Chapter 4. Bayesian Analysis (C3)

♠ Empirical Bayes Analysis

• Parametric Empirical Bayes (PEB) for Normal Means — The Exchangeable Case

Let $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)'$. Assume

$$X_i \sim N(\theta_i, \sigma_f^2)$$

independently, $i = 1, 2, \ldots, p$, and the θ_i are considered to be exchangeable, modelled by supposing

$$\theta_i \overset{\text{i.i.d.}}{\sim} N(\mu_\pi, \sigma_\pi^2),$$

the hyperparameters μ_π and σ_π^2 being unknown.

The marginal distribution of X_i is given by

$$X_i \sim N(\mu_\pi, \sigma_f^2 + \sigma_\pi^2).$$

The Maximum likelihood Estimates of μ_π and σ_π^2 are

$$\hat{\mu}_\pi = \bar{x} = \frac{1}{p}\sum_{i=1}^{p} x_i,$$

and

$$\hat{\sigma}_\pi^2 = \max\left\{0, \frac{1}{p}s^2 - \sigma_f^2\right\},$$

where $s^2 = \sum_{i=1}^{p}(x_i - \bar{x})^2$.

Formally, one could then pretend that the θ_i are (independently) $N(\hat{\mu}_\pi, \hat{\sigma}_\pi^2)$, and proceed with a Bayesian analysis. This indeed works well when p is large. However, this approach ignores the fact that μ_π and σ_π^2 were estimated. The errors undoubtedly introduced in the hyperparameter estimation will not be reflected.

When p is small or moderate, Morris (1983) has developed empirical Bayes approximations to the hierarchical Bayes answers which do take into account the uncertainty in $\hat{\mu}_\pi$ and $\hat{\sigma}_\pi^2$. These approximations can best be described as providing an estimated posterior distribution (rather than an estimated prior).

To describe these approximations, recall that the posterior distribution of θ_i, for given μ_π and σ_π^2, is

$$N(\mu_i(x_i), V),$$

where letting $B = \sigma_f^2/(\sigma_f^2 + \sigma_\pi^2)$,

$$\mu_i(x_i) = x_i - B(x_i - \mu_\pi),$$

and

$$V = \frac{\sigma_\pi^2 \sigma_f^2}{\sigma_f^2 + \sigma_\pi^2} = \sigma_f^2 (1 - B).$$

The estimates Morris (1983) suggests for $\mu_i(x_i)$ and V (when $P \geq 4$)

$$\mu_i^{\boldsymbol{EB}}(x_i) = x_i - \hat{B}(x_i - \bar{x}),$$

and

$$V_i^{\boldsymbol{EB}}(\boldsymbol{x}) = \sigma_f^2 \left(1 - \frac{(p-1)}{p} \hat{B}\right) + \frac{2}{(p-3)} \hat{B}^2 (x_i - \bar{x})^2,$$

where the estimate of B is

$$\hat{B} = \left(\frac{p-3}{p-1}\right) \frac{\sigma_f^2}{(\sigma_f^2 + \tilde{\sigma}_\pi^2)},$$

and

$$\tilde{\sigma}_\pi^2 = \max\left\{0, \frac{s^2}{(p-1)} - \sigma_f^2\right\}.$$

Note that $V_i^{\boldsymbol{EB}}(\boldsymbol{x})$ and $\hat{B}$ are approximations to the posterior variances of θ_i and B given the data $\boldsymbol{X}$. The factor $(p-3)/(p-1)$ in $\hat{B}$ has to do with adjusting for the error in the estimation of σ_π^2 and $(p-1)/p$ and the last term in $V_i^{\boldsymbol{EB}}(\boldsymbol{x})$ have to do with the error in estimating μ_π.

