DSSV 2023 Data Science, Statistics & Visualisation
ECDA 2023 European Conference on Data Analysis

Improving the Ternary Diagrams for Compositions

Paul H.C. Eilers[1]

[1]Erasmus University Medical Center, Rotterdam, The Netherlands

The ternary diagram is a familiar display for compositions with three components. It uses the fact that the sum of the distances to the sides of an equilateral triangle is constant for all points inside the triangle. It is a convenient device, as long as none of the three fractions are close to zero or one. If that is not the case, data points are pushed into a corner or to one of the sides of the triangle.

De Rooij and Eilers [1] proposed an alternative display, which they called TrioScale. In the ternary diagram, the projection on each axis is determined by the arithmetic difference of one pair of of the three components. The TrioScale display is based on the differences of their logarithm, i.e. the log-ratio.

Log-ratios do not accept zeros. If a data set contains low counts, with one or more zeros, the TrioScale display cannot be used directly. As a remedy, I propose to compute a penalized log-linear fit to the data and to replace counts by their expected values. This idea is inspired by the PRIDE model of Perperoglou and Eilers [2]. Let $\mu_{ij} = E(y_{ij})$. We fit the saturated log-linear model $\eta_{ij} = \log(\mu_{ij}) = \alpha_i + \beta_j + \gamma_{ij}$, with a ridge penalty $\lambda \sum_i \sum_j \gamma_{ij}^2$ on γ. The penalty parameter λ is tuned using AIC.

The figure below shows an application, using the data set *Pigs* in the R package *zCompositions*. It contains counts of the movements of pigs in three sections of a stable. The 12 dots in the left panel represent 29 observed compositions. There are many zeros, forcing the dots to the axes, and many dots overlap exactly. They become the crosses in the right panel, in which more details of their structure can be seen.

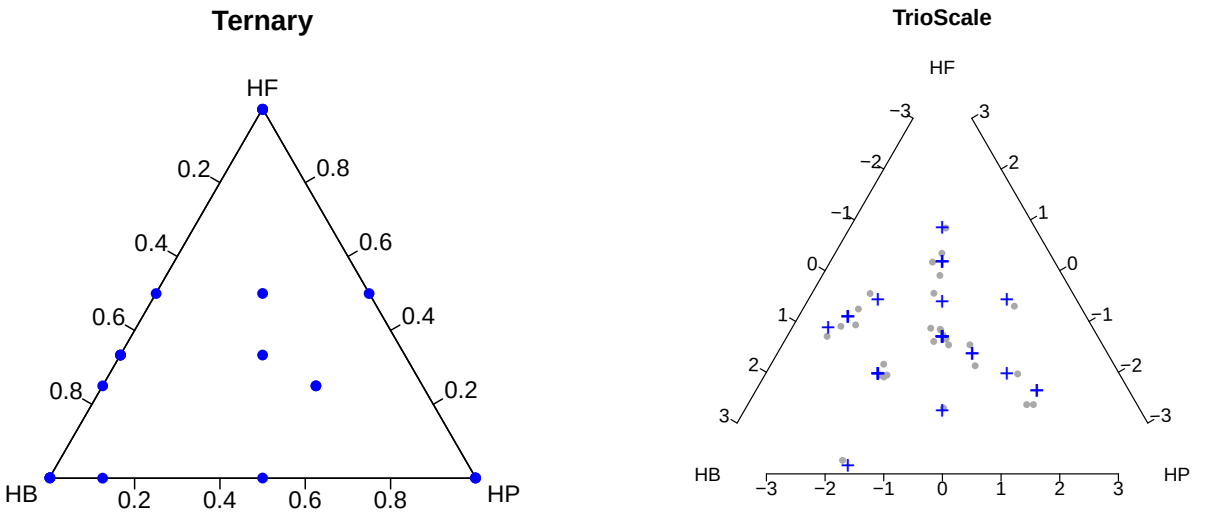

Figure 1: The pigs data in ternary and zero-corrected TrioScale plots. Many points overlap. In the right panel the gray dots are jittered to give an impression of the numbers of overlapping points. The labels along the scales refer to the decimal logarithms of ratios.

[1] M. de Rooij and P. Eilers, "Trioscale: A new diagram for compositional data.," in *Proceedings of CODAWork 2013, Vorau.*, 2013.
[2] A. Perperoglou and P. H. C. Eilers, "Penalized regression with individual deviance effects," *Computational Statistics*, vol. 25, pp. 341–361, 2010.