<div align="center">

STATISTICS HONOURS
THEORY OF STATISTICS
LIKELIHOOD ASSIGNMENT 2018

</div>

This assignment will count 25% towards your final mark for this section.

Show all your work. If you don't hand in an R Markdown file, attach your code as an appendix. Handin an electronic version of your project on Vula (under Assignments), and a hard copy at reception. I want to see that you understand what you are doing, this means you should give good interpretations and explanations of results. You are allowed to discuss the work amongst each other. However, your code must be your own work and the write-up must be your own work. Also, I expect you to not just copy the code that I have given to you, but be a bit creative in how you use and adapt it.

Your methods should state clearly what you have done and assumed (e.g. in all plots, calculations), in your own words, but not using R code within the report.

Models for count data

Count data are often much more variable than can be modelled by a simple Poisson distribution or Poisson regression model. This *overdispersion* can be detected in a goodness of fit test, often this is done by comparing the residual deviance to the residual degrees of freedom (chi-squared test). Overdispersion refers to the observations being much more variable than expected if they would really be assumed to follow a Poisson distribution. For Poisson distributions the variance = mean. For regression models this mean that the variance of the observations around the fitted line (residuals) should increase with fitted value (and equal the fitted value). If the variance is much larger than the fitted value, we call this overdispersion. However, usually overdispersion is caused by something. Possible reasons include 1) some important explanatory variables are missing in the regression model, 2) there are different groups in the population, or large heterogeneity, i.e. the observations at a given fitted value do not all come from the same population, 3) factors like spatial or temporal dependence between observations cause clustering.

One can of course ignore overdispersion and lack of fit of the simple Poisson model. The problem with this is that uncertainty estimates may be seriously over- or underestimated.

Alternatively, one can choose a different model that can deal with such data.

Mixture of Poissons

This assumes that the population sampled from is a mixture of groups. Each observation comes from one of these groups, which is not known. The proportion from each group is one of the model parameters. See your notes, Appendix, for the likelihood of (finite) mixture models.

The parameters of the model are the mixture proportions plus the parameters for each of the group models.

$$P(y) = \sum_{i=1}^{m} f(y|g_i)p(g_i)$$

Zero-inflated Poisson

Often there are many zeros in the data, which also contributes to overdispersion. One can model these extra zeros using a mixture of a Poisson and a zero process, i.e. there are two processes that could result in zeros. For example, the zeros could come from individuals that just behave differently and always result in a zero, whereas other individuals follow a Poisson process.

$$P(y_i = 0) = \pi + (1 - \pi)e^{-\lambda}$$

$$P(y_i = x) = (1 - \pi)\frac{\lambda^x e^{-\lambda}}{x!}, \qquad x \geq 1$$

Negative Binomial

A negative binomial model is an alternative model for count data. There are two common parameterizations of the negative binomial. The one has parameters mean m and shape r, with variance $m + m^2/r$. Note that the variance also increases with the mean, but there is an additional scaling factor to increase the variance, and this is why this form of the negative binomial is often used to model extra variation in count data.

One can also mix negative binomial distributions, or Poisson with negative binomial distributions.

Instructions

The data (`accidents.csv`) are number of accidents on two-lane (same direction) road segments in Cape Town over a five-year period. The segments differ in length between 0.2 and 7.2 km; unfortunately, the segment lengths are not provided here.

Not necessarily in this order your analysis and report should include the following:

- Exploratory data analysis.

- Find a good model for the count data. Try at least the following models:

    - Poisson

- negative binomial
  - mixture of 2 Poissons
  - zero-inflated Poisson

- Give an expression for the likelihood and the log-likelihood. Define all parameters.

- Choose the model with strongest support from the data.

- How good is this model? Can you illustrate your results?

- Find the profile likelihoods for all parameters in your selected model. Give confidence intervals for the parameters. Plot likelihood surface (two parameters at a time if necessary, fixing the other parameters at their MLEs).

- Are parameter estimates correlated?

- Write a report, with introduction, methods, results and conclusions. Put into context. <mark>What would your next steps be in trying to understand these data? How could you improve these models?</mark>

- You should program all likelihoods, profile likelihoods, confidence intervals yourselves, <mark>i.e. don't use existing R software to fit these models, although you can use these to check your programs. You can use R's optimization functions.</mark>

- A note on parameter estimation: optim and nlm in R by default assume that parameters are unbounded, i.e. can take on any value on the real number line. However, many statistical parameters are bounded, e.g. probabilities can only take on values between 0 and 1, variances are always ≥ 0.

  It is a good idea to reparameterize such parameters. Not only will the optimization not go into invalid regions, but mostly, the log-likelihoods are much better behaved for the transformed versions (not so flat, not near the boundary).

  For variance parameters the transformation (reparameterization) mostly used is

  $$\lambda = \log(\sigma)$$

  so that

  $$\sigma = \exp(\lambda)$$

  In the above transformation σ is always positive, and λ can take on any real value. The likelihood is parameterized in terms of λ, any σ is replaced with $\exp(\lambda)$.

  For mixture probabilities $\delta_i \geq 0$ and $\sum \delta_i = 1$. The following transformation can be used

  $$\tau_i = \log\left(\frac{\delta_i}{1 - \sum_{j=2}^{m} \delta_j}\right), \qquad i = 2, ..., m$$

  and back:

  $$\delta_i = \frac{\exp(\tau_i)}{1 + \sum_{j=2}^{m} \exp(\tau_j)}, \qquad i = 2, ..., m$$

  The last probability can be derived as 1 minus the rest (Zucchini & MacDonald 2009, pg. 10). <mark>Alternatively, you could use an algorithm with constrained optimization.</mark>