

# MAST30027: Modern Applied Statistics

## Week 6 Lab

1. In the question 2 of the week 5 lab, we fitted a rate model (a type of Poisson regression) to data on the effect of gamma radiation on chromosomal abnormalities.

Show that these data are overdispersed compared to a Poisson distribution. Next test for an interaction between `doserate` and `doseamt`, firstly without allowing for overdispersion (fixing this dispersion  $\phi = 1$ ), and secondly allowing for overdispersion. Do you get different answers?

2. 1) Simulate 1000 samples from the following mixture model. For  $i = 1 \dots, 1000$ ,

$$Z_i \sim \text{categorical}(0.3, 0.7),$$

$$X_i|Z_i = 1 \sim \text{Pois}(\lambda = 2) \text{ and } X_i|Z_i = 2 \sim \text{Pois}(\lambda = 10).$$

Please set a seed using `'set.seed(30027)'`. Save the information about which component each observation come from. Make a histogram of the simulated samples.

2) We pretend that we just obtain the simulated samples without their true component information. Cluster the simulated samples into two clusters using the mixture of two Poisson models. We assume we have already obtained MLE of the parameters in the model. We will use true values of parameters as MLE here ( $\hat{\pi}_1 = 0.3, \hat{\lambda} = 2$  for the first component, and  $\hat{\lambda} = 10$  for the second component). If  $P(Z_i = 1|X_i) > 0.5$ , assign a sample  $X_i$  to the first cluster. How many samples are assigned to wrong clusters?