Two Distribution Families for the Compensation of Over- and Underdispersed Frequencies Compared with the Binomial Distribution

Verena Feirer, Ulrich Hirn, Herwig Friedl, Wolfgang Bauer

Institute for Paper, Pulp and Fibre Techonolgy & Institute for Statistics
Graz University of Technology

The motivation for this work originates in an application of generalized linear models to a problem from the paper and printing industries. The aim is to unravel dependencies between the successful transfer of the printing ink and certain properties of the paper (e.g. topography) in order to gain a better knowledge of the underlying processes. Various samples from different sack papers were considered and surveyed with logistic regression models. Of the resulting models, some display overdispersion whereas others feature underdispersion.

Which is why this work tries to employ two distributions that can account for the additional or lesser variability of the data compared to the classic binomial distribution. Those two distributions are the multiplicative binomial distribution as first introduced by Altham (1978) and the double binomial distribution as introduced by Efron (1986). Both distributions originate from the classic binomial distribution but carry an additional parameter that enables the distribution to react to variabilities smaller or larger than in the binomial case.

An overview of these distributions will be given, followed by an embedding into the framework of generalized linear models. Simulation results will show the performance of the parameter estimation via maximum likelihood estimation with special focus on success rates of almost 1 as the rate for successful ink transfer can be found in this section. Finally, the presentation will conclude with a comparison of the three models resulting from the different distributions (classic, multiplicative and double binomial), applied to the dataset of successful or failed ink transfer.

References

Altham, P. (1978). Two generalizations of the binomial distribution. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *27*(2), 162-197.

Efron, B. (1986). Double exponential families and their use in generalized linear regression. *Journal of the American Statistical Association*, *81*(395), 709-721.