Bayesian Estimation and Shrinkage

- The posterior mean of μ_j (given ϕ, τ^2, σ^2 and $\mathbf{y}_j$) is

$$E[\mu_j|\mathbf{y}_j, \phi, \tau^2, \sigma^2] = \frac{\frac{n_j \bar{y}_j}{\sigma^2} + \frac{\phi}{\tau^2}}{\frac{n_j}{\sigma^2} + \frac{1}{\tau^2}}$$
$$= \left(\frac{n_j/\sigma^2}{n_j/\sigma^2 + 1/\tau^2}\right)\bar{y}_j + \left(\frac{1/\tau^2}{n_j/\sigma^2 + 1/\tau^2}\right)\phi$$

- So the posterior mean of μ_j is pulled **away from** $\bar{y}_j$ and **toward** ϕ, the **mean** of the distribution of **all** the μ_j's.
- This is called **shrinkage**.
- How much is each μ_j shrunk? It depends on n_j.
- For schools with a large sample size (large n_j), shrinkage is minimal.
- For schools with a few students (small n_j), shrinkage is substantial.

Bayesian Estimation and Shrinkage

- **Example 1**: (Schools 82 vs. 46)

$$\text{Data: } \bar{y}_{82} = 38.76, \quad n_{82} = 5, \quad \hat{\mu}_{82} = 42.53$$
$$\bar{y}_{46} = 40.18, \quad n_{46} = 21, \quad \hat{\mu}_{46} = 41.31$$

- Note $\hat{\phi} = 48.12$.
- For school 82, we have substantial shrinkage toward $\hat{\phi}$.
- For school 46, we have less shrinkage toward $\hat{\phi}$.
- We might then rank school 82 ahead of school 46, because we doubt that $\bar{y}_{82}$ is a good estimate of school 82's true mean, being based on only 5 students.

Bayesian Estimation and Shrinkage

- **Example 2**: (Schools 67 and 51)

$$\text{Data: } \bar{y}_{67} = 65.02, \quad n_{67} = 4, \quad \hat{\mu}_{67} = 57.14$$
$$\bar{y}_{51} = 64.37, \quad n_{51} = 19, \quad \hat{\mu}_{51} = 61.84$$

- School 67 is shrunk down more toward $\hat{\phi}$.
- We expect school 51 to have a higher true mean even though its sample mean was lower.
- **Intuition**: Whom would you trust more to make a free throw, someone who has made 4 out of 4, or someone who has made 96 out of 100?

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) ψ, then the posterior is:

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

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

Examples: Empirical Bayes Estimation

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

$\lambda_i \stackrel{\text{iid}}{\sim}$ Gamma(α, β) with α known, β unknown.

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

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

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