Bayesian Statistics

Empirical Bayes

Nan Lin
Department of Mathematics
Washington University in St. Louis

Motivation

- Sometimes, after one chooses the form of the prior distribution, e.g. conjugate prior, there is not enough prior information to decide the hyperparameters.
- (Parametric) Empirical Bayesians then use data to determine the hyperparameters.

- Empirical Bayes can be viewed as a compromise between frequentist and Bayesian approaches.

General formulation

- Data model: $y|\theta \sim p(y|\theta)$
- Prior: $\theta|\eta \sim p(\theta|\eta)$
  - Hyperparameter: η
- Posterior: $p(\theta|y) = \frac{p(y|\theta)p(\theta|\eta)}{m(y|\eta)}$
- Data marginal distribution:

$$m(y|\eta) = \int p(y|\theta)p(\theta|\eta)d\theta$$

- How to decide η?

Solution

- Fully Bayesian approach → hierarchical Bayes
  - Adopt a hyperprior distribution $h(\eta)$
  - Posterior becomes a mixture of conditional posterior with mixing via the marginal posterior distribution of η :

$$p(\theta|\eta) = \frac{\int p(y|\theta)p(\theta|\eta)h(\eta)d\eta}{\iint p(y|u)p(u|\eta)h(\eta)dud\eta} = \int p(\theta|y,\eta)h(\eta|y)d\eta$$

- Empirical Bayes
  - Use the marginal distribution $m(y|\eta)$ to estimate η by $\hat{\eta} \equiv \hat{\eta}(y)$
  - For example, one may maximize the data marginal likelihood → **type-II maximum likelihood** or **marginal MLE**
  - **method of moments**

Example: estimating batting averages

- Efron and Morris (1975) reported the batting averages of 18 major league baseball players after their first 45 at bats.
  - Observed batting average: X_i
- Question: estimate their final batting average
- Data model: $X_i|\theta_i \sim N(\theta_i, \sigma^2)$ ← based on normal approximation
  - True batting average: θ_i
  - $\sigma^2 = (0.0659)^2$ is known
- Prior: $\theta_i \sim N(\mu, \tau^2)$
- How to get the Empirical Bayes estimate of θ_i?

Gaussian/Gaussian model

- Data model: $X_i|\theta_i \sim N(\theta_i, \sigma^2)$ i.i.d., $i = 1, \dots, p$
- Prior: $\theta_i|\mu \sim N(\mu, \tau^2)$ i.i.d. , $i = 1, \dots, p$
- Now, assume both σ^2 and τ^2 are known
- First, we need to drive the marginal distribution
  - $X_i|\mu \sim N(\mu, \sigma^2 + \mu^2)$
  - Jointly, $X_1, \dots, X_p|\mu$ are i.i.d.
- Next, we may estimate μ by either marginal MLE of method of moments, and one gets $\hat{\mu} = \bar{x}$
- Using this choice, the posterior is

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

where $B = \frac{\sigma^2}{\sigma^2+\tau^2}$. $\rightarrow \widehat{\theta_i} = B\bar{x} + (1 - B)x_i$

Gaussian/Gaussian model (cont)

- Now assume that τ^2 is unknown
- We will still have the same estimate $\hat{\mu} = \bar{x}$
- How to estimate τ^2?
- Using method of moments, one can have
  $E\left[\frac{(p-3)\sigma^2}{\sum_{i=1}^{p}(x_i-\bar{x})^2}\right] = \frac{\sigma^2}{\sigma^2+\tau^2} = B$
  - B and τ^2 are one-to-one
- Use $\widehat{B} = \frac{(p-3)\sigma^2}{\sum_{i=1}^{p}(x_i-\bar{x})^2}$, the Empirical Bayes estimate is then

$$\delta_i^E(X) = \left[\frac{(p-3)\sigma^2}{\Sigma(X_i - \bar{X})^2}\right]\bar{X} + \left[1 - \frac{(p-3)\sigma^2}{\Sigma(X_i - \bar{X})^2}\right]X_i.$$

Gaussian/Gaussian model (cont)

- Alternatively, if one uses marginal MLE to estimate τ^2, one will get $\hat{\tau}^2 = (s^2 - \sigma^2)^+ = \max(0, s^2 - \sigma^2)$
  - where $s^2 = \frac{\sum_{i=1}^{p}(x_i - \bar{x})^2}{p}$
- The method of moments estimate leads to the James-Stein estimator
- The marginal MLE leads to positive James-Stein estimator, which has a better performance then the James Stein estimator under the squared error loss.