

## This is an Example of an Article Title

A.U. Thor\*

Institute of Data Science, CA 560034, USA

\*email: email@address.com

and

A. Other\*

Building, Institute, Street Address, City, Code, Country

\*email: email1aa@address.com

**SUMMARY:** The world will little note, nor long remember, what we say here, but can never forget what they did here. It is for us, the living, rather to be dedicated here to the unfinished work which they have, thus far, so nobly carried out. It is rather for us to be here dedicated to the great task remaining before us—that from these honored dead we take increased devotion to that cause for which they here gave the last full measures of devotion—that we here highly resolve that these dead shall not have died in vain; that this nation shall have a new birth of freedom; and that this government of the people, by the people, for the people, shall not perish from the earth. The world will little note.

**KEY WORDS:** Colostrum; Milk; Milk oligosaccharide; Non-human mammal.

### 1. Introduction

The generalized linear model (GLM) formulated by Nelder and Wedderburn (1972) provides an elegant and encompassing mathematical framework to model response variables whose distribution belongs to the exponential family (such as normal, binomial, Poisson, gamma and inverse Gaussian). A feature of exponential family distributions is the *mean-variance* relationship, i.e., the fact that the variance is a function of the mean. When analysing count data the Poisson model is a natural first choice, but the model imposes equality of mean and variance and this assumption is not always present.

Usually, count data are overdispersed, due to the deficiency of relevant covariates, or heterogeneity of samples or repeated measures, thus there are already many models proposed in the literature that incorporate this overdispersion, such as quasi-Poisson, negative binomial among others (Hinde et al., 1998; Ver Hoef and Boveng, 2007).

In continuous data, however, a very present character is the heterogeneity of variances or heteroscedastic errors. Several models were proposed to model these errors, such as, the Mixed Models, that can model besides, heteroscedasticity, correlated data (Carroll and Ruppert, 1988; Pinheiro and Bates, 2006). Breslow and Clayton extended these mixed models to the generalized linear case, and proposed the generalized linear mixed models (Breslow and Clayton, 1993).

### References

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of*

*the American statistical Association* **88**, 9–25.

Carroll, R. J. and Ruppert, D. (1988). *Transformation and weighting in regression*, volume 30. CRC Press.

Hinde, J., Demétrio, C. G., et al. (1998). Overdispersion: models and estimation. *Computational Statistics and Data Analysis* **27**, 151–170.

Nelder, J. A. and Wedderburn, R. W. (1972). Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)* **135**, 370–384.

Pinheiro, J. and Bates, D. (2006). *Mixed-effects models in S and S-PLUS*. Springer Science & Business Media.

Ver Hoef, J. M. and Boveng, P. L. (2007). Quasi-poisson vs. negative binomial regression: how should we model overdispersed count data. *Ecology* **88**, 2766–2772.

### Supporting Information

Web Appendix 1 referenced in Section ?? is available with this paper at the Biometrics website on Wiley Online Library.

Received October 2004. Revised February 2005.

Accepted March 2005.

### Appendix

Computation of  $E_i\{\alpha_i\}$

(This appendix was not part of the original paper by A.V. Raveendran and is included here just for illustrative purposes. The references are not relevant to the text of the appendix, they are references from the bibliography used to illustrate text before and after citations.)

Here is an equation; note how it is numbered:

$$A = B + C. \quad (\text{A.1})$$

Equation (A.1) is the only numbered equation in this appendix.

Spectroscopic observations of bright quasars show that the mean number density of Ly $\alpha$  forest lines, which satisfy certain criteria, evolves like  $dN/dz = A(1+z)^\gamma$ , where  $A$  and  $\gamma$  are two constants. Given the above intrinsic line distribution we examine the probability of finding large gaps in the Ly $\alpha$  forests. We concentrate here only on the statistics and neglect all observational complications such as the line blending effect (see ?, for example). We concentrate here only on the statistics and neglect all observational complications such as the line blending effect (see ?, for example). We concentrate here only on the statistics and neglect all observational complications such as the line blending effect (see ?, for example).

The references are not relevant to the text of the appendix, they are references from the bibliography used to illustrate text before and after citations.)

Spectroscopic observations of bright quasars show that the mean number density of Ly $\alpha$  forest lines, which satisfy certain criteria, evolves like  $dN/dz = A(1+z)^\gamma$ , where  $A$  and  $\gamma$  are two constants. Given the above intrinsic line distribution we examine the probability of finding large gaps in the Ly $\alpha$  forests. We concentrate here only on the statistics and neglect all observational complications such as the line blending effect (see ?, for example). We concentrate here only on the statistics and neglect all observational complications such as the line blending effect (see ?, for example). We concentrate

here only on the statistics and neglect all observational complications such as the line blending effect (see ?, for example).

Suppose we have observed a Ly $\alpha$  forest between redshifts  $z_1$  and  $z_2$  and found  $N - 1$  lines. For high-redshift quasars  $z_2$  is usually the emission redshift  $z_{\text{em}}$  and  $z_1$  is set to  $(\lambda_{\text{Ly}\beta}/\lambda_{\text{Ly}\alpha})(1+z_{\text{em}}) = 0.844(1+z_{\text{em}})$  to avoid contamination by Ly $\beta$  lines. We want to know whether the largest gaps observed in the forest are significantly inconsistent with the above line distribution. To do this we introduce a new variable  $x$ :

$$x = \frac{(1+z)^\gamma - (1+z_1)^\gamma}{(1+z_2)^\gamma - (1+z_1)^\gamma}.$$

$x$  varies from 0 to 1. We then have  $dN/dx = \lambda$ , where  $\lambda$  is the mean number of lines between  $z_1$  and  $z_2$  and is given by

$$\lambda \equiv \frac{A[(1+z_2)^\gamma - (1+z_1)^\gamma]}{\gamma + 1}.$$

This means that the Ly $\alpha$  forest lines are uniformly distributed in  $x$ .