Series 4

1. **Empirical Bayes**[1]

   Assume that the data $X_1, \dots, X_n$ are independent random variables with a binomial distribution $\text{Binomial}(m, \theta_i)$ and that the parameters θ_i have priors θ_i i.i.d. $\sim \text{Beta}(\alpha, \beta)$. 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.

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: $\pi(\sigma^2 \mid y)$.

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.

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

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

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