
Fisher dispersion index for multivariate count distributions: A review and a new proposal

Célestin C. Kokonendji, Pedro Puig

PII: S0047-259X(17)30383-4
DOI: https://doi.org/10.1016/j.jmva.2017.12.010
Reference: YJMVA 4319

To appear in: *Journal of Multivariate Analysis*

Received date : 22 June 2017

# Fisher dispersion index for multivariate count distributions: A review and a new proposal

Célestin C. Kokonendji[a,*], Pedro Puig[b]

[a]Laboratoire de mathématiques de Besançon, Université Bourgogne Franche-Comté, Besançon, France
[b]Departament de Matemàtiques, Universitat Autònoma de Barcelona, Barcelona, Spain

**Abstract**

The Fisher dispersion index is very widely used to measure the departure of any univariate count distribution from the equidispersed Poisson model. A multivariate extension has not yet been well defined and discussed in the literature. In this paper, a new definition of the multivariate Fisher index through the generalized dispersion index is proposed. This is a scalar quantity, defined as a ratio of two quadratic forms of the mean vector and the covariance matrix. A multiple marginal dispersion index and its relative extension for a given reference count distribution are discussed, and the asymptotic behavior and other properties are studied. Illustrative examples and practical applications on count datasets are analyzed under several scenarios. Some concluding remarks are made, including challenging problems.

*Keywords:* Dependence, equidispersion, multivariate Poisson distribution, overdispersion, relative index, scaled generalized variance, underdispersion
*2010 MSC:* 62E10, 62F10, 62H05, 62H12, 62H99, 62-07

## 1. Introduction

The univariate concepts of over- and under-dispersion with respect to the Poisson distribution are widely used in practice; see, e.g., Kokonendji [27] for a review. A simple measure, the ratio of variance to mean, is the Fisher [16] dispersion index which makes it possible to classify distributions and make inference; see, e.g., Mizère et al. [38] for statistical tests and also Weiß et al. [51] in count time series. There are many interpretable mechanisms leading to univariate over- and also under-dispersed distributions. But what is the meaning of over-, equi- and under-dispersion in a multivariate framework? Such an extension is not an easy task.

Consider for instance the bivariate data set described in [14] counting the number of live-born ($y_1$) and dead-born ($y_2$) piglets along births. Data came from a farm in Lleida (Spain) at 9-year follow-up, and the objective of the analysis was the early detection of Porcine Reproductive and Respiratory Syndrome. This is a viral disease with negative impact on sow reproduction, being associated to a decrease in the number of live-born piglets ($y_1$) and an increase in the number of dead-born piglets ($y_2$). The number of analyzed births was $n = 889$ (sample size), coming from female pigs with a parity from 3 to 5, recorded during outside outbreak periods. The empirical dispersion indexes of live-born and dead-born piglets were 0.94 and 1.95, respectively, with a correlation coefficient equal to $-0.18$. Descriptive statistics are shown in Table 1 (Dataset No 10). In this example one marginal is underdispersed, the other is overdispersed and the correlation is negative. This is not a usual behavior for bivariate count distributions. In general, comparison of only the marginal dispersion indexes ignores the crucial dependency information. Therefore, the first question that arises is how to appropriately define a bi/multivariate dispersion index by giving a reference and taking into account the correlation structure. A multivariate dispersion index could be a powerful descriptive tool for outbreak detection.

*Corresponding author at: Laboratoire de mathématiques de Besançon - UMR 6623 CNRS-UFC, UFR Sciences et Techniques - UFC, 16, route de Gray, 25030 Besançon Cedex, France; Tel. +33.3.81.66.63.41, Fax +33.3.81.66.66.23
*Email addresses:* celestin.kokonendji@univ-fcomte.fr (Célestin C. Kokonendji), ppuig@mat.uab.cat (Pedro Puig)