Publicacions més rellevants de la línia de recerca:
Disseny d'Experiments i anàlisi estadística de dades discretes

Abstract: The inverse Gaussian-Poisson mixture model is very useful when modelling highly skewed non-negative integer data in fields as diverse as linguistics, ecology, market research, bibliometry, engineering and insurance. When using this statistical model on the frequency of word or species frequency data, one typically truncates its sample space at zero to accommodate for the ignorance about the number of words or species that are not observed. In this paper, we show that by truncating the sample space of the inverse Gaussian–Poisson model, one is allowed to extend its parameter space and in that way improve its fit when the frequency of one is larger and the right tail is heavier than is allowed by the unextended model. By fitting the extended model to word frequency count data, we find many instances where the maximum likelihood estimates fall in the extension of the parameter space.

Abstract: The information in a statistical experiment is often measured through the determinant of its information matrix. Under first order normal linear models the determinant of the information matrix of a two-level factorial experiment neither depends on where the experiment is centered nor on how it is oriented, and balanced allocations are more informative than unbalanced ones with the same number of runs. In contrast, under binary response models none of these properties hold. The performance of two-level experiments for binomial responses is explored by investigating the dependence of the determinant of their information matrix on their location, orientation, range, presence or absence of interactions and on the relative allocation of runs to support points, and in particular on the type of fractionating involved. Conventional wisdom about two-level factorial experiments, which is deeply rooted on normal response models, does not apply to binomial models. In binary response settings factorial experiments should not be used for screening or as building blocks for binary response surface exploration, and there is no alternative to the optimal

design theory approach to planning experiments.

Abstract: The analysis of word frequency count data can be very useful in authorship attribution problems. Zero-truncated generalized inverse Gaussian-Poisson mixture models are very helpful in the analysis of this kind of data because their model mixing density estimates can be used as estimates of the density of the word frequencies of the vocabulary. It is found that this model provides excellent fits for the word frequency counts of very long texts, where the truncated inverse Gaussian-Poisson special case fails because it does not allow for the large degree of over-dispersion in the data. The role played by the three parameters of this truncated GIG-Poisson model is also explored. Our second goal is to compare the fit of the truncated GIG-Poisson mixture model with the fit of the model that results from switching the order of the mixing and truncation stages. An heuristic interpretation of the mixing distribution estimates obtained under this alternative GIG-Truncated Poisson mixture model is also provided.