The resultant $N(\mu_i^{\boldsymbol{EB}}(x_i), V_i^{\boldsymbol{EB}}(\boldsymbol{x}))$ (estimated) posterior for θ_i can be used in the standard Bayesian way. A $100(1-\alpha)\%$ HPD interval for θ_i is

$$C_i^{\boldsymbol{EB}}(\boldsymbol{x}) = \left(\mu_i^{\boldsymbol{EB}}(x_i) + z\left(\frac{\alpha}{2}\right)\sqrt{V_i^{\boldsymbol{EB}}(\boldsymbol{x})}, \right.$$

$$\left. \mu_i^{\boldsymbol{EB}}(x_i) - z\left(\frac{\alpha}{2}\right)\sqrt{V_i^{\boldsymbol{EB}}(\boldsymbol{x})} \right).$$

Example 1: Consider the example of the child who scores $x_7 = 115$ on a $N(\theta_7, 100)$ IQ test. Also available are intelligence test scores of the child for six previous years. These six scores are, 105, 127, 115, 130, 110, and 135, are assumed to be observations X_1, X_2, ..., X_6 from independent $N(\theta_i, 100)$. It can be calculated that $\bar{x} = 121$ and $s^2 = 762$. Since $\sigma_f^2 = 100$, we have

$$\tilde{\sigma}_\pi^2 = 27, \quad \hat{B} = \left(\frac{4}{6}\right)\left(\frac{100}{127}\right) = 0.525,$$

$$\mu_7^{\boldsymbol{EB}}(\boldsymbol{x}) = 115 - (0.525)(115 - 113) = 118.150$$

and

$$V_7^{\boldsymbol{EB}}(\boldsymbol{x}) = 100[1 - (6/7)(0.525)] + (2/4)(0.525^2)(115 - 121)^2 = 59.96.$$

The 95% HPD interval for θ_7 is

$$C_7^{\boldsymbol{EB}}(\boldsymbol{x}) = (118.15 \pm 1.96\sqrt{59.96}) = (102.97, 133.33).$$

Note that the classical 95% CI for θ_7, based on X_7 alone is

$$C_7(x_7) = (115\pi 1.96\sqrt{100}) = (95.40, 134.60).$$

When σ_f^2 is unknown,

$$X_i^1, X_i^2, \ldots, X_i^n \overset{i.i.d.}{\sim} N(\theta_i, \sigma^2),$$

and thus

$$\overline{X}_i \sim N(\theta_i, \sigma_f^2),$$

where $\sigma_f^2 = \frac{\sigma^2}{n}$, and we estimate σ_f^2 by

$$\hat{\sigma}_f^2 = \frac{1}{n(n-1)p} \sum_{i=1}^{n} \sum_{j=1}^{n} (x_i^j - \bar{x}_i)^2.$$

- **Parametric Empirical Bayes (PEB) for Normal Means — The General Case**

Consider

$$x_i \sim N(\theta_i, \sigma_i^2),$$

$$\theta_i = \boldsymbol{y}_i' \boldsymbol{\beta} + \epsilon_i,$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_l)'$ is a vector of unknown regression coefficients ($l < p - 2$), $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{il})'$ is a known set of regressors for each i, and

$$\epsilon_i \sim N(0, \sigma_\pi^2).$$

The simplest empirical Bayes analysis entails estimating the hyperparameters, $\boldsymbol{\beta}$ and σ_π^2 based on the marginal density of $\boldsymbol{X} = (X_1, \ldots, X_p)'$.

Since

$$X_i \sim N(\boldsymbol{y}_i' \boldsymbol{\beta}, \sigma_i^2 + \sigma_\pi^2)$$

independently, the marginal distribution of $\boldsymbol{X}$ is given by

$$m(\boldsymbol{x}) = \left(\prod_{i=1}^{p} [2\pi(\sigma_i^2 + \sigma_\pi^2)]^{-1/2} \right) \times \exp\left\{ -\frac{1}{2} \sum_{i=1}^{p} (x_i - \boldsymbol{y}_i' \boldsymbol{\beta})^2 / (\sigma_i^2 + \sigma_\pi^2) \right\}.$$

