



Advances in Analysis of Count Data (and AGM)

Wednesday 28th November 2018

This meeting, with the theme of Advances in the Analysis Count Data, will include the Presidential Address by Martin Ridout and be preceded by the AGM.

Location: The Hardy Room, De Morgan House, 57-58 Russell Square, London WC1B 4HS

Date: 1.30-5.15pm, Wednesday 28th November 2018

Registration fees: £25 for full or retired members; £10 for student members; £50 for non-members or £90 for non-members including Biometric Society membership for 2019

Organiser: Martin Ridout (contact via [Committee page](#))

Programme

13:30 -
14:00 **Annual General Meeting of BIR**

Martin Ridout (University of Kent)

Presidential Address: Musings on matching

14:00 -
14:40 The classical matching distribution arises when a list of items is rearranged in a random order; it describes the number of 'matches', that is items that occur in the same position in the new list as in the old list. One statistical application is in testing for matching ability - for example to assess whether a graphologist can match personality sketches of people to samples of their handwriting more successfully than might be expected by chance. In some applications, it is clear that matching is better than random, and it is then of interest to formulate alternative distributions that can model this. In this talk, I describe some distributions of this type and show that there are close connections with models for ranking data and models for species sampling. I will also talk about the biometrical problem that sparked my interest in this area, and the challenges that it still presents.

Angela Noufaily (University of Warwick)

Taylor's power law and counts of infectious organisms

14:40 -
15:20 Surveillance data collected on counts of several hundred different infectious organisms over 20 years have revealed strikingly universal power relationships between their variance and mean in successive time periods. Such patterns are common in ecology, where they are referred to collectively as Taylor's power law. In this talk we will describe these relationships, which appear to match those expected of Tweedie distributions, and will discuss how they might inform the modelling of infectious disease data.

15:20 -
15:50 **Tea / Coffee Break**

Alina Peluso (Imperial College)

Discrete Weibull-based regression models for count data

15:50 - In a parametric framework, Poisson regression is the simplest model for count data though it is often found not adequate in real applications, particularly in the presence of excessive zeros and in the case of dispersion, i.e. when the conditional mean is different to the conditional variance. Negative Binomial regression is the standard model for over-dispersed data, but it fails in the presence of under-dispersion. Poisson-Inverse Gaussian regression can be used in the case of over-dispersed data, Generalised-Poisson regression can be employed in the case of under-dispersed data, and Conway-Maxwell Poisson regression can be employed in both cases of over- or under-dispersed data, though the interpretability of these models is not straightforward, and they are often found computationally demanding. While Jittering is the default non-parametric approach for count data, inference must be made for each individual quantile, separate quantiles may cross, and the underlying uniform random sampling can generate instability in the estimation. These features motivate the development of a novel parametric regression model for counts via a Discrete Weibull distribution. This distribution can adapt to different types of dispersion relative to Poisson, and it also has the advantage of having a closed form expression for the quantiles. Simulated and real data applications with different type of dispersion show a good performance of Discrete Weibull-based regression models compared with existing regression approaches for count data.

Peter Neal (University of Lancaster)

Time series models for count data: Which model to choose?

16:30 - Temporal data arise in a wide range of situations and numerous time series models have been developed to analyse such data. For low-frequency count data traditional real-valued time series models are inappropriate and hence the development of integer valued time series models. There are two broad classes of integer valued time series models, observation-driven models, such as the INAR(p) model, the discrete analogue of the AR(p) model, and parameter-driven models, such as the autoregressive Poisson regression model.

17:10 - In this talk we will introduce observation-driven and parameter-driven models for low-frequency count data. We will discuss key features of time series generated by these different models and a general approach for selecting between competing time series models for a given data set. The methods will be illustrated with example data sets from epidemiology and ecology.