Series 4

1. **Empirical Bayes**[1]

Assume that the data $X_1, \dots, X_n$ are independent random variables with a binomial distribution Binomial(m, θ_i) and that the parameters θ_i have priors θ_i i.i.d. $\sim$ Beta(α, β). Compute posterior means for θ_i using the empirical Bayes method. Follow the steps in the following to do this.

a. Show that the marginal distribution of X_i is beta-binomial (see Series 1 for details). Then, use the method of moments to obtain estimates $\hat{\alpha}$ and $\hat{\beta}$.

b. Next, calculate the posterior means $\hat{\theta}_i$ as point estimates for θ_i using $\hat{\alpha}$ and $\hat{\beta}$. Compare them with the corresponding MLEs.

Solution

a. We can show that the marginal distribution of X_i is the beta-binomial distribution with density

$$\binom{m}{x} \frac{B(\alpha + x, \beta + m - x)}{B(\alpha, \beta)}.$$

It has the following first two moments

$$\begin{aligned}
\mu_1 = E(X_i) &= m \frac{\alpha}{\alpha + \beta}, \\
\mu_2 = E(X_i^2) &= m \frac{\alpha(m(1 + \alpha) + \beta)}{(\alpha + \beta)(1 + \alpha + \beta)}.
\end{aligned}$$

The corresponding empirical moments are

$$\begin{aligned}
\hat{\mu}_1 &= \frac{1}{n} \sum_{i=1}^{n} x_i, \\
\hat{\mu}_2 &= \frac{1}{n} \sum_{i=1}^{n} x_i^2,
\end{aligned}$$

Solving for α and β, we obtain the following moment estimators

$$\begin{aligned}
\hat{\alpha} &= \frac{(m\hat{\mu}_1 - \hat{\mu}_2)\hat{\mu}_1}{m(\hat{\mu}_2 - \hat{\mu}_1(\hat{\mu}_1 + 1)) + \hat{\mu}_1^2}, \\
\hat{\beta} &= \frac{(m\hat{\mu}_1 - \hat{\mu}_2)(m - \hat{\mu}_1)}{m(\hat{\mu}_2 - \hat{\mu}_1(\hat{\mu}_1 + 1)) + \hat{\mu}_1^2}.
\end{aligned}$$

[1] Based on Exercise 12 in Section 6.8 of Held and Sabanes Bove (2014).

b. From the lecture, we know that $\theta_i|x_i \sim \text{Beta}(\alpha+x_i, \beta+m-x_i)$. The posterior mean as empirical Bayes estimator is thus given by

$$\hat{\theta}_i = \frac{\hat{\alpha}+x_i}{\hat{\alpha}+\hat{\beta}+m}.$$

In contrast, the MLE is given by

$$\hat{\theta}_{i,MLE} = \frac{x_i}{m}.$$

Hence, the empirical Bayes estimator is equal to the MLE if and only if $\hat{\alpha} = \hat{\beta} = 0$ which corresponds to an improper prior distribution. In the other cases, the empirical Bayes estimator

$$\frac{\hat{\alpha}+x_i}{\hat{\alpha}+\hat{\beta}+m} = \frac{\hat{\alpha}+\hat{\beta}}{\hat{\alpha}+\hat{\beta}+m}\frac{\hat{\alpha}}{\hat{\alpha}+\hat{\beta}} + \frac{m}{\hat{\alpha}+\hat{\beta}+m}\frac{x_i}{m}$$

is a weighted average of the prior mean $\frac{\hat{\alpha}}{\hat{\alpha}+\hat{\beta}}$ and the MLE $\frac{x_i}{m}$. The weights are proportional to the prior sample size $m_0 = \hat{\alpha}+\hat{\beta}$ and the data sample size m, respectively.

2. **Bayesian linear regression model**

Consider the linear regression model:

$$y = X\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2 I_n),$$

where y is an $n \times 1$ vector of responses, X is the $n \times p$ design matrix, β the $p \times 1$ regression parameter, and ε is the $n \times 1$ vector of errors. Note that in this formulation the intercept term is either included in the design matrix X or there is no intercept.

