Prelims Statistics and Data Analysis – Sheet 5 TT 2019

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

Show that the maximum likelihood estimates of α and β 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

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

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

$$\widehat{\mu}(x_0) = \widehat{a} + \widehat{b}(x_0 - \overline{x}).$$

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

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

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

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

where $\widehat{\alpha}, \widehat{\beta}$ are the usual estimates of α, β.

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

4. Let

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

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

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 $E[(Y_0 - \widehat{f}(x_0))^2]$. Here $\epsilon_0 \sim N(0, \sigma^2)$ independent of $\epsilon_1, \ldots, \epsilon_n$. Show that

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

where

$$\text{Bias}\big(\widehat{f}(x_0)\big) = E\big[\widehat{f}(x_0)\big] - f(x_0)$$
$$\text{Var}\big(\widehat{f}(x_0)\big) = E\big[\big\{\widehat{f}(x_0) - E\big[\widehat{f}(x_0)\big]\big\}^2\big].$$

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

5. (Optional: using R or Matlab) Complete Q4 on Sheet 4.