Linear Algebra and its Applications 473 (2015) 261–283

Contents lists available at ScienceDirect

Linear Algebra and its Applications

www.elsevier.com/locate/laa

Linear estimating equations for exponential families with application to Gaussian linear concentration models

Peter G.M. Forbes, Steffen Lauritzen *

Department of Statistics, University of Oxford, 1 South Parks Road, Oxford OX1 3TG, United Kingdom

ARTICLE INFO

Article history:
Received 4 November 2013
Accepted 21 August 2014
Available online 2 September 2014
Submitted by M. Drton

MSC:
62H12
62F10

Keywords:
Gaussian graphical models
Jordan algebra models
Score matching
Scoring rules
Symmetry

ABSTRACT

In many families of distributions, maximum likelihood estimation is intractable because the normalization constant for the density which enters into the likelihood function is not easily available. The score matching estimator [35] provides an alternative where this normalization constant is not required. For an exponential family, e.g. a Gaussian linear concentration model, the corresponding estimating equations become linear [2,36] and the score matching estimator is shown to be consistent and asymptotically normally distributed as the number of observations increase to infinity, although not necessarily efficient. For linear concentration models that are also linear in the covariance [37] we show that the score matching estimator is identical to the maximum likelihood estimator, hence in such cases it is also efficient. Gaussian graphical models and graphical models with symmetries [32] form particularly interesting subclasses of linear concentration models and we investigate the potential use of the score matching estimator for this case.

* Corresponding author.
E-mail address: steffen@stats.ox.ac.uk (S. Lauritzen).

http://dx.doi.org/10.1016/j.laa.2014.08.015

1. Introduction

The increasing interest in analysis of high-dimensional data has necessitated the development of parsimonious multivariate models with reliable, computationally efficient estimation procedures. For example, sparse Gaussian graphical models [19,39,44,11,13] have drawn significant interest and several computationally efficient estimation procedures have been developed [6,5,22]. Gaussian graphical models with symmetry [32] form another example, though no efficient estimation procedures have yet been developed. Here we describe and exploit a method which provides linear estimating equations when applied to any exponential family, in particular to any Gaussian graphical model with symmetry. In contrast to the maximum likelihood estimator, which often requires iterative methods, this *score matching estimator* is computationally efficient for such families and therefore has potential for initial model screening. Even when the maximum likelihood estimator is desired, it must be computed iteratively and the score matching estimator may provide a useful initial value for the iterations.

2. Preliminaries

Consider a random quantity taking values in an open subset $\mathcal{X}$ of $\mathbb{R}^p$ which we consider equipped with the standard inner product $\langle\cdot,\cdot\rangle_p$, the associated norm $\|\cdot\|_p$, and the canonical basis.

Throughout the paper, $\mathcal{P}$ denotes a class of distributions over $\mathcal{X}$ with twice continuously differentiable densities with respect to the Lebesgue measure on $\mathcal{X}$. The general developments below are equally valid for $\mathcal{X}$ being a Riemannian manifold with associated geometric measure [17], but as our main focus is the multivariate Gaussian distribution we shall refrain from working at this level of generality. We use ∇ for the gradient and Δ for the Laplace operator on $\mathcal{X}$ so that

$$\nabla f(x) = \left\{\frac{\partial}{\partial x_i} f(x)\right\} \in \mathbb{R}^p, \qquad \Delta f(x) = \sum_{i=1}^{p} \frac{\partial^2}{\partial x_i^2} f(x).$$

2.1. Scoring rules

A *scoring rule* $S(x, Q)$ is a real-valued function which quantifies the accuracy of a predictive distribution $Q \in \mathcal{P}$ upon observing the realized value $x \in \mathcal{X}$. It is (strictly) *proper* if the expected value $E_{X\sim P}S(X, Q)$ is (uniquely) minimized over $\mathcal{P}$ at $Q = P$. Two scoring rules are *equivalent* if they differ by a positive scalar multiple and a function of x alone.

Every proper scoring rule induces a *divergence* [15,28]:

$$d(P, Q) = E_{X\sim P}\bigl\{S(X, Q) - S(X, P)\bigr\}.$$