On the limiting distribution of sample central moments

Georgios Afendras · **Nickos Papadatos** · **Violetta E. Piperigou**

Abstract We investigate the limiting behavior of sample central moments, examining the special cases where the limiting (as the sample size tends to infinity) distribution is degenerate. Parent (non-degenerate) distributions with this property are called *singular*, and we show in this article that the singular distributions contain at most three supporting points. Moreover, using the *delta*-method, we show that the (second order) limiting distribution of sample central moments from a singular distribution is either a multiple, or a difference of two multiples of independent chi-square random variables with one degree of freedom. Finally, we present a new characterization of normality through the asymptotic independence of the sample mean and all sample central moments.

Keywords sample central moments · singular distributions · second order approximation · characterization of normality · *delta*-method

1 Introduction

Let X be a random variable with distribution function F and finite moment of order k, for some positive integer $k \geq 2$. Then, X has finite central moment of order k.

Based on a random sample of size n from F, a natural estimator of the kth central moment of X is the kth sample central moment, and the strong law of large numbers implies that the kth sample central moment is a strongly consistent estimator of the

Georgios Afendras
Department of Mathematics, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece
E-mail: gafendra@math.auth.gr

Nickos Papadatos (Corresponding Author)
Department of Mathematics, National and Kapodistrian University of Athens, Panepistemiopolis, 15784 Athens, Greece.
E-mail: npapadat@math.uoa.gr

Violetta E. Piperigou
Department of Mathematics, University of Patras, 26500 Rio, Greece.
E-mail: vpiperig@math.upatras.gr

population kth central moment. If, in addition, X has finite moment of order $2k$, its asymptotic normality is also known (see, for example, Lehmann 1999, pp. 297–8).

In the particular case where $k = 2$ and the sample size $n \geq 2$ is fixed, Kourouklis (2012) proved that the usual unbiased estimator for the variance, S^2, is inadmissible in the class $\mathcal{C} = \{cS^2 : c > 0\}$, showing that the estimator $n(n-1)[n(n-1)+2]^{-1}S^2$ has smaller mean squared error than S^2 for all F with finite fourth moment; see also Yatracos (2005).

On the other hand, various authors provide statistical inference based on the asymptotic (as $n \to \infty$) distribution of a function of the sample central moments. Such results have several applications, including the evaluation of the limiting distribution of process capability indices, which have been widely used to measure the improvement of quality and productivity (see, e.g., Wu and Liang 2010). In a different context, Pewsey (2005) and Afendras (2013) provide hypothesis testing, including normality-testing, based on a function of the first four central moments of a distribution.

Haug et al. (2007) suggest moment estimators for the parameters of a continuous time $\mathrm{GARCH}(1,1)$. The asymptotic normality of these estimators plays an important role in their analysis.

Investigating the M-estimation procedure, Stefanski and Boos (2002) present cases in which central moment-based estimates may be presented as M-estimators. The asymptotic analysis and approximate inference are an important issue for large-sample inference.

The sequence of sample central moments, after suitable centering, has a k-variate limiting normal distribution; this result arises easily from the multivariate central limit theorem and the *delta*-method. However, there are cases where the asymptotic distribution is degenerate. In such cases, the order of convergence is faster than $\sqrt{n}$, specifically, the convergence is of order n. Therefore, a deeper study of the asymptotic behavior of the sample central moments is required.

This paper is organized as follows. Section 2 provides the basic notation and terminology that will be used through the paper. Section 3 presents a motivation of the problems that are studied and lists our contributions. Section 4 provides general asymptotics of the sample central moments. Specifically, we introduce the property of asymptotic independence, and investigate this property for the random vector of the first k central moments; an asymptotic independence-based characterization for the normal distribution is also given. In Section 5, we introduce the notion of a singular distribution and we study the class of such distributions, while Section 6 contains results associated with the asymptotic distribution of sample central moments under singularity. Some of the proofs are presented in the Appendix.

