BST 140.752
Problem Set 5

1 LMM

1. Let $Y_{ij} = \mu + u_i + \epsilon_{ij}$ for $u_i \sim N(0, \sigma_u^2)$ and $\epsilon_{ij} \sim N(0, \sigma^2)$. Calculate the BLUP for u_i

2. Let $Y = X\beta + Zu + \epsilon$ for $u \sim N(0, \Sigma_u)$ and $\epsilon \sim N(0, \sigma^2 I)$. Calculate the BLUP for u.

3. Load the Rail data set in R. Fit a mixed model of the form from question 1. Compare a the estimates of the mean for each rail with the empirical mean.

4. Load the pixel data set in R. Fit a linear mixed effect model where you have $Y_{ijk} = \beta_0 + \beta_1 x_k + u_i + u_{ij} + \epsilon_{ijk}$ where Y_{ijk} is pixel, i is dog, j is side and k is day index and x_k is day. Fit the model and interpret the results.

5. Consider the model $Y_i = \mu + \epsilon_{ij}$. Consider putting a so-called "flat" prior on μ. That is acting like a distribution that is 1 from $-\infty$ to $+\infty$ is a valid density. Calculate the distribution marginalized over μ and show that it is the same likelihood used to obtain the REML estimates.

6. Derive the BLUP estimates for a linear mixed effect model of the form $Y = X\beta + ZU + \epsilon$ where $U \sim N(0, \Sigma_u)$ and $\epsilon \sim N(0, \sigma^2 I)$ and assuming known variance components.

7. Find the BLUPs for the Rail data set in R. Compare the BLUP estimates with those obtained using a fixed effect estimate for the U_i.

8. Consider the model from Question 1. Show that the estimate for σ^2 is not consistent as $i \to \infty$ for equal numbers of observations per cluster.

2 Smoothing

1. Let `seq(0, 6 * pi, length = 1000)`. Simulate a sin wave and do the following:

   A. Fit a smoother using linear and quadratic regression splines.

   B. Fit a smoother using bsplines and natural splines.

   C. Plot the columns of the design matrix for the different bases.

2. Take the `mtcars` dataset. Fit a penalized spline model using `mgcv` with `mpg` as the outcome and additive smooth functions of weight and horsepower. Visualize the fits.

3. Show that the function $\sum_{k=0}^{p} \beta_k x^k + \sum_{k=1}^{K} \gamma_k (x - \xi_k)_+^p$ has $p - 1$ derivatives for $p > 1$.

3 Bayesian models

1. Let $Y \mid \mu, \theta \sim N(\mu J_n, \theta^{-1} I)$ be an $n \times 1$ vector. Further let $\mu \mid \theta \sim N(\mu_0, \theta^{-1}\sigma_0^2)$ and $\theta \sim$ Gamma(α_0, β_0). Derive the joint posterior distribution for μ and θ as well as the marginal posterior distribution for μ and θ.

2. Let p denote the unknown proportion of rocks in a riverbed that are sedimentary in type. Suppose that $X = 12$ of a sample of $n = 20$ rocks collected in random locations are found to be sedimentary in type. Display the Bayes interval for a uniform and Jeffrey's prior. Give 95% HPD and equi-tail intervals for both.

3. A cohort study was performed to evaluate the relation between oral-contraceptive use and breast cancer in the age group 40-44. At enrollment, women were classified as current users, past users or never users of oral contraceptives. Current users had 13 cancer cases out of 4,671 person years while never users 113 cancer cases out of 98,091 years. Assume that cancer counts given t person years is Poisson(λt) where λ is the incidence rate. Assume independent Gamma priors for the incident rate for the never and current users and create the following (trying different values for the hyper-parameters):

   a. The posterior for the incidence rate for the never users.

   b. The posterior for the incidence rate for the current users.

   c. The posterior for the relative rate comparing the never to current users.

   d. Give 95% equi-tail credible intervals for all three above.

4. Take the `mtcars` dataset. Fit a Bayesian regression model with `mpg` as the outcome and `wt` as the predictor. Display the posterior for the slope, intercept and residual variance. Assume a normal/gamma prior.