M.L. Puri and J.P. Vilaplana, editors

The Kagan Classification of Multivariate Distributions and the Central Limit Theorem

J. WESOLOWSKI
Politechnika Warszawska, Instytut Matematyki, Plac Jednosci Robotniczej 1, 00661 Warszawa, Poland 00661

Abstract – A version of the Lapounov type central limit theorem for triangular array of dependent random variables – belonging row-wisely to the Kagan $\mathcal{D}_{k_n,2}$ classes – is proved.

1. INTRODUCTION

In Kagan (1988) the following classification of multivariate probability distributions has been introduced: A n-dimensional random vector X (or its distribution) belongs to the class $\mathcal{D}_{n.k}$ $(k \leq n)$ if its characteristic function has the form

$$\phi(\mathrm{t}) = \prod_{1 \leq i_1 < \ldots < i_k \leq n} R_{i_1 \ldots i_k}(t_{i_1}, \ldots, t_{i_k}) \tag{1}$$

for any $\mathrm{t} = (t_1, \ldots, t_n) \in \mathrm{R}^n$, where $R_{i_1,\ldots,i_k}$ is a continuous complex function with $R_{i_1,\ldots,i_k}(0, \ldots, 0) = 1$, $1 \leq i_1 < \ldots < i_k \leq n$. If (1) holds in a neighborhood of the origin then X belongs to the class $\mathcal{D}_{n,k}(loc)$.
Many interesting properties of these classes, together with some applications and examples are given in the above mentioned paper. Its main result lies in some extensions of the celebrated Darmois–Skitovitch Theorem (on characterizing the normal law by independence of linear statistics in independent random variables) and the Cramèr Theorem (on decomposition of the normal law). The core of these extensions was a weakening of the independence assumption by applying $\mathcal{D}_{n.k}$ classes. And this is the essential feature of these classes – they are gradual generalizations of independence of components of a random vector. Observe that if $X \in \mathcal{D}_{n,1}$ then its components are independent and if $X \in \mathcal{D}_{n,1}(loc)$ then its components are quasi-independent.
The investigations have been continued in Kagan (1989) where a similar concept of (n, k)-equivalence has been studied (being an extension of the identity of distributions). In Wesolowski (1991a) a formula expressing the characteristic function of $\mathcal{D}_{n,k}$ distribution in terms of characteristic functions of its k-dimensional marginals was obtained. Its immediate consequences are some results on determining the distribution of $\mathcal{D}_{n,k}$ random vectors by k-dimensional marginal distributions. This formula was also applied in Wesolowski (1991b) where the relations between Gaussian conditional structure of the second order and $\mathcal{D}_{n,k}$ classes have been studied.
The factorization of the characteristic function given in Wesolowski (1991a) is also the core of the proof of the Lapounov type central limit theorem for $\mathcal{D}_{k_n,2}$ classes. This is the

subject of the present paper. We consider random vectors $X_n = (X_{n,1}, \ldots, X_{n,k_n})$ with k_n tending to infinity together with n.

2. CENTRAL LIMIT THEOREM OF THE LAPOUNOV TYPE

Recall that the classical Lapounov central limit theorem for triangular array has the following form:

Theorem 1 *Let $X_n = (X_{n,1}, \ldots, X_{n,k_n})$ be a random vector with independent zero-mean components, $n = 1, 2, \ldots$ If*

$$\lim_{n\to\infty} \sum_{i=1}^{k_n} E X_{n,i}^2 = \sigma^2 > 0,$$

$$\lim_{n\to\infty} \sum_{i=1}^{k_n} E \mid X_{n,i} \mid^3 = 0$$

then $S_n = X_{n,1} + \ldots + X_{n,k_n}$ converges in distribution to the normal law with the mean zero and the variance σ^2 ($S_n \xrightarrow{d} \mathcal{N}(0, \sigma^2)$) as $n \to \infty$.

Observe that in the above theorem X_n belongs to the class $\mathcal{D}_{k_n,1}$. Now we are interested in the next Kagan class i.e. we try to obtain the Lapounov type theorem replacing the assumption of independence by considering X_n from the class $\mathcal{D}_{k_n,2}$.

