{y}) = E_{\eta | \mathbf{y}}[Var(\theta_i | y_i, \eta)] + Var_{\eta | \mathbf{y}}[E(\theta_i | y_i, \eta)]$$
- in normal/normal case, 95% naive EB CI would be

$$E(\theta_i | y_i, \hat{\eta}) \pm 1.96 \sqrt{Var(\theta_i | y_i, \hat{\eta})}$$

- i.e. naive EB CI ignores the posterior uncertainty about η .
- so naive interval is very likely to be too short, and to have lower coverage probability than claimed
- substantial statistical literature on how to correct this problem