2 Notation and Terminology

Let $X \sim F$ with $\mathbb{E}|X|^k < \infty$ for some (fixed) $k \in \{1,2,\ldots\}$; and let us consider a random sample $X_1,\ldots,X_n$ from F. To avoid trivialities we further assume that X is non-degenerate, that is, the set of points of increase of F,

$$S_F = \{x \in \mathbb{R} : F(x+\varepsilon) - F(x-\varepsilon) > 0 \text{ for all } \varepsilon > 0\},$$

contains at least two elements. The first k central moments of X around its mean, $\mu = \mathbb{E}(X)$, are well-defined and finite. In the sequel, we shall use the notation

$$\mu_j = \mathbb{E}(X-\mu)^j, \quad j = 0,\dots,k.$$

The centered sample moments are

$$m_{j,n} = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^j, \quad j = 1,\dots,k.$$

The moment estimator of μ_k (for $k \geq 2$) when μ is unknown (as is usually the case) is its sample counterpart,

$$M_{k,n} = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X}_n)^k, \quad \text{where} \quad \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n}X_i;$$

for convenience, we set $M_{1,n} = \bar{X}_n - \mu$.

Now, we define the vectors

$$\boldsymbol{\mu}_k = (\mu_1,\dots,\mu_k)' = \left(0,\sigma^2,\mu_3,\dots,\mu_k\right)' \quad \text{and} \quad \boldsymbol{\mu}_k^* = \left(\sigma^2,\mu_3,\dots,\mu_k\right)',$$

as well as the random vectors

$$\boldsymbol{M}_{k,n} = (\bar{X}_n-\mu, M_{2,n},\dots,M_{k,n})', \quad \boldsymbol{M}_{k,n}^* = (M_{2,n},\dots,M_{k,n})', \quad \boldsymbol{m}_{k,n} = (m_{1,n},\dots,m_{k,n})';$$

it is worth noting that it is convenient to find the asymptotic distribution of $\sqrt{n}(\boldsymbol{M}_{k,n} - \boldsymbol{\mu}_k)$ instead of $\sqrt{n}(\boldsymbol{M}_{k,n}^* - \boldsymbol{\mu}_k^*)$.

Observe that, by Newton's formula, $M_{j,n} = g_{j,k}(\boldsymbol{m}_{k,n})$, where for $\boldsymbol{x}_k = (x_1,\dots,x_k)'$

$$g_{j,k}(\boldsymbol{x}_k) = (-1)^{j-1}(j-1)x_1^j + \sum_{i=2}^{j-1}(-1)^{j-i}\binom{j}{i}x_i x_1^{j-i} + x_j, \quad j = 1,\dots,k,$$

and an empty sum should be treated as zero. Therefore, $\boldsymbol{M}_{k,n} = \boldsymbol{g}_k(\boldsymbol{m}_{k,n})$, where $\boldsymbol{g}_k = (g_{1,k},\dots,g_{k,k})'$.

Finally, let $\boldsymbol{X}_n$ be a sequence of random vectors. The terminology $\boldsymbol{X}_n$ $\sqrt{n}$-converges in distribution to a distribution, say F_0, means that there exists $\boldsymbol{\mu}$ such that $\sqrt{n}(\boldsymbol{X}_n - \boldsymbol{\mu}) \xrightarrow{\text{d}} F_0$ as $n \to \infty$; similarly, we define n-convergence. In the rest of the paper, all limiting behaviors (limits, convergence in distribution or in probability as well as $o(\cdot)$, $O(\cdot)$ and $o_p(\cdot)$ functions) will be with respect to $n \to \infty$.

3 Motivation and our contributions

Based on the asymptotic distribution of the vector of the sample skewness and kurtosis, Pewsey (2005) gave an asymptotic result for testing normality. Afendras (2013) establishes moment-based estimators of the parameter vector of the characteristic quadratic polynomial for both, integrated Pearson and cumulative Ord families of distributions, and obtained the asymptotic distribution of those estimators. Using this asymptotic distribution, he provides a number of hypothesis testing, including a normality test. In both cases, i.e., sample skewness and kurtosis (Pewsey 2005) and parameter vector of the characteristic quadratic polynomial (Afendras 2013), the estimator is a function of $\boldsymbol{M}_{4,n}$. Thus, it is of some interest to obtain the asymptotic distribution of $\sqrt{n}(\boldsymbol{M}_{k,n}-\boldsymbol{\mu}_k)$ for any value of k.

