

1. Suppose that in the model  $Y_i = \alpha + \beta x_i + \epsilon_i$ ,  $i = 1, \dots, n$ , the errors  $\epsilon_i$  are independent and normally distributed with mean 0, but that  $\text{var}(\epsilon_i) = \sigma^2/w_i$  where  $w_1, \dots, w_n > 0$  are known constants.

Show that the maximum likelihood estimates of  $\alpha$  and  $\beta$  can be found by minimizing

$$\sum_{i=1}^n w_i (y_i - \alpha - \beta x_i)^2.$$

Can you give two examples of situations in which this model might arise?

2. Suppose the straight-line model

$$Y_i = a + b(x_i - \bar{x}) + \epsilon_i, \quad i = 1, \dots, n$$

is fitted using maximum likelihood, where  $\epsilon_1, \dots, \epsilon_n \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$ . Suppose we estimate the position of the line at new value  $x_0$  of  $x$  by  $\hat{\mu}(x_0)$ , where

$$\hat{\mu}(x_0) = \hat{a} + \hat{b}(x_0 - \bar{x}).$$

Derive an expression for the variance of  $\hat{\mu}(x_0)$ .

Sketch the regression line  $y = \hat{\mu}(x)$  together with  $y = \hat{\mu}(x) + 2 \text{SE}(\hat{\mu}(x))$  and  $y = \hat{\mu}(x) - 2 \text{SE}(\hat{\mu}(x))$  as a function of  $x$ .

3. (a) In the model  $Y_i = \alpha + \beta x_i + \epsilon_i$ ,  $i = 1, \dots, n$ , where  $\epsilon_i \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$ , show that the maximum likelihood estimator  $\hat{\sigma}^2$  of  $\sigma^2$  is given by

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta} x_i)^2$$

where  $\hat{\alpha}, \hat{\beta}$  are the usual estimates of  $\alpha, \beta$ .

- (b) Show that  $E(\hat{\sigma}^2) = \left(\frac{n-2}{n}\right) \sigma^2$  and deduce an unbiased estimator of  $\sigma^2$ .  
 [Hint: use the result from lectures that  $\text{var}(e_i) = \sigma^2(1 - h_i)$ .]

4. Let

$$Y_i = f(x_i) + \epsilon_i, \quad i = 1, \dots, n$$

where  $f(x)$  is some function (not necessarily  $f(x) = \alpha + \beta x$ ) and where  $\epsilon_i \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$ . Suppose that  $f$  is estimated by some estimator  $\hat{f}$  (where  $\hat{f}$  depends on  $(x_i, y_i)$ ,  $i = 1, \dots, n$ ).

The mean squared error for a new  $Y$  at a new value of  $x$ , say  $Y_0 = f(x_0) + \epsilon_0$ , is defined by  $E[(Y_0 - \hat{f}(x_0))^2]$ . Here  $\epsilon_0 \sim N(0, \sigma^2)$  independent of  $\epsilon_1, \dots, \epsilon_n$ . Show that

$$E[(Y_0 - \hat{f}(x_0))^2] = \text{Var}(\hat{f}(x_0)) + [\text{Bias}(\hat{f}(x_0))]^2 + \sigma^2$$

where

$$\begin{aligned} \text{Bias}(\hat{f}(x_0)) &= E[\hat{f}(x_0)] - f(x_0) \\ \text{Var}(\hat{f}(x_0)) &= E[\{\hat{f}(x_0) - E[\hat{f}(x_0)]\}^2]. \end{aligned}$$

[Hint: start from  $E[(Y_0 - \hat{f}(x_0))^2] = E[\{(Y_0 - f(x_0)) + (f(x_0) - \hat{f}(x_0))\}^2]$ .]

5. (Using R or Matlab) Complete Q4 on Sheet 4.