Project 4 of ST 755, Spring 2008

Due: Tuesday, 4/29/2008

Consider the following nonparametric regression problem

$$y_i = f(t_i) + e_i \quad i = 1, ..., n$$

where $0 < t_1 < t_2 < \cdots < t_n < 1$, $f(t)$ is a smooth function and e_i's are iid from $\mathrm{N}(0, \sigma^2)$. For an $m \geq 1$, we would like to estimate $f(t)$ by minimizing the following penalized residual sum of squares

$$Q(f) = \sum_{i=1}^{n} [y_i - f(t_i)]^2 + \lambda \int_0^1 [f^{(m)}(t)]^2 dt.$$

In class, we show that the minimizer $f_\lambda(t)$ takes the following form

$$f_\lambda(t) = \sum_{j=1}^{m} d_j \phi_j(t) + \sum_{i=1}^{n} c_i R^1(t_i, t),$$

where $\phi_j(t) = t^{j-1}/(j-1)!$ and

$$R^1(t, s) = \frac{1}{[(m-1)!]^2} \int_0^1 (t-u)_+^{m-1} (s-u)_+^{m-1} du.$$

Do the following:

1. What kind of functions $f_\lambda(t)$ will be in $[0, t_1]$, $[t_i, t_{i+1}]$ and $[t_n, 1]$ for $m = 1$ and $m = 2$?

2. Suppose we only want to get the estimate of $f = (f_\lambda(t_1), f_\lambda(t_2), \cdots, f_\lambda(t_n))^T$. Show that

   $$\hat{f} = T\hat{d} + \hat{b},$$

   where $\hat{d}$ and $\hat{b}$ are BLUE and BLUP from the following linear mixed model

   $$y = Td + b + e,$$

   with $T = \phi_j(t_{i_{n \times m}}$, $b \sim \mathrm{N}(0, \sigma_b^2 \Sigma)$, $e \sim \mathrm{N}(0, \sigma_e^2 I)$, and $\sigma_b^2 = \sigma_e^2 \lambda$. Find the expression for the (i, j)th element of Σ for $m = 1$ and $m = 2$.

3. For given data, we can estimate σ_b^2 and σ_e^2 using REML methodology. Develop Fisher scoring algorithm for estimating σ_b^2 and σ_e^2 and hence $\hat{f}$.

4. Assume $n = 99$, $t_i = i/100$, $f(t) = sin(2\pi t)$ and $\sigma_e^2 = 1$. Implement the REML estimation in (3) for $m = 1$ and $m = 2$. Conduct a simulation study with 20 simulation runs and get the average of 20 estimated $f(t)$ for $m = 1$ and $m = 2$. Superimpose the bias in a plot. Any comment?