

MATH3821 Assignment 1

Stephen Sung

Question 1

For Simple Linear Regression model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ where $\epsilon_i \sim N(0, \sigma^2)$.

- a) Let $\beta_0 = \alpha - \beta_1 \bar{x}$. Then the SLR model can be expressed as $y_i = \alpha + \beta_1(x_i - \bar{x}) + \epsilon_i$.
- b) α is the intercept of the new model?
- c) To find the closed form formula of the LSE,

$$RSS(\beta_1) = \sum_{i=1}^n [y_i - (\alpha + \beta_1(x_i - \bar{x}))]^2$$
$$\frac{dRSS(\beta_1)}{d\alpha} = -2 \sum_{i=1}^n (y_i - (\alpha + \beta_1(x_i - \bar{x}))) \quad (1)$$

$$\frac{dRSS(\beta_1)}{d\beta_1} = -2 \sum_{i=1}^n (y_i - (\alpha + \beta_1(x_i - \bar{x}))(x_i - \bar{x})) \quad (2)$$

Let Equation (1) = 0

$$\begin{aligned} -2 \sum_{i=1}^n (y_i - (\hat{\alpha} + \hat{\beta}_1(x_i - \bar{x}))) &= 0 \\ \sum_{i=1}^n y_i - \sum_{i=1}^n \hat{\alpha}_i - \sum_{i=1}^n \hat{\beta}_1 x_i + \sum_{i=1}^n \hat{\beta}_1 \bar{x} &= 0 \\ n\hat{\alpha}_i = \sum_{i=1}^n y_i - \hat{\beta}_1 \sum_{i=1}^n x_i + n\hat{\beta}_1 \bar{x} \\ \hat{\alpha}_i &= \frac{1}{n} \sum_{i=1}^n y_i - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 \bar{x} \\ \hat{\alpha}_i &= \bar{y} \end{aligned}$$

Let Equation (2) = 0

$$\begin{aligned} -2 \sum_{i=1}^n (y_i - (\hat{\alpha} + \hat{\beta}_1(x_i - \bar{x}))(x_i - \bar{x})) &= 0 \\ \sum_{i=1}^n y_i(x_i - \bar{x}) - \sum_{i=1}^n \hat{\alpha}(x_i - \bar{x}) - \sum_{i=1}^n \hat{\beta}_1(x_i - \bar{x})^2 &= 0 \\ \hat{\beta}_1 \sum_{i=1}^n (x_i - \bar{x})^2 &= \sum_{i=1}^n (y_i - \hat{\alpha})(x_i - \bar{x}) \end{aligned}$$

since $\hat{\alpha} = \bar{y}$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

d)

$$\begin{aligned} \text{Var}[\hat{\alpha}] &= \text{Var}\left[\frac{1}{n} \sum_{i=1}^n y_i\right] \\ &= \frac{1}{n^2} \text{Var}\left[\sum_{i=1}^n y_i\right] \end{aligned}$$

Since all y_i 's are uncorrelated

$$\begin{aligned} &= \frac{1}{n^2} \sum_{i=1}^n \text{Var}[y_i] \\ &= \frac{1}{n^2} \sum_{i=1}^n \sigma^2 \\ &= \frac{n\sigma^2}{n^2} \\ &= \frac{\sigma^2}{n} \end{aligned}$$

Therefore $\text{Var}[\hat{\alpha}] = \frac{\sigma^2}{n}$.

We note that $\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x}) = \sum_{i=1}^n (x_i - \bar{x})y_i$

Question 2

Given n independent binary random variables $Y_1 \cdots Y_n$ with

$$P(Y_i = 1) = \pi_i \text{ and } P(Y_i = 0) = 1 - \pi_i$$

The probability function of Y_i is:

$$\pi_i^{Y_i} (1 - \pi_i)^{1-Y_i}$$

where $Y_i = 0$ or 1

- a) For a probability function to belong to the exponential family of distributions, it must follow the formula:

$$f(y; \theta, \phi) = K(y, \frac{p}{\phi}) \exp\left(\frac{p}{\phi} \{y\theta - c(\theta)\}\right)$$

For the given probability density function:

$$\begin{aligned} f(y; \pi) &= \pi_i^y (1 - \pi_i)^{1-y} \\ &= \exp(\log \pi_i^y (1 - \pi_i)^{1-y}) \\ &= \exp(\log \pi_i^y + \log(1 - \pi_i)^{1-y}) \\ &= \exp(y \log \pi_i + (1 - y) \log(1 - \pi_i)) \\ &= \exp\left(y \log\left(\frac{\pi}{1 - \pi}\right) + \log(1 - \pi)\right) \end{aligned}$$

With $p = 1$ and $\phi = 1$, the above equation follows the form of the exponential family of distribution where $K(y, \frac{p}{\phi}) = 1$, $\theta = \log(\frac{\pi}{1-\pi})$ and $c(\theta) = -\log(1 - \pi) = -\log(1 - \frac{e^\theta}{1+e^\theta})$ where $\pi = \frac{e^\theta}{1+e^\theta}$.

