Empirical Bayes

Haining

December 2021

- Empirical Bayes is a procedure for statistical inference where the prior distribution is estimated from the data.
- This is not "purely" Bayesian, since in a sense we are using the data to determine the prior specification.

Standard Bayesian method: the prior distribution is fixed before any data are observed.

Empirical Bayesian:

1. making prior (with hyperparameters $\boldsymbol{\eta}$) on the parameter $\boldsymbol{\theta}$ of the prior
2. estimate the hyperparameters using techniques such as maximum likelihood
3. can be viewed as a hierarchical model $\boldsymbol{\eta} \to \boldsymbol{\theta} \to \mathbf{X}$

Example 1

Prior: Identical Independent $X_i \sim \text{Poi}(\lambda_i), i = 1, ..., n$

Assume a prior over the prior $\lambda_i \sim \text{Gamma}(\alpha, \beta)$ with α known, β unknown hyperparameters (iid)

$$
\begin{aligned}
p(X_i = x_i|\beta) &= \int_0^\infty p(X_i|\lambda_i)p(\lambda_i|\alpha,\beta)\,d\lambda_i \\
&= \int_0^\infty (\frac{\lambda_i^{x_i}}{x_i!}e^{-\lambda_i})(\frac{\beta^\alpha}{\Gamma(\alpha)}\lambda_i^{\alpha-1}e^{-\beta\lambda_i})\,d\lambda_i \\
&= \frac{\beta^\alpha}{x_i!\Gamma(\alpha)}\int_0^\infty \lambda_i^{x_i+\alpha-1}e^{-(\beta+1)\lambda_i}\,d\lambda_i \\
&= \frac{\beta^\alpha}{x_i!\Gamma(\alpha)}\frac{\Gamma(x_i+\alpha)}{(\beta+1)^{x_i+\alpha}}\int_0^\infty \text{Gamma}(x_i+\alpha,\beta+1)\,d\lambda_i \\
&= \frac{\beta^\alpha}{x_i!\Gamma(\alpha)}\frac{\Gamma(x_i+\alpha)}{(\beta+1)^{x_i+\alpha}} \\
&= \frac{\Gamma(x_i+\alpha)}{x_i!\Gamma(\alpha)}(\frac{\beta}{\beta+1})^\alpha(\frac{1}{\beta+1})^{x_i} \\
&= \binom{x_i+\alpha-1}{\alpha-1}(\frac{\beta}{\beta+1})^\alpha(\frac{1}{\beta+1})^{x_i} \qquad \text{(Negative binomial)}
\end{aligned}
$$

The joint likelihood

$$p(X_1, X_2, ..., X_n|\beta) = (\prod_{i=1}^{n} \binom{x_i + \alpha - 1}{\alpha - 1})(\frac{\beta}{\beta+1})^{n\alpha}(\frac{1}{\beta+1})^{\sum x_i}$$

Estimate hyperparameter β using maximum likelihood:

Set the derivative to 0, obtain

$$\begin{aligned} n\alpha(\frac{\beta}{\beta+1})^{n\alpha-1}\frac{1}{(\beta+1)^2}(\frac{1}{\beta+1})^{\sum x_i} - \sum x_i(\frac{\beta}{\beta+1})^{n\alpha}(\frac{1}{\beta+1})^{\sum x_i+1} = 0 \\ n\alpha = \beta\sum x_i \Rightarrow \hat{\beta} = \alpha/\bar{x} \end{aligned}$$

For each λ_i, using prior $\lambda_i \sim \text{Gamma}(\alpha, \hat{\beta})$, the posterior

$$\begin{aligned} p(\lambda_i|x_i) &= p(x_i|\lambda_i)p(\lambda_i)/\int_0^\infty p(x_i|\lambda_i)p(\lambda_i)\,d\lambda_i \\ &\propto (\frac{\lambda_i^{x_i}}{x_i!}e^{-\lambda_i})(\frac{\hat{\beta}^\alpha}{\Gamma(\alpha)}\lambda_i^{\alpha-1}e^{-\hat{\beta}\lambda_i}) \\ &\propto \lambda_i^{x_i+\alpha-1}e^{-(\hat{\beta}+1)\lambda_i} \\ &\sim \text{Gamma}(x_i+\alpha, \hat{\beta}+1) \end{aligned}$$

The empirical Bayes estimator for λ_i can be the posterior mean

$$\frac{x_i+\alpha}{\hat{\beta}+1} = \frac{x_i+\alpha}{\alpha/\bar{x}+1}$$