

Assignment 3
Due at 18:00 March 23, 2020

1. Suppose $\{X_n, n \geq 1\}$ is a sequence of random variables with $X_n \sim \text{Binomial}(n, p_n)$ such that $np_n \rightarrow \lambda > 0$ as $n \rightarrow \infty$. Show that as $n \rightarrow \infty$,

$$P(X_n = k) \rightarrow \frac{e^{-\lambda} \lambda^k}{k!}, \text{ for } k = 0, 1, 2, \dots$$

The dataset "Long.txt" contains data on the productivity of biochemistry PhD students. The variables are as follows

- **art** Number of articles published by the student during last three years of PhD
- **fem** Gender: 1 if female, 0 if male
- **mar** Marital status: 1 if married, 0 if not
- **kid5** Number of children five years old or younger
- **phd** Prestige rating of PhD department
- **ment** Number of articles published by mentor during last three years

Read the dataset into R and attach the data frame.

2. Since the exposure of all students is fixed at three years, we can model the students' article count directly, using a Poisson log-linear model with no offset. Investigate the bivariate relationships of $\log(\text{art})$ with the other variables. Which variables does the article count appear to depend on?
3. Fit a Poisson log-linear model regressing **art** on the linear effect of the other variables. Notice that the deviance is much greater than the degrees of freedom. Could this be due to a need for second order terms?
4. Fit a quasi-Poisson model regressing **art** on the linear effect of the other variables. Do the data appear to be overdispersed?
5. Using a Poisson or quasi-Poisson model as you see fit, select an appropriate model for **art**. Interpret your final model.

6. ISL 4.7 Exercise 5
7. ISL 4.7 Exercise 6
8. ISL 4.7 Exercise 7
9. ISL 4.7 Exercise 10
10. ISL 4.7 Exercise 11