Journal of Multivariate Analysis 144 (2016) 200–217

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

New algorithms for M-estimation of multivariate scatter and location

Lutz Dümbgen [a,*], Klaus Nordhausen [b], Heike Schuhmacher [a]

[a] University of Bern, Switzerland

[b] University of Turku, Finland

ARTICLE INFO

Article history:
Received 21 May 2015
Available online 27 November 2015

AMS subject classifications:
62H12
65C60

Keywords:
Fixed-point algorithm
Matrix exponential function
Newton–Raphson algorithm
Taylor expansion

ABSTRACT

We present new algorithms for M-estimators of multivariate scatter and location and for symmetrized M-estimators of multivariate scatter. The new algorithms are considerably faster than currently used fixed-point and other algorithms. The main idea is to utilize a Taylor expansion of second order of the target functional and devise a partial Newton–Raphson procedure. In connection with symmetrized M-estimators we work with incomplete U-statistics to accelerate our procedures initially.

1. Introduction

Robust estimation of multivariate location and scatter for a distribution P on $\mathbb{R}^q$ is a recurring topic in statistics. For instance, different estimators of multivariate scatter are an important ingredient for independent component analysis (ICA) or invariant coordinate selection (ICS), see Nordhausen et al. [10] and Tyler et al. [18] and the references therein. Of particular interest are M-estimators and their symmetrized versions as defined in Sections 2.1 and 2.3, respectively, because they offer a good compromise between robustness and computational feasibility. The most popular algorithm to compute M-estimators of multivariate scatter is to iterate a fixed-point equation, see Huber [7, Section 8.11], Tyler [17] and Kent and Tyler [8]. This algorithm has nice properties such as guaranteed convergence for any starting point. However, as discussed later, it can be rather slow for high dimensions and large data sets. We introduce two alternative methods, a gradient descent method with approximately optimal stepsize and a partial Newton–Raphson method, which turn out to be substantially faster.

Computation time becomes a major issue in connection with symmetrized M-estimators. These estimators are important because of a desirable "block independence property" as explained in Section 2.3; see also Dümbgen [3] and Sirkiä et al. [16]. If applied to a sample of n observations $X_1, \ldots, X_n \in \mathbb{R}^q$, symmetrized M-estimators utilize the empirical distribution of all $\binom{n}{2}$ differences $X_i - X_j$, $1 \le i < j \le n$.

In Section 2 we describe briefly the various M-estimators we are interested in. Then we introduce a general target functional on the space of symmetric and positive definite matrices in $\mathbb{R}^{q\times q}$ which has to be minimized. Section 3 presents some analytical properties of the latter functional which are essential to understand existing algorithms and to devise new ones. These parts follow closely a recent survey of multivariate M-functionals by Dümbgen et al. [5]. In Section 4

* Corresponding author.
E-mail address: duembgen@stat.unibe.ch (L. Dümbgen).

http://dx.doi.org/10.1016/j.jmva.2015.11.009
0047-259X/

we discuss the aforementioned fixed-point algorithm of Kent and Tyler [8] and explain rigorously why it is suboptimal. Then we introduce two alternative methods, a gradient descent method with approximately optimal stepsize and a partial Newton–Raphson method. Numerical experiments in Section 5 show that the new algorithms are substantially faster than the fixed-point algorithms or the algorithms by Arslan et al. [1]. Proofs are deferred to Section 6.

Some notation. The space of symmetric matrices in $\mathbb{R}^{q\times q}$ is denoted by $\mathbb{R}^{q\times q}_{\mathrm{sym}}$, and $\mathbb{R}^{q\times q}_{\mathrm{sym},>0}$ stands for its subset of positive definite matrices. The identity matrix in $\mathbb{R}^{q\times q}$ is written as I_q. The Euclidean norm of a vector $v \in \mathbb{R}^q$ is denoted by $\|v\| = \sqrt{v^\top v}$. For matrices M, N with identical dimensions we write

$$\langle M, N\rangle := \mathrm{tr}(M^\top N) \quad \text{and} \quad \|M\| := \sqrt{\langle M, M\rangle},$$

so $\|M\|$ is the Frobenius norm of M.

2. The *M*-estimators and the target functional

Let $X_1, \ldots, X_n$ be independent random vectors with unknown distribution P on $\mathbb{R}^q$. Our task is to define and then estimate a certain center $\boldsymbol{\mu}(P) \in \mathbb{R}^q$ and scatter matrix $\boldsymbol{\Sigma}(P) \in \mathbb{R}^{q\times q}_{\mathrm{sym},>0}$.