Further, assume a uniform prior distribution on $(\beta, \log(\sigma))$. I.e.,

$$p(\beta, \sigma^2) \propto \frac{1}{\sigma^2}.$$

Derive the following distributions:

a. The conditional posterior of β given σ^2: $\pi(\beta \mid \sigma^2, y)$.

b. The marginal posterior of σ^2: $\pi(\sigma^2 \mid y)$.

Solution

As in the lecture, we can show that the likelihood can be written as

$$(\sigma^2)^{-n/2} \exp\left(-\frac{s^2 + (\beta-\widehat{\beta})X^T X(\beta-\widehat{\beta})}{2\sigma^2}\right),$$

where
$$s^2 = (y - X\widehat{\beta})^T (y - X\widehat{\beta}),$$
and
$$\widehat{\beta} = (X^T X)^{-1} X^T y.$$
The joint posterior is therefore given by
$$\pi(\beta, \sigma^2 \mid y) \propto (\sigma^2)^{-n/2-1} \exp\left(-\frac{s^2 + (\beta - \widehat{\beta}) X^T X (\beta - \widehat{\beta})}{2\sigma^2} \right). \tag{1}$$

a. Based on (1), we find that
$$\beta \mid y, \sigma^2 \sim \mathcal{N}\left(\widehat{\beta}, \sigma^2 (X^T X)^{-1}\right)$$

b. Integrating β out from the joint posterior in (1), gives
$$\pi(\sigma^2 \mid y) \propto (\sigma^2)^{-n/2-1+p/2} \exp\left(-\frac{s^2}{2\sigma^2} \right).$$
From this, we conclude that
$$\sigma^{-2} \mid y \sim \text{Gamma}\left(\frac{n-p}{2}, \frac{s^2}{2} \right).$$

3. **Regularization and Bayesian regression model**

Consider again the linear regression model:
$$y = X\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2 I_n),$$
where y is an $n \times 1$ vector of responses, X is the $n \times p$ design matrix, β the $p \times 1$ regression parameter, and ε is the $n \times 1$ vector of errors.

a. In frequentist statistics, the Ridge regression coefficients are chosen as minimizers of
$$\|y - X\beta\|^2 + \lambda \|\beta\|^2, \quad \lambda \geq 0.$$
Show that for some λ, the Ridge regression coefficients are equivalent to the posterior mode, also called the maximum a posteriori (MAP) estimator, if one assumes the following normal prior for the coefficients
$$\beta \sim N(0, \sigma_\beta^2 I_p).$$

b. The (frequentist) Lasso estimates are defined as minimizers of
$$\|y - X\beta\|^2 + \lambda \|\beta\|_1,$$

where

$$\|\beta\|_1 = \sum_{k=1}^{p} |\beta_k|.$$

Show that for some λ, the Lasso coefficients are equivalent to the MAP estimator if one assumes the following independent double exponential, also called Laplace, priors for the coefficients

$$\pi(\beta) = \prod_{k=1}^{p} \frac{1}{2b} e^{-|\beta_k|/b}.$$

Solution

a. We have

$$\begin{aligned}
\arg\max_{\beta} \pi(\beta \mid y) &= \arg\max_{\beta} \left(\log(f(y \mid \beta)\pi(\beta)\right) \\
&= \arg\max_{\beta} \left(-\|y - X\beta\|^2/\sigma^2 - \|\beta\|^2/\sigma_\beta^2\right) \\
&= \arg\min_{\beta} \left(\|y - X\beta\|^2 + \frac{\sigma^2}{\sigma_\beta^2}\|\beta\|^2\right).
\end{aligned}$$

b. We have

$$\begin{aligned}
\arg\max_{\beta} \pi(\beta \mid y) &= \arg\max_{\beta} \left(\log(f(y \mid \beta)\pi(\beta)\right) \\
&= \arg\max_{\beta} \left(-\frac{\|y - X\beta\|^2}{2\sigma^2} - \sum_{k=1}^{p} |\beta_k|/b\right) \\
&= \arg\min_{\beta} \left(\|y - X\beta\|^2 + \frac{2\sigma^2}{b}\|\beta\|_1\right).
\end{aligned}$$