- b) As seen in 2a, the naturalised parameter is $\theta = \log(\frac{\pi}{1-\pi})$
- c) As seen in 2a, the cumulant generator is $c(\theta) = -\log(1 - \frac{e^\theta}{1+e^\theta})$. Since $E[Y] = c'(\theta)$, $c'(\theta) = -(\frac{e^\theta}{1+e^\theta}) = -(-\pi) = \pi$. Therefore, $E[Y] = \pi$.
- d) Given the link function:

$$g(\pi) = \log(\frac{\pi}{1-\pi}) = e^{x^T \beta}$$

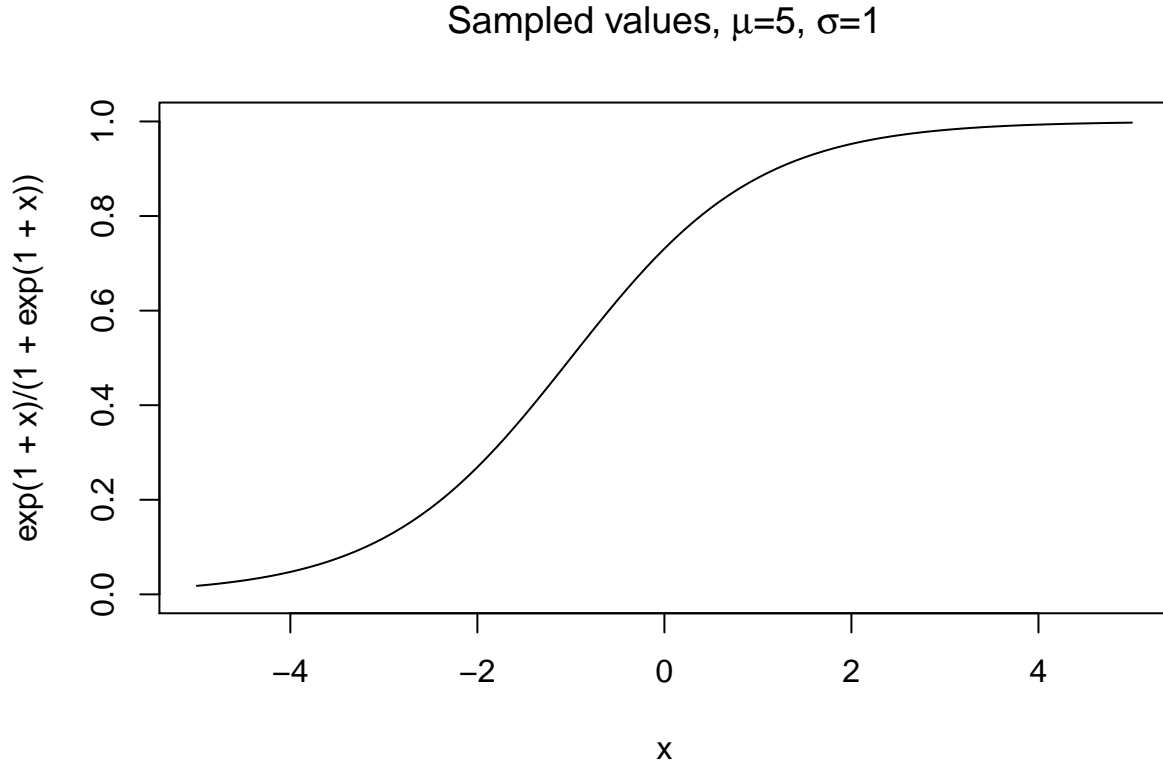
it can be rearranged in terms of the probability π ,

$$\begin{aligned} e^{x^T \beta} &= \log(\frac{\pi}{1-\pi}) \\ e^{x^T \beta} - \pi e^{x^T \beta} &= \pi \\ \pi &= \frac{e^{x^T \beta}}{1 + e^{x^T \beta}} \end{aligned}$$

e)

```
curve(exp(1+x)/(1+exp(1+x)), xlim = c(-5, 5), ylim = c(0, 1), main=expression(paste("Sampled values, ",
"=1")))

```



It shows the log odds of the insecticide working with a given probability?

f) The following probability density function:

$$f(y; \theta) = \frac{1}{\phi} \exp \left(\frac{(y - \theta)}{\phi} - \exp \left[\frac{(y - \theta)}{\phi} \right] \right)$$

is NOT in the exponential family of distributions as it does not follow the form of a probability density function in the exponential family.