2.1. The scatter-only problem

Let us start with the assumption that $\boldsymbol{\mu}(P) = 0$. To define and estimate a scatter functional $\boldsymbol{\Sigma}(P)$ we consider a simple working model consisting of elliptically symmetric probability densities f_Σ on $\mathbb{R}^q$ depending on a parameter $\Sigma \in \mathbb{R}^{q\times q}_{\mathrm{sym},>0}$:

$$f_\Sigma(x) = C^{-1}(\det \Sigma)^{-1/2} \exp\{-\rho(x^\top \Sigma^{-1} x)/2\},$$

where $\rho : [0,\infty) \to \mathbb{R}$ is a given function such that $C := \int \exp\{-\rho(\|x\|^2)/2\}\, dx$ is finite. Assuming temporarily that this working model is correct, one could estimate the true underlying matrix parameter by a maximizer of the corresponding log-likelihood function for this model,

$$\Sigma \mapsto -n \ln C - \frac{1}{2}\sum_{i=1}^n \rho(X_i^\top \Sigma^{-1} X_i) - \frac{n}{2}\ln\det \Sigma.$$

With the empirical distribution $\widehat{P} = n^{-1}\sum_{i=1}^n \delta_{X_i}$ of the data $X_1, \ldots, X_n$, the log-likelihood at Σ may be written as $n\int \ln f_\Sigma\, d\widehat{P}$. Thus maximization of the log-likelihood function over $\mathbb{R}^{q\times q}_{\mathrm{sym},>0}$ is equivalent to minimization of $\Sigma \mapsto L(\Sigma, \widehat{P})$, where

$$\begin{aligned} L(\Sigma, Q) &:= 2\int \ln(f_{I_q}/f_\Sigma)\, dQ \\ &= \int \{\rho(x^\top \Sigma^{-1} x) - \rho(x^\top x)\}\, Q(dx) + \ln\det \Sigma \end{aligned}$$

for a generic distribution Q on $\mathbb{R}^q$. We include f_{I_q} and $\rho(x^\top x)$, respectively, because often this increases the range of distributions Q such that $L(\Sigma, Q)$ is well-defined in $\mathbb{R}$. If $L(\cdot, Q)$ has a unique maximizer over $\mathbb{R}^{q\times q}_{\mathrm{sym},>0}$, we denote it with $\boldsymbol{\Sigma}(Q)$. The resulting mapping $Q \mapsto \boldsymbol{\Sigma}(Q)$ is called an M-functional of scatter. In particular, $\boldsymbol{\Sigma}(\widehat{P})$ serves as an estimator of the scatter parameter $\boldsymbol{\Sigma}(P)$, assuming that both exist. If P happens to have a density f_{Σ_*} in our working model, then $\boldsymbol{\Sigma}(P) = \Sigma_*$. If P is merely elliptically symmetric with center 0 and scatter matrix Σ_*, for instance, if it has a density f of the form

$$f(x) = (\det \Sigma_*)^{-1/2} g_*(x^\top \Sigma_*^{-1} x)$$

with $g_* : [0,\infty) \to [0,\infty)$, then at least $\boldsymbol{\Sigma}(P) = \gamma\Sigma_*$ for some $\gamma > 0$.

An important example is multivariate t distributions with $\nu > 0$ degrees of freedom. Here $\rho = \rho_{\nu,q}$ with

$$\rho_{\nu,q}(s) = (\nu + q)\ln(\nu + s) \quad \text{for } s \geq 0. \tag{1}$$

Note that $\rho(x^\top \Sigma^{-1} x) - \rho(x^\top x)$ equals $(q+\nu)\ln\{(\nu + x^\top \Sigma^{-1} x)/(\nu + x^\top x)\}$, a bounded and smooth function of $x \in \mathbb{R}^q$.

2.2. The location–scatter problem

Now our working model consists of probability densities $f_{\mu,\Sigma}$ on $\mathbb{R}^q$ with parameters $\mu \in \mathbb{R}^q$ and $\Sigma \in \mathbb{R}^{q\times q}_{\mathrm{sym},>0}$, namely,

$$f_{\mu,\Sigma}(x) = C^{-1}(\det \Sigma)^{-1/2} \exp\big[-\rho\{(x-\mu)^\top \Sigma^{-1}(x-\mu)\}/2\big].$$