Following the ML-II approach, we can estimate $\boldsymbol{\beta}$ and σ_π^2 by differentiating $m(\boldsymbol{x})$ with respect to $\boldsymbol{\beta}$ and σ_π^2, and setting the equations equal to zero. Letting $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}_\pi^2$ denote these ML-II estimates, the equations obtained can be written

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{y}'V^{-1}\boldsymbol{y})^{-1}(\boldsymbol{y}'V^{-1}\boldsymbol{x}),$$

where $\boldsymbol{y}$ is the $(p \times l)$ matrix with rows $\boldsymbol{y}_i'$ and V is the $(p \times p)$ diagonal matrix with diagonal elements $V_{ii} = \sigma_i^2 + \hat{\sigma}_\pi^2$, and

$$\hat{\sigma}_\pi^2 = \frac{\sum_{i=1}^p \left\{[(x_i - \boldsymbol{y}_i'\hat{\boldsymbol{\beta}})^2 - \sigma_i^2]/[\sigma_i^2 + \hat{\sigma}_\pi^2]^2\right\}}{\sum_{i=1}^p (\sigma_i^2 + \hat{\sigma}_\pi^2)}$$

Unfortunately, the above equations do not provide closed form expressions for $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}_\pi^2$; but they do provide an easy iterative scheme for calculating $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}_\pi^2$.

Start out with a guess for $\hat{\sigma}_\pi^2$, and use this guess to calculate an approximate $\hat{\boldsymbol{\beta}}$. Then, plug the guess and the approximate $\hat{\boldsymbol{\beta}}$ into the expression of $\hat{\sigma}_\pi^2$ to obtain a new estimate for $\hat{\sigma}_\pi^2$. Repeat this procedure with the updated estimates until the numbers stabilize. If the convergence is to a negative value of $\hat{\sigma}_\pi^2$, the ML-II estimate of σ_π^2 is probably zero, and $\hat{\boldsymbol{\beta}}$ is then given by the least squares estimate with $\hat{\sigma}_\pi^2 = 0$.

Empirical Bayes analyses based on pretending that the θ_i have $N(\boldsymbol{y}_i'\hat{\boldsymbol{\beta}}, \hat{\sigma}_\pi^2)$ priors will work well if $p - l$ is large (and the model assumptions are valid). For smaller p, however, the estimation of $\boldsymbol{\beta}$ and σ_π^2 can again introduce substantial errors that must be taken into account. Morris (1983) accordingly develops approximations to the posterior means and variances of the θ_i which do take these additional errors into account. The approximations are given by

$$\mu_i^{\boldsymbol{EB}}(\boldsymbol{x}) = x_i \hat{B}_i(x_i - \boldsymbol{y}_i'\hat{\boldsymbol{\beta}}$$

and

$$\begin{aligned} V_i^{\boldsymbol{EB}}(\boldsymbol{x}) =& \sigma_i^2 \left[1 - \frac{(p - \hat{l}_i)}{p} \hat{B}_i\right] \\ &+ \frac{2}{(p - l - 2)} \hat{B}_i^2 \left(\frac{\bar{\sigma}^2 + \tilde{\sigma}_\pi^2}{\sigma_i^2 + \tilde{\sigma}_\pi^2}\right) (x_i - \boldsymbol{y}_i'\hat{\boldsymbol{\beta}})^2, \end{aligned}$$

where $\hat{\boldsymbol{\beta}} = (\boldsymbol{y}'V^{-1}\boldsymbol{y})^{-1}(\boldsymbol{y}'V^{-1}\boldsymbol{x})$,

$$\tilde{\sigma}_\pi^2 = \frac{\sum_{i=1}^p \left\{[(p/(p-1))(x_i - \boldsymbol{y}_i'\hat{\boldsymbol{\beta}})^2 - \sigma_i^2]/[\sigma_i^2 + \tilde{\sigma}_\pi^2]^2\right\}}{\sum_{i=1}^p (\sigma_i^2 + \tilde{\sigma}_\pi^2)^{-2}},$$

