Journal of Multivariate Analysis 141 (2015) 197–216

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Risk aggregation with empirical margins: Latin hypercubes, empirical copulas, and convergence of sum distributions

Georg Mainik

RiskLab, Department of Mathematics, ETH Zurich, Raemistrasse 101, 8092 Zurich, Switzerland

HIGHLIGHTS

- Asynchronously sampled data can be endowed with any copula by a reordering technique.
- Popular since 1982, this method gets a rigorous convergence proof in the present paper.
- Related estimates of sum distribution functions converge uniformly with rate $O_P(1/\sqrt{n})$.
- The underlying problem is not covered by classic empirical process results.
- CLT fails in this case. This issue affects many real-world applications.

ARTICLE INFO

Article history:
Received 4 March 2012
Available online 28 July 2015

AMS subject classifications:
62G09
62G20
62G30
62P05
62H99

Keywords:
Risk aggregation
Sum distribution
Empirical margins
Empirical copula
Functional CLT
Iman–Conover
Latin hypercube sampling

ABSTRACT

This paper studies convergence properties of multivariate distributions constructed by endowing empirical margins with a copula. This setting includes Latin Hypercube Sampling with dependence, also known as the Iman–Conover method. The primary question addressed here is the convergence of the component sum, which is relevant to risk aggregation in insurance and finance. This paper shows that a CLT for the aggregated risk distribution is not available, so that the underlying mathematical problem goes beyond classic functional CLTs for empirical copulas. This issue is relevant to Monte-Carlo based risk aggregation in all multivariate models generated by plugging empirical margins into a copula. Instead of a functional CLT, this paper establishes strong uniform consistency of the estimated sum distribution function and provides a sufficient criterion for the convergence rate $O(n^{-1/2})$ in probability. These convergence results hold for all copulas with bounded densities. Examples with unbounded densities include bivariate Clayton and Gauss copulas. The convergence results are not specific to the component sum and hold also for any other componentwise non-decreasing aggregation function. On the other hand, convergence of estimates for the joint distribution is much easier to prove, including CLTs. Beyond Iman–Conover estimates, the results of this paper apply to multivariate distributions obtained by plugging empirical margins into an exact copula or by plugging exact margins into an empirical copula.

1. Introduction

In various real-world applications, multivariate stochastic models are constructed upon empirical marginal data and an assumption on the dependence structure between the margins. This dependence assumption is often formulated in terms of copulas. The major reason for this set-up is the lack of multivariate data sets, as it is often the case in insurance and finance.

E-mail address: research@georgmainik.com.

http://dx.doi.org/10.1016/j.jmva.2015.07.008

This approach may appear artificial from the statistical point of view, but it arises naturally in the context of stress testing. In addition to finance and insurance, relevant application areas include engineering and environmental studies. Sometimes the marginal data is not even based on observations, but is generated by a univariate model that is considered reliable. Many of these models are so complex that the resulting distributions cannot be expressed analytically. In such cases exact marginal distributions are replaced by empirical distributions of simulated univariate samples. These empirical margins are endowed with some dependence structure to obtain a multivariate distribution. The computation of aggregated risk or other characteristics of this multivariate model is typically based on Monte-Carlo techniques.

Iman–Conover: dependence "injection" by sample reordering

Related methods include generation of synthetic multivariate samples from univariate data sets. Whilst the margins of such a synthetic sample accord with the univariate data, its dependence structure is modified to fit the application's needs. The most basic example is the classic Latin Hypercube Sampling method, which mimics independent margins. It is a popular tool for removing spurious correlations from multivariate data sets. This method is also applied to variance reduction in the simulation of independent random variables (cf. [15,23,18,12]). Similar applications to dependent random variables include variance reduction in Monte-Carlo methods [19] and in copula estimation [10].

