Empirical Bayes Estimation

- In this approach, we again do not specify particular values for the prior parameters in $\boldsymbol{\psi}$.
- Instead of placing a (hyperprior) distribution on $\boldsymbol{\psi}$ as in hierarchical Bayes, the empirical Bayes approach is to **estimate** $\boldsymbol{\psi}$ from the data.
- This is not "purely" Bayesian, since in a sense we are using the data to determine the prior specification.
- Furthermore, the estimation of $\boldsymbol{\psi}$ must be done with non-Bayesian techniques (like maximum likelihood or method of moments).

Empirical Bayes Estimation

- If the prior on $\boldsymbol{\theta}$ depends on hyperparameter(s) $\boldsymbol{\psi}$, then the posterior is:

$$\pi(\boldsymbol{\theta}|\mathbf{X},\boldsymbol{\psi}) = \frac{p(\boldsymbol{\theta}|\boldsymbol{\psi})L(\boldsymbol{\theta}|\mathbf{X})}{\int_{\Theta} p(\boldsymbol{\theta}|\boldsymbol{\psi})L(\boldsymbol{\theta}|\mathbf{X})\,\mathrm{d}\boldsymbol{\theta}} = \frac{p(\boldsymbol{\theta}|\boldsymbol{\psi})L(\boldsymbol{\theta}|\mathbf{X})}{q(\mathbf{X}|\boldsymbol{\psi})}$$

- Now we use as the hyperparameter(s) some estimate of $\boldsymbol{\psi}$, such as the MLE of $\boldsymbol{\psi}$ based on $q(\mathbf{X}|\boldsymbol{\psi})$.

Examples: Empirical Bayes Estimation

- **Example 1**: Let $X_i \overset{\text{iid}}{\sim}$ Pois(λ_i), $i = 1, \ldots, n$, and let

$$\lambda_i \overset{\text{iid}}{\sim} \text{Gamma}(\alpha, \beta) \text{ with } \alpha \text{ known, } \beta \text{ unknown.}$$

$$\text{Then } q(X_i|\beta) = \int_0^\infty \left[\frac{\beta^\alpha}{\Gamma(\alpha)} \lambda_i^{\alpha-1} e^{-\beta\lambda_i}\right] \left[\frac{e^{-\lambda_i} \lambda_i^{x_i}}{x_i!}\right] \mathrm{d}\lambda_i$$

$$= \frac{\beta^\alpha}{x_i!\Gamma(\alpha)} \int_0^\infty \lambda_i^{x_i+\alpha-1} e^{-(\beta+1)\lambda_i} \, \mathrm{d}\lambda_i$$

$$= \frac{\beta^\alpha \Gamma(x_i + \alpha)}{x_i!\Gamma(\alpha)(\beta+1)^{x_i+\alpha}}$$

$$= \binom{x_i + \alpha - 1}{\alpha - 1} \left(\frac{\beta}{\beta+1}\right)^\alpha \left(\frac{1}{\beta+1}\right)^{x_i}$$

which is negative binomial.

Examples: Empirical Bayes Estimation

$$\Rightarrow q(\mathbf{X}|\beta) = \left[\prod_{i=1}^{n} \binom{x_i + \alpha - 1}{\alpha - 1}\right] \left(\frac{\beta}{\beta + 1}\right)^{n\alpha} \left(\frac{1}{\beta + 1}\right)^{\sum x_i}$$

and it can be shown that the MLE of β is $\hat{\beta} = \dfrac{\alpha}{\bar{x}}$.

- Using the prior $\lambda_i \sim \text{Gamma}(\alpha, \hat{\beta})$, the posterior for λ_i is thus

$$\lambda_i | x_i, \hat{\beta} \sim \text{Gamma}(x_i + \alpha, 1 + \hat{\beta})$$

- Hence the Empirical Bayes estimator for λ_i ($i = 1, \ldots, n$) is the posterior mean

$$\frac{X_i + \alpha}{1 + \alpha/\bar{X}} = \left(\frac{\bar{X}}{\bar{X} + \alpha}\right)(X_i + \alpha).$$

Examples: Empirical Bayes Estimation

Example 2(a): (One-way classification, 1 observation per group)

$$X_i|\mu_i \overset{\text{indep}}{\sim} N(\mu_i, \sigma^2), \quad i = 1, \ldots, m$$

$$\mu_i \overset{\text{iid}}{\sim} N(\phi, \tau^2), \quad \sigma^2, \tau^2 \text{ known.}$$

- Then it can be shown

$$q(\mathbf{x}|\phi) = [2\pi(\sigma^2 + \tau^2)]^{-m/2} e^{-\frac{1}{2(\sigma^2+\tau^2)} \sum_{i=1}^{m} (x_i - \phi)^2}$$

- Hence the MLE of ϕ is clearly $\hat{\phi} = \bar{X}$.
- The empirical Bayes estimator turns out to be

$$E[\mu_i|\mathbf{x}, \hat{\phi}] = \frac{\tau^2}{\sigma^2 + \tau^2} x_i + \frac{\sigma^2}{\sigma^2 + \tau^2} \bar{x}.$$

Examples: Empirical Bayes Estimation

Example 2(b):

- If we have a one-way classification with m groups and n observations per group, the previous example extends to

$$X_{ij}|\mu_i \overset{\text{indep}}{\sim} N(\mu_i, \sigma^2), \quad i = 1, \ldots, m, \, j = 1, \ldots, n$$
$$\mu_i \overset{\text{iid}}{\sim} N(\phi, \tau^2), \quad \sigma^2, \tau^2 \text{ known.}$$

- Then note that

$$\bar{X}_i \sim N\Big(\phi, \frac{\sigma^2}{n} + \tau^2\Big)$$

- Hence the empirical Bayes estimate of $\mu_i \, (i = 1, \ldots, m)$ is

$$\hat{\mu}_i = \frac{n\tau^2}{\sigma^2 + n\tau^2}\bar{x}_i + \frac{\sigma^2}{\sigma^2 + n\tau^2}\bar{x}$$

Examples: Empirical Bayes Estimation

- If τ^2 is unknown, note that $\dfrac{m-3}{\sum(\bar{X}_i - \bar{X})^2}$ is an unbiased estimator of $\dfrac{1}{\sigma^2 + n\tau^2}$, so we can use

$$\hat{\mu}_i = \left[1 - \frac{(m-3)\sigma^2}{\sum(\bar{X}_i - \bar{X})^2}\right]\bar{X}_i + \left[\frac{(m-3)\sigma^2}{\sum(\bar{X}_i - \bar{X})^2}\right]\bar{X}$$

  as the empirical Bayes estimator.

- On the other hand, if σ^2 is unknown, we can use

$$\hat{\mu}_i = \left[\frac{(m-3)n\tau^2}{\sum(\bar{X}_i - \bar{X})^2}\right]\bar{X}_i + \left[1 - \frac{(m-3)n\tau^2}{\sum(\bar{X}_i - \bar{X})^2}\right]\bar{X}$$

  as the empirical Bayes estimator.

Empirical Bayes vs./ Hierarchical Bayes Estimation

- Hierarchical Bayes (HB) and Empirical Bayes (EB) estimators both typically involve **shrinkage**.
- Some Bayesians feel EB is "less honest" since EB plugs in estimates of the hyperparameters **without** accounting for the **variability** associated with the estimate.
- HB places a **distribution** on the hyperparameters, and thus models the **uncertainty** in the hyperparameter values.
- See HB/EB Comparison for the Italian Marriage Data example on course web page.