

Stat W 4201, Spring 2016

Assignment #8: Due April 6

1. Chapter 12, problem 17
2. Chapter 12, problem 20
3. Chapter 20, problem 11
4. Chapter 20, problem 15
5. Consider a random vector (X, Y, Z) , where Y and Z are marginally Bernoulli random variables such that

$$\begin{aligned} P(Y = 1|X) &= \frac{e^{\alpha + X^T \beta}}{1 + e^{\alpha + X^T \beta}}, \\ P(Z = 1|X, Y) &= \pi(X)\phi(Y). \end{aligned}$$

Show that

$$P(Y = 1|X, Z = 1) = \frac{e^{\alpha^* + X^T \beta}}{1 + e^{\alpha^* + X^T \beta}},$$

and give the expression of α^* .

6. Consider a properly normalized density function $f(x)$ on the real line and its natural exponential family,

$$f(x|\theta) = f(x)e^{\theta x - \varphi(\theta)}.$$

- (a) What conditions do you need for $f(x|\theta)$ to be a well defined density function?
- (b) What is the choice of $\varphi(\theta)$ such that $f(x|\theta)$ is a properly normalized density function, that is,

$$\int f(x|\theta)dx = 1.$$

- (c) Show that with the choice of φ in (b) satisfies

$$\begin{aligned} \varphi'(\theta) &= \int x f(x|\theta)dx, \\ \varphi''(\theta) &= \int (x - \varphi'(\theta))^2 f(x|\theta)dx, \end{aligned}$$

where φ' and φ'' are the first and second derivatives of φ . If $\varphi'(\theta)$ is a monotone strictly increasing function, the variance is determined by its mean for natural exponential families.

- (d) Consider the following dataset

Y	X
0	-2
0	-1
1	1
1	2

and a generalized linear model

$$\text{logit}^{-1} \{P(Y = 1|X)\} = \beta_1 X.$$

Report the “maximum likelihood estimate” of β_1 from computer using `glm(Y~X-1,binomial)` (or any other softwares you use). Does the estimate make sense to you? Can we obtain a reasonable estimate of $P(Y = 1|X = 0.5)$ from the data? Why? Plot the log-likelihood as a function of β_1 . Does the plot explain this phenomenon? This is one of the few situations when the MLE does not exist. It happens sometimes especially when the sample size is small. Provide a sufficient and necessary condition that the MLE of logistic regression exists.

This problem provide a warning for your future work. When the model becomes more complicated, some parameters may not be estimable/identifiable for different reasons. The data may not have enough information to identify the best model, which is our current situation. It can also be caused by the existence of redundant parameters, that is $f(x|\theta_1) = f(x|\theta_2)$, for some $\theta_1 \neq \theta_2$. You should pay particular attention especially to the first situation. The estimates provided by software may not be meaningful and yield misleading results.