Empirical Bayes Appoaches

PUBH 8442: Bayes Decision Theory and Data Analysis

Eric F. Lock
UMN Division of Biostatistics, SPH
elock@umn.edu

03/13/2024

The Empirical Bayes Approach

- An empirical Bayes (EB) method is one in which the prior is estimated from the data.
- Useful when borrowing strength across multiple related parameters $\theta = (\theta_1, \ldots, \theta_k)$
- Model $\theta_i's$ as iid from shared (estimated) prior $\hat{p}_\theta(\cdot)$.
  - *Parametric EB*: Estimate p_θ from a parametric family with hyperparameters η: $p(\cdot \mid \eta)$
  - *Nonparametric EB*:Assume no parametric form for p_θ

Parametric empirical Bayes

- For the parametric EB approach we consider prior $p_\theta(\cdot \mid \eta)$ and estimate η.
- A natural estimate is the maximizer of the marginal density

$$\hat{\eta} = \underset{\eta}{\operatorname{argmax}}\ p(\mathbf{y} \mid \eta) = \underset{\eta}{\operatorname{argmax}} \int p(\mathbf{y} \mid \theta) p(\theta \mid \eta) d\theta$$

  - Used in the previous regression example, with $\theta := \beta$ and $\eta := \tau^2$.
- Alternatively, could estimate η by matching moments or some other approach.

Example: Liver disease (cont.)

- Recall: mice from 50 genetic strains
  - n_i is the total number of mice from strain i
  - y_i is the number of mice with liver disease from strain i.
  - θ_i is probability a mouse from strain i develops liver disease:
  $$y_i \sim \text{Binomial}(n_i, \theta_i).$$
  - $\theta_i's$ are iid from Beta(a, b) distribution.
- Would like to estimate a, b empirically.

Example: Liver disease (cont.)

- Consider reparameterization $\mu = \frac{a}{a+b}$ and $M = a + b$.
- The mean and variance of a beta distributed variable θ are

$$E(\theta) = \mu \quad \text{and} \quad \text{Var}(\theta) = \frac{\mu(1-\mu)}{M+1}$$

- To compute mean and variance of Y in terms of μ and M, use
  - Law of total expectation:

$$E_y Y = E_\theta[E_{Y|\theta} Y]$$

  - Law of total variance:

$$\text{Var}_y Y = E_\theta[\text{Var}_{Y|\theta} Y] + \text{Var}_\theta[E_{Y|\theta} Y]$$

Example: Liver disease (cont.)

- Let $\hat{\theta}_i = y_i/n_i$. Then,

$$E(\hat{\theta}_i) = \mu,$$
$$\text{Var}(\hat{\theta}_i) = \frac{\mu(1-\mu)}{n_i}\left(1 + \frac{n_i - 1}{M+1}\right).$$

Example: Liver disease (cont.)

- Consider "iterated method of moments" estimator based on $y_1, \ldots, y_m$

$$E(\hat{\theta}_i) \approx \hat{\mu} = \frac{1}{m}\sum_{j=1}^{m}\frac{y_j}{n_j}$$

Plugging in $\hat{\mu}$, note that

$$\frac{1}{m}\sum_{i=1}^{m}\text{Var}(\hat{\theta}_i) = \frac{1}{m}\sum_{i=1}^{m}\frac{\hat{\mu}(1-\hat{\mu})}{n_i}\left(1+\frac{n_i-1}{M+1}\right) \qquad (1)$$

Equating (1) with the sample variance of the $\hat{\theta}_i's$, s^2, gives

$$\hat{M} = \frac{\hat{\mu}(1-\hat{\mu}) - s^2}{s^2 - \frac{\hat{\mu}(1-\hat{\mu})}{m}\sum_{i=1}^{m} 1/n_i}.$$

Example: Liver disease (cont.)

- Can solve for $\hat{a}$ and $\hat{b}$:

$$\hat{a} = \hat{\mu}\hat{M}$$

$$\hat{b} = \frac{\hat{a}(1-\hat{\mu})}{\hat{\mu}}$$

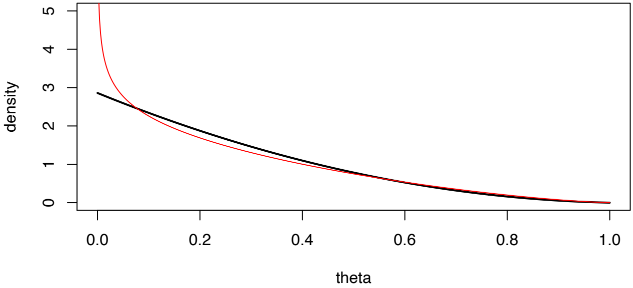
- Note that $\tilde{\theta}_i = E(\theta_i \mid y_i, n_i, \hat{M}, \hat{\mu}) =$

- Can write $\tilde{\theta}_i = \hat{B}_i\hat{\mu} + (1-\hat{B}_i)\hat{\theta}_i$

  - The "shrinkage factor" is $\hat{B}_i =$

Example: Liver disease (cont.)

- For mouse liver data $\hat{a} = 0.81, \hat{b} = 2.35$
- Plot of empirical prior Beta$(0.81, 2.35)$ red, with formerly obtained posterior predictive for θ_{51} using a discrete uniform prior for $a = 1, 2, 3$ and $b = 1, 2, 3$:



- - http://www.ericfrazerlock.com/Emperical_Bayes_Approaches_Rcode1.r

Non-parametric empirical Bayes

- A non-parametric empirical Bayes method does not assume a parametric form for p_θ.
- Base inference directly on marginal distribution

$$p(\mathbf{y}) = \int p(\mathbf{y} \mid \theta) p(\theta) d\theta$$

- Example: Robbins' method.
  - If $Y_i \mid \theta_i \sim \text{Poisson}(\theta_i)$, and θ_i have unspecified prior. Estimate

$$\hat{\theta}_i = (y_i + 1) \frac{\# ys \text{ equal to } y_i + 1}{\# ys \text{ equal to } y_i}$$