Theorem 2 *Let $X_n = (X_{n,1}, \ldots, X_{n,k_n}) \in \mathcal{D}_{k_n,2}$ be a zero-mean random vector, $n = 1, 2, \ldots$ If*

$$\lim_{n\to\infty} \sum_{i,j=1}^{k_n} E X_{n,i} X_{n,j} = \sigma^2 > 0, \tag{2}$$

$$\lim_{n\to\infty} k_n \sum_{i=1}^{k_n} E X_{n,i}^2 = \tau^2, \tag{3}$$

$$\lim_{n\to\infty} k_n \sum_{i=1}^{k_n} E \mid X_{n,i} \mid^3 = 0, \tag{4}$$

then $S_n \xrightarrow{d} \mathcal{N}(0, \sigma^2)$ as $n \to \infty$.

The proof of the above theorem is given in Section 3. Note that, in spite of the inclusions $\mathcal{D}_{n,1} \subset \mathcal{D}_{n,2}$, due to the relations (2) and (3), Theorem 2 can not be applied for X_n's with independent components. The formulas (2) and (3) jointly are the analogues of the first condition in the Lapounov Theorem and (4) works as the Lapounov condition.
Consider now zero-mean random vector $X_n \in \mathcal{D}_{k_n,k}$, $n = 1, 2, \ldots$, where k is an arbitrary natural number. Then by Theorem 1 in Wesolowski (1991a) in some neighborhood of the origin we have

$$E \exp(itS_n) = \prod_{r=1}^{k} \left(\prod_{1 \le i_1 < \ldots < i_r \le k_n} \phi_{n;i_1,\ldots,i_r}(t, \ldots, t) \right)^{a(k_n,k,r)}, \tag{5}$$

where $a(m,k,r) = \sum_{i=0}^{k-r} \binom{m-r}{i} (-1)^i$ and $\phi_{n;i_1,\ldots,i_r}$ is a characteristic function of the marginal $(X_{n,i_1}, \ldots, X_{n,i_r})$, $1 \le i_1 < \ldots < i_r \le k_n$.
We hope that by applying to (5) a reasoning similar to the proof given in Section 3 the following general theorem may be proved, which is formulated as a conjecture since we have not been able to prove it yet.

Conjecture 1 *Let $X_n = (X_{n,1}, \dots, X_{n,k_n}) \in \mathcal{D}_{k_n,k}$ be a zero-mean random vector, $n = 1, 2, \dots$, where k is a fixed natural number and $k_n \geq k$ for each n. If* (2) *and*

$$\lim_{n\to\infty} \binom{k_n}{k-1} \sum_{i=1}^{k_n} E X_{n,i}^2 = \tau^2 < \infty,$$

$$\lim_{n\to\infty} \binom{k_n}{k-1} \sum_{i=1}^{k_n} E \mid X_{n,i} \mid^3 = 0,$$

hold then $S_n \xrightarrow{d} \mathcal{N}(0, \sigma^2)$ as $n \to \infty$,

Note that Theorems 1 and 2 are immediate consequences of the above conjecture if only it is true. It suffices to take $k = 1, 2$, respectively.

3. PROOF

Consider independent random variables $Y_{n,j}$, $j = 1, 2, \dots, k_n^2$ such that

$$Y_{n,(j-1)k_n+i} \stackrel{d}{=} X_{n,j}, \qquad j, i = 1, \dots, k_n,$$

where $\stackrel{d}{=}$ denotes equation of distributions, $n = 1, 2, \dots$ Observe that

$$\sum_{i=1}^{k_n^2} E Y_{n,i}^2 = k_n \sum_{i=1}^{k_n} E X_{n,i}^2$$

and

$$\sum_{i=1}^{k_n^2} E \mid Y_{n,i} \mid^3 = k_n \sum_{i=1}^{k_n} E \mid X_{n,i} \mid^3 .$$