$$\hat{B} = \frac{(p - l - 2)}{(p - l)} \cdot \frac{\sigma_i^2}{\sigma_i^2 + \tilde{\sigma}_\pi^2},$$

$$\hat{l}_i = p[\boldsymbol{y}(\boldsymbol{y}'V^{-1}\boldsymbol{y})^{-1}\boldsymbol{y}']_{ii}/(\sigma_i^2 + \tilde{\sigma}_\pi^2),$$

and

$$\bar{\sigma}^2 = \frac{\sum_{i=1}^{p} \sigma_i^2/(\sigma_i^2 + \tilde{\sigma}_\pi^2)}{\sum_{i=1}^{p} 1/(\sigma_i^2 + \tilde{\sigma}_\pi^2)}.$$

Then one assume that θ_i has a $N(\mu_i^{\boldsymbol{EB}}(\boldsymbol{x}), V_i^{\boldsymbol{EB}}(\boldsymbol{x}))$ posterior distribution, and proceeds with the analysis.

- **Nonparametric Empirical Bayes Analysis**

The nonparametric empirical Bayes approach supposes that a large amount of data is available to estimate the prior, and hence places no (or minimal) restrictions on the form of the prior. One can try for direct estimation of the prior, as discussed in the ML-II approach and the moment approach to prior selection, and, if implementable, this is probably the best approach.

A mathematically appealing alternative introduced by Robbins (1955) is to seek a representation of the desired Bayes rule in terms of the marginal distribution, $m(\boldsymbol{x})$, of $\boldsymbol{x}$, and then use the data to estimate m, rather than π.

Example 2: Suppose X_1, X_2, ..., X_p are independent $\mathcal{P}(\theta_i)$, and that the θ_i are i.i.d. from a common prior π_0. Then X_1, ..., X_p can (unconditionally) be considered to be a sample from the marginal distribution

$$m(x_i) = \int f(x_i|\theta)\pi_0(\theta)d\theta,$$

where $f(x_i|\theta)$ is the $\mathcal{P}(\theta)$ density. This m can be approximated by the empirical estimate

$$\hat{m}(j) = (\text{the number of } x_i \text{ equal to } j)/p.$$

Suppose now that it is desired to estimate θ_p using the posterior mean. Observe that the posterior mean can be written

$$\begin{aligned}
\delta^{\pi_0}(x_p) =& E^{\pi(\theta_p|x_p)}[\theta_p] = \int \theta_p \pi(\theta_p|x_p)d\theta_p \\
=& \int \theta_p f(x_p|\theta_p)\pi_0(\theta_p)d\theta_p/m(x_p) \\
=& \int \theta_p^{x_p+1} \exp(-\theta_p)(x_p!)^{-1}\pi_0(\theta_p)d\theta_p/m(x_p) \\
=&(x_p+1)\int f(x_p+1|\theta_p)\pi_0(\theta_p)d\theta_p/m(x_p) \\
=&(x_p+1)m(x_p+1)/m(x_p).
\end{aligned}$$

Replacing m by the estimate $\hat{m}$ results in the estimated Bayes rule

$$\delta^{\boldsymbol{EB}}(\boldsymbol{x}) = \frac{(x_p + 1)(\#\text{ of } x_i \text{ equal to } x_p + 1)}{(\#\text{ of } x_i \text{ equal to } x_p)}.$$

The general approach can be stated as

(i) find a representation (for the Bayes rule) of the form $\delta^\pi(x) = \psi(x, \phi(m))$, where ψ and ϕ are known functionals;

(ii) estimate $\phi(m)$ by $\widehat{\phi(m)}$; and

(iii) use $\delta^{\boldsymbol{EB}}(x) = \psi(x, \widehat{\phi(m)})$.