Journal of Multivariate Analysis 135 (2015) 71–88

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Transformation-based nonparametric estimation of multivariate densities

Meng-Shiuh Chang [a], Ximing Wu [b,c,∗]

[a] School of Public Finance and Taxation, Southwest University of Finance and Economics, Sichuan 611130, China
[b] Department of Agricultural Economics, Texas A&M University, College Station, TX 77843, USA
[c] School of Economics, Xiamen University, China

HIGHLIGHTS

- We propose a transformation based nonparametric multivariate density estimator.
- We establish convergence in terms of the Kullback–Leibler Information Criterion.
- We propose a supervised hierarchical procedure of model specification.
- Our simulations demonstrate good performance of the proposed estimator.
- Our analysis reveals interesting features of global financial markets.

ARTICLE INFO

Article history:
Received 10 May 2014
Available online 10 December 2014

AMS subject classifications:
62G07
62G20
62H12

Keywords:
Multivariate density estimation
Nonparametric estimation
Copula
Kullback–Leibler information criterion

ABSTRACT

We present a probability-integral-transformation-based estimator of multivariate densities. Given a sample of random vectors, we first transform the data into their corresponding marginal distributions. The marginal densities and the joint density of the transformed data are estimated nonparametrically. The joint density of the original data is constructed as the product of the density of the transformed data and marginal densities, which coincides with the copula representation of multivariate densities. We show that the Kullback–Leibler Information Criterion (KLIC) between the true density and its estimate can be decomposed into the KLIC of the marginal densities and that of the copula density. We derive the convergence rate of the proposed estimator in terms of the KLIC and propose a supervised hierarchical procedure of model selection. Monte Carlo simulations demonstrate the good performance of the estimator. An empirical example on the US and UK stock markets is presented. The estimated conditional copula density provides useful insight into the joint movements of the US and UK markets under extreme Asian markets.

1. Introduction

Estimating probability distributions is one of the most fundamental tasks in statistics. With the advance of modern computer technology, multidimensional analysis has played an increasingly important role in many fields of science. For instance, the recent financial crisis has called for a more comprehensive approach of risk assessment of the financial markets, in which multivariate analysis of the markets is of fundamental importance. Estimating the joint distribution of stock returns

∗ Corresponding author at: Department of Agricultural Economics, Texas A&M University, College Station, TX 77843, USA.
E-mail addresses: mchang@swufe.edu.cn (M.-S. Chang), xwu@tamu.edu (X. Wu).

http://dx.doi.org/10.1016/j.jmva.2014.11.010
0047-259X/

is of independent interest by itself, and constitutes a useful exploratory step that may provide guidance for subsequent analysis.

This paper concerns the estimation of multivariate density functions. Density functions can be estimated by either parametric or nonparametric methods. Parametric estimators are asymptotically efficient if they are correctly specified, but are inconsistent under erroneous distributional assumptions. In contrast, nonparametric estimators are consistent, although they converge at a slower rate. In nonparametric estimations, the number of (nuisance) parameters generally increases with sample size. This so called *curse of dimensionality* is particularly severe for multivariate analysis, where the number of parameters increases exponentially with both the sample size and the dimension of random vectors.

In this paper, we propose a method that uses the probability-integral-transformation of the marginals to facilitate the estimation of multivariate densities. The same transformation is used by Ruppert and Cline [10] in univariate density estimations. Let $\{\boldsymbol{X}_t\}_{t=1}^n$, where $\boldsymbol{X}_t = (X_{1t}, \ldots, X_{dt})$, be an iid random sample from a d-dimensional distribution F with density f. We first transform X_{jt} to $\hat{F}_j(X_{jt})$ for $j = 1, \ldots, d$ and $t = 1, \ldots, n$, where $\hat{F}_j$ is an estimate of the jth marginal distribution. Let $\hat{c}$ be an estimate of the joint density of the transformed data. We then calculate the joint density of the original data by

$$\hat{f}(\boldsymbol{x}) = \hat{c}\left(\hat{F}_1(x_1), \ldots, \hat{F}_d(x_d)\right) \prod_{j=1}^{d} \hat{f}_j(x_j), \tag{1}$$

where $\hat{f}_j$'s, $j = 1, \ldots, d$, are the estimated marginal densities of the original data.

Eq. (1) indicates that the joint density can be constructed as the product of marginal densities and the density of the transformed data. Interestingly this construction coincides with the copula representation of multivariate density according to the celebrated Sklar's Theorem (1959), in which the first factor of (1) is termed the copula density function. This representation allows one to assemble an estimator of a joint density by estimating the marginal densities and copula density separately. A valuable by-product of this approach is the copula density, which completely summarizes the dependence structure among variables.

Our estimator consists of two steps. We first estimate the marginal densities nonparametrically. In the second step we estimate the joint density of the transformed data, which is equivalent to the copula density. We propose a nonparametric estimator of the empirical copula density using the Exponential Series Estimator (ESE) of Wu [15]. The ESE is particularly suitable for copula density estimations since it is defined explicitly on a bounded support and mitigates the boundary biases of the usual kernel density estimators, which are particularly severe for copula densities that peak toward the boundaries. This estimator has an appealing information-theoretic interpretation and lends itself to asymptotic analysis in terms of the Kullback–Leibler Information Criterion (KLIC).

We present a decomposition of the KLIC between two multivariate densities into the KLIC between the marginal densities and that between their respective copula densities, plus a remainder term. This result provides a convenient framework for the asymptotic analysis of the proposed estimator. We show that the KLIC convergence rate of the proposed estimator is the sum of the KLIC of the marginal density estimates and that of the copula density estimate, and the remainder term is asymptotically negligible.

As is common in series estimations, the number of basis functions in the ESE increases with the dimension of $\boldsymbol{x}$ rapidly. To facilitate the selection of basis functions, we propose a supervised hierarchical approach of basis function selection, which is an incremental model selection process coupled with a preliminary subset selection procedure at each step. We then use some information criterion such as the Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC) to select a preferred model.

We conduct two sets of Monte Carlo simulations. The first experiment compares the proposed estimator with a multivariate kernel density estimator in terms of overall performance. The second experiment examines the estimation of joint tail probabilities using the proposed method, the kernel estimator and the empirical distribution. In both experiments, the proposed method outperforms its competing methods, oftentimes by substantial margins.

Lastly we apply the proposed method to estimating the joint distribution of the US and UK stock markets under a variety of Asian market conditions. Our analysis reveals how fluctuations and extreme movements of the Asian market influenced the western markets in some asymmetric manner. We note that the asymmetric relation, albeit obscure in the joint densities of the US and UK markets, is quite evident in the estimated copula densities.

This paper proceeds as follows. Section 2 proposes a two-stage transformation-based estimator of multivariate densities. Section 3 presents its large sample properties and Section 4 discusses a method of model specification. Sections 5 and 6 present results of Monte Carlo simulations and an empirical application to global financial markets. Section 7 concludes. Proofs of theorems are gathered in the Appendix.

2. Estimator

Let $\{\boldsymbol{X}_t\}_{t=1}^n$ be a d-dimensional iid random sample from an unknown distribution F with density f, $d \geq 2$. We are interested in estimating f. There exist two general approaches: parametric and nonparametric. The parametric approach entails functional form assumptions up to a finite set of unknown parameters. The multivariate normal distribution is commonly used due to its simplicity. The nonparametric approach provides a flexible alternative that seeks a functional approximation to the unknown density, which is guided by data-driven principles.