Accepted Manuscript

Simplicial variances, potentials and Mahalanobis distances

Luc Pronzato, Henry P. Wynn, Anatoly A. Zhigljavsky

PII: S0047-259X(17)30531-6
DOI: https://doi.org/10.1016/j.jmva.2018.08.002
Reference: YJMVA 4397

To appear in: *Journal of Multivariate Analysis*

Received date : 5 September 2017

Simplicial variances, potentials and Mahalanobis distances

Luc Pronzato[a,*], Henry P. Wynn[b], Anatoly A. Zhigljavsky[c]

[a] *Université Côte d'Azur, CNRS, I3S, Sophia Antipolis, France*
[b] *London School of Economics, London, UK*
[c] *School of Mathematics, Cardiff University, Cardiff, UK*

Abstract

The average squared volume of simplices formed by k independent copies from the same probability measure μ on $\mathbb{R}^d$ defines an integral measure of dispersion $\psi_k(\mu)$, which is a concave functional of μ after suitable normalization. When $k = 1$ it corresponds to $\text{tr}(\Sigma_\mu)$ and when $k = d$ we obtain the usual generalized variance $\det(\Sigma_\mu)$, with Σ_μ the covariance matrix of μ. The dispersion $\psi_k(\mu)$ generates a notion of simplicial potential at any $x \in \mathbb{R}^d$, dependent on μ. We show that this simplicial potential is a quadratic convex function of x, with minimum value at the mean a_μ for μ, and that the potential at a_μ defines a central measure of scatter similar to $\psi_k(\mu)$, thereby generalizing results by Wilks (1960) and van der Vaart (1965) for the generalized variance. Simplicial potentials define generalized Mahalanobis distances, expressed as weighted sums of such distances in every k-margin, and we show that the matrix involved in the generalized distance is a particular generalized inverse of Σ_μ, constructed from its characteristic polynomial, when $k = \text{rank}(\Sigma_\mu)$. Finally, we show how simplicial potentials can be used to define simplicial distances between two distributions, depending on their means and covariances, with interesting features when the distributions are close to singularity.

Keywords: Bregman divergence, Characteristic polynomial, Dispersion, Generalized variance, Mahalanobis distance, Potential, Scatter
MSC: 94A17, 62B10, 62K05

1. Introduction

A rather common problem in multivariate statistical data analysis involves measuring the scatter of a dataset. Classical approaches rely on the empirical covariance matrix or a robust version of it. Most frequently, this matrix is close to being degenerate, with several small eigenvalues. In such situations, many standard methods, including analysis via the generalized variance, may not be applicable. Hence, the need of methods that concentrate their attention on subspaces of appropriate dimensions. In [17], we introduced a class of extended generalized k-variances for a probability measure μ on $\mathbb{R}^d$ with covariance matrix $\Sigma = \Sigma_\mu$. These measures of dispersion are indexed by an integer parameter $k \in \{1, \ldots, d\}$. When $k = 1$ the generalized k-variance becomes $\text{tr}(\Sigma)$ and when $k = d$ we obtain the usual generalized variance $\det(\Sigma)$. For general $k \in \{1, \ldots, d\}$, the k-variance is the sum of the determinants of all the $k \times k$ principal minors of Σ, i.e., the sum of generalized variances for all k-dimensional minors.

The simplicial nature of the results stems from a theorem which, up to a circumstantial multiplier, equates the extended generalized variance to the expected squared volume of simplices formed from independent copies of the random vector associated with μ; for the value k we take $k + 1$ copies.

A main idea of this paper is that an integral measure of dispersion generates a notion of potential at a general point x and dependent on μ. A main result relates the notion of simplicial potential obtained here to a generalized Mahalanobis distance, expressed as a weighted sum of such distances in every k-margin. We show also that the potential arises

*Corresponding author
URL: `Luc.Pronzato@cnrs.fr` (Anatoly A. Zhigljavsky), `H.Wynn@lse.ac.uk` (Anatoly A. Zhigljavsky), `ZhigljavskyAA@cf.ac.uk` (Anatoly A. Zhigljavsky)