1 Pre-quiz for Advanced Data Analysis

1. Derive the maximum likelihood estimators of β_0, β_1 and σ^2 for the simple linear regression model of the form $y_i = \beta_0 + \beta_1 x_i + e_i$ with $e_i \sim N(0, \sigma^2)$.

2. Derive (the method of) moment estimators of β_0, β_1 and σ^2 based on the first moment of Y_i and on the equations $\frac{1}{n}\sum_{i=1}^{n} x_i e_i = 0$ and $\frac{1}{n}\sum_{i=1}^{n} e_i^2 = \sigma^2$ with $e_i = y_i - (\beta_0 + \beta_1 x_i)$

3. What distributions have the following mean-variance relationships (of a random variable Y with mean μ)? a) $\mathrm{Var}(Y) = \mu$ b) $\mathrm{Var}(Y) = constant$, c) $\mathrm{Var}(Y) = \mu(1-\mu)$?

4. A certain disease occurs 4% in the population. A diagnostic test gives a positive result with 95% probability if a subject has the disease, and with 2% if the subject does not have the disease. If a randomly selected person obtains a positive test result, using Bayes' theorem what is the probability that this person actually has the disease?

1. $L(\beta_0, \beta_1, \sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - (\beta_0 + \beta_1 x_i))^2$
Differentiate

$$\frac{\partial L}{\partial \beta_0} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - (\beta_0 + \beta_1 x_i)) = 0 \tag{1}$$

$$\frac{\partial L}{\partial \beta_1} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - (\beta_0 + \beta_1 x_i))x_i = 0 \tag{2}$$

$$\frac{\partial L}{\partial \sigma^2} = \frac{n}{2}\frac{1}{\sigma^2} - \frac{1}{\sigma^4}\sum_{i=1}^{n}(y_i - (\beta_0 + \beta_1 x_i))^2 = 0 \tag{3}$$

setting to zero and simplifying

$$\bar{y} = \beta_0 + \beta_1\bar{x} \tag{4}$$

$$\bar{xy} = \beta_0\bar{x} + \beta_1\bar{x^2} \tag{5}$$

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} e_i^2 \tag{6}$$

It follows $\beta_0 = \bar{y} - \beta_1\bar{x}$ now plugging β_0 into (5) gives

$$\begin{aligned}\bar{xy} &= (\bar{y} - \beta_1\bar{x})\bar{x} + \beta_1\bar{x^2} \\ &= \bar{y}\bar{x} + \beta_1(\bar{x^2} - \bar{x}^2)\end{aligned}$$

Re-arranging gives

$$\beta_1 = \frac{\bar{xy} - \bar{x}\bar{y}}{\bar{x^2} - \bar{x}^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{C_{XY}}{S_X^2}$$

So the estimators are

$$\begin{aligned}\hat{\beta}_1 &= \frac{C_{XY}}{S_X^2} \\ \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1\bar{x} \\ \hat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n}\hat{e_i}^2\end{aligned}$$

2. We see that the maximum likelihood equations (1), (2) and (3) are equivalent to the moment equations

$$\frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}\sum_{i=1}^{n} \beta_0 + \beta_1 x_i$$

$$\frac{1}{n}\sum_{i=1}^{n} x_i e_i = 0$$

$$\frac{1}{n}\sum_{i=1}^{n} e_i^2 = 0$$

and hence it the estimators are identical.

3. What distributions have the following mean-variance relationships? a) Poisson b) Normal c) Binary.

4. Let D be the event the person has disease, T the event of receiving a positive result.

$$P(D|T) = \frac{P(D)P(T|D)}{P(D)P(T|D) + P(\bar{D})P(T|\bar{D})}$$
$$= \frac{0.04 \times 0.95}{0.04 \times 0.95 + (1 - 0.04) \times 0.02} = 0.664$$