An extension of Latin Hypercube Sampling that brings dependence into the samples was proposed by Iman and Conover [13]. The original description of the Iman–Conover method uses random reordering of marginal samples, and the intention there was to control the rank correlations in the synthetic multivariate sample. The reordering is performed according to the vectors of marginal ranks in an i.i.d. sample of some multivariate distribution, say, H, with continuous margins. Thus rank correlations of H are "injected" into the synthetic sample. This procedure is equivalent to plugging empirical margins (obtained from asynchronous observations) into the rank based empirical copula of a sample of H [2]. Moreover, it turned out that the Iman–Conover method allows to introduce not only the rank correlations of H into the synthetic samples, but the entire copula of H (cf. [2,16]). In somewhat weaker sense, these results are related to the approximation of stochastic dependence by deterministic functions and to the pioneering result by Kimeldorf and Sampson [14]. Further developments in that area include measure preserving transformations [26] and shuffles of min [7]. In statistical optimization, reordering techniques were also used by Rüschendorf [21]. A very recent, related application in quantitative risk management is a rearrangement algorithm that computes worst-case bounds for the aggregated loss quantiles in a portfolio with given marginal distributions (cf. [8], and references therein).

Using explicit reorderings of univariate marginal samples, the Iman–Conover method has a unique algorithmic tractability. It is implemented in various software packages, and it serves as a standard tool in dependence modelling and uncertainty analysis. The reordering algorithm allows even to construct synthetic samples with hierarchical dependence structures that meet the needs of risk aggregation in insurance and reinsurance companies [2]. The distribution of the aggregated risk is estimated by the empirical distribution of the component sums $\widetilde{X}_1^{(k)} + \cdots + \widetilde{X}_d^{(k)}$ of the synthetic samples $\widetilde{X}^{(k)} = (\widetilde{X}_1^{(k)}, \ldots, \widetilde{X}_d^{(k)})$ for $k = 1, \ldots, n$. This Monte-Carlo approach has computational advantages. The resulting convergence rate of $n^{-1/2}$ (or even faster with Quasi-Monte-Carlo using special sequences) allows to outperform explicit calculation of sum distributions already for moderate dimensions $d \geq 4$ (cf. [1]).

Challenge and contribution: convergence proofs

Despite its popularity, some applications of the Iman–Conover method have been justified by simulations rather than by mathematical proofs. The original publication [13] derives its conclusions from promising simulation results for the distribution of the following function of a 4-dimensional random vector: $f(X_1, \ldots, X_4) = X_1 + X_2(X_3 - \log|X_1|) + \exp(X_4/4)$. Yet a rigorous proof is still missing. The present paper provides a convergence proof for Iman–Conover estimates of the component sum distribution. It also includes a proof sketch for the much simpler case of the estimated joint distribution. Both problems have been open until now.

The solutions given in this paper are derived from the empirical process theory as presented in [24]. Under appropriate regularity assumptions, Iman–Conover estimates of the sum distribution are strongly uniformly consistent with convergence rate $O_P(n^{-1/2})$ (see Theorems 4.1 and 4.2). The convergence of Iman–Conover estimates for the joint distribution is discussed in cf. Remark 4.8. All these findings are not specific to the component sum and extend immediately to all componentwise non-decreasing functions (see Corollary 4.10). Moreover, Theorems 4.1 and 4.2 also cover the convergence of aggregated risk distributions obtained by Monte-Carlo sampling of a multivariate model constructed by plugging empirical margins into a copula (see Remark 4.9). In fact, both sampling methods (reordering by Iman–Conover and classic top-down sampling with empirical margins instead of the exact ones) lead to the same mathematical problem. This is discussed in Remark 3.2(d).

The regularity assumptions used here to establish the $O_P(n^{-1/2})$ convergence rate for Iman–Conover estimates of sum distributions are satisfied for all copulas with bounded densities. This case includes the independence copula in arbitrary dimension $d \geq 2$. The assumptions are also satisfied for all bivariate Clayton copulas and for bivariate Gauss copulas with correlation parameter $\rho \geq 0$. The convergence rate for $\rho < 0$ is, if at all, only slightly weaker. The best bound that is currently available for $\rho < 0$ is $O_P(n^{-1/2}\sqrt{\log n})$.