Consequently by (3) and (4) the assumptions of Theorem 1 are fulfilled. Hence

$$\left(\prod_{1 \leq i \leq k_n} \phi_{n;i}(t) \right)^{k_n} \to \exp\left(-\frac{t^2 \tau^2}{2} \right) \tag{6}$$

for any real t as $n \to \infty$.
Define now independent random variables $Z_{n;i,j}$, $1 \leq i < j \leq k_n$ such that

$$Z_{n;i,j} \stackrel{d}{=} X_{n,i} + X_{n,j}$$

for any $n = 1, 2, \dots$ Since

$$\sum_{1 \leq i < j \leq k_n} E(X_{n,i} + X_{n,j})^2 = (k_n - 2) \sum_{1 \leq i \leq k_n} E X_{n,i}^2 + \sum_{i,j=1}^{k_n} E X_{n,i} X_{n,j}$$

then by (2) and (3)

$$\lim_{n\to\infty} \sum_{1 \leq i < j \leq k_n} E Z_{n;i,j}^2 = \tau^2 + \sigma^2, \tag{7}$$

On the other hand

$$\begin{aligned} \sum_{1 \leq i < j \leq k_n} E \mid X_{n,i} + X_{n,j} \mid^3 &\leq 4 \sum_{1 \leq i < j \leq k_n} (E \mid X_{n,i} \mid^3 + E \mid X_{n,j} \mid^3) \\ &\leq 4(k_n - 1) \sum_{1 \leq i \leq k_n} E \mid X_{n,i} \mid^3 . \end{aligned}$$

Consequently by (4)

$$\lim_{n\to\infty} \sum_{1\le i<j\le k_n} E \mid Z_{n;i,j} \mid^3 = 0. \tag{8}$$

The formulas (7) and (8) together with Theorem 1 yield

$$\lim_{n\to\infty} \prod_{1\le i<j\le k_n} \phi_{n;i,j}(t) = \exp\left(-\frac{t^2(\tau^2+\sigma^2)}{2}\right) \tag{9}$$

for any real t.
The equations (6) and (9) imply that there is a neighborhood V of the origin such that for any $n > N$, where N is sufficiently large $\phi_{n;i}(t)$, $i = 1, 2, \ldots, k_n$ and $\phi_{n;i,j}(t,t)$, $1 \le i < j \le k_n$ are non-zero for every $t \in V$. Hence by Theorem 1 from Wesolowski (1991a) - put $k = 2$ in the formula (5) above - we obtain for any $t \in V$ and any $n > N$

$$E\,\exp(itS_n) = \frac{\prod_{1\le i<j\le k_n} \phi_{n;i,j}(t,t)}{\left(\prod_{1\le i\le k_n} \phi_{n;i}(t)\right)^{k_n-2}}$$

Once again applying (6) and (9) we find that for $t \in V$

$$E\,\exp(itS_n) \to \exp\left(-\frac{t^2\sigma^2}{2}\right). \tag{10}$$

To complete the proof observe that by the Tschebyshev inequality

$$P(\mid S_n \mid > \lambda) \le \lambda^{-2} E\, S_n^2$$

and (2) the sequence of distributions of $S_n, n = 1, 2, \ldots$ is tight. Consequently from the Prokhorov Theorem it follows that each subsequence contains a convergent subsequence. Thus we can assume that S_n converges in distribution. Hence the characteristic functions of $S_n, n = 1, 2, \ldots$ are convergent to some characteristic function ψ which, by (10), in a neighborhood of the origin is equal to the characteristic function of $\mathcal{N}(0, \sigma^2)$. It is well known that in such a case for any real t

$$\psi(t) = \exp\left(-\frac{t^2\sigma^2}{2}\right). \qquad \blacksquare$$

REFERENCES

[1] KAGAN, A.M. (1988) New Classes of Dependent Random Variables and a Generalization of the Darmois–Skitovitch Theorem to the Case of Several Forms. *Teor. Veroyat. i Primen.*, **33**, 305–315 (in Russian).

[2] KAGAN, A.M. (1989) Generalized Condition of the Identity of Distributions of Random Vectors in Connection with the Analytic Theory of Linear Forms in Independent Random Variables. *Teor. Veroyat. i Primen.*, **34**, 370–375 (in Russian).

[3] WESOLOWSKI, J. (1991a) On Determination of Multivariate Distribution by its Marginals for the Kagan Class. *J. Multivar. Anal.*, **36**, 314–319.

[4] WESOLOWSKI, J. (1991b) Gaussian Conditional Structure of the Second Order and the Kagan Classification of Multivariate Distributions. *J. Multivar. Anal.*, **39**, 79–86.