Our contributions are as follows.

1. We give some more light on the limiting behavior of the vector $\boldsymbol{M}_{k,n}$. In particular, we present results related to the rate of convergence of the first and second moments of $\boldsymbol{M}_{k,n}$. Furthermore, we investigate in some detail the singular cases, i.e., the cases where $v_k^2=0$ (see (3) below), characterizing the distributions with this property.
2. We introduce the notion of *asymptotic independence* between the components of a sequence of k-dimensional random vectors, and we investigate the asymptotic properties of $\boldsymbol{M}_{k,n}$ in view of this notion. Specifically, we show that, among the distributions having finite moments of any order, the asymptotic independence of $\bar{X}_n$ and the sequence $\{M_{k,n}, k\geq 2\}$ characterizes the normal distribution. This fact provides, in a sense, a limiting counterpart of the well-known result that independence of $\bar{X}_n$ and $M_{2,n}=(1-1/n)S_n^2$ (for some fixed $n\geq 2$) characterizes normality (see Geary 1936; Zinger 1958; Laha et al. 1960; Kagan et al. 1973). Here, the assumption of independence is weakened to *asymptotic independence* but, of course, the requirement of the existence of all moments and the fact that $\bar{X}_n$ has to be asymptotically independent *of all* $M_{k,n}, k\geq 2$ (and not only $k=2$), seems to be quite restricted. However, this result is best possible. Indeed, as we shall show, for any fixed $k\geq 2$ there are (infinitely many) non-normal distributions for which $\bar{X}_n$ and $\boldsymbol{M}^*_{k,n}$ are asymptotically independent.
3. Let $k=2,3,\ldots$ be fixed such that $\mathbb{E}|X|^{2k}<\infty$. Under non-singularity of order k, that is $v_k^2\neq 0$, the $\sqrt{n}$-convergence of $M_{k,n}$ is a well-known result, i.e., $\sqrt{n}(M_{k,n}-\mu_k)$ converges in distribution to $N(0,v_k^2)$. Under singularity of order k we shall verify the n-convergence of $M_{k,n}$, i.e., $n(M_{k,n}-\mu_k)$ converges in distribution to a non-normal distribution.

4 The limiting distribution and a characterization of normality

Assume that $k\geq 2$ and $\mathbb{E}|X|^{2k}<\infty$. The multivariate central limit theorem immediately yields that

$$\sqrt{n}(\boldsymbol{m}_{k,n}-\boldsymbol{\mu}_k)\xrightarrow{\text{d}} N_k(\mathbf{0}_k,\boldsymbol{\Sigma}_k), \tag{1}$$

where $\mathbf{0}_k=(0,\ldots,0)'\in\mathbb{R}^k$ and $\boldsymbol{\Sigma}_k=(\sigma_{ij})\in\mathbb{R}^{k\times k}$ with $\sigma_{ij}=\mu_{i+j}-\mu_i\mu_j$. Since $\boldsymbol{M}_{k,n}=\boldsymbol{g}_k(\boldsymbol{m}_{k,n})$, the asymptotic distribution of $\sqrt{n}(\boldsymbol{M}_{k,n}-\boldsymbol{\mu}_k)$ easily arises by a simple application of *delta*-method and is a k-dimensional normal distribution (see, e.g., van der Vaart 1998, Theorem 3.1). This result is presented in the following proposition.

Proposition 1 *If* $\mathbb{E}|X|^{2k}<\infty$, *then*

$$\sqrt{n}(\boldsymbol{M}_{k,n}-\boldsymbol{\mu}_k)\xrightarrow{\text{d}} N_k(\mathbf{0}_k,\mathbf{V}_k), \tag{2}$$

where the variance-covariance matrix $\mathbf{V}_k = (v_{ij}) \in \mathbb{R}^{k\times k}$ *has elements*

$$v_{11} = \sigma^2, \tag{3a}$$
$$v_{1j} = v_{j1} = \mu_{j+1} - j\sigma^2\mu_{j-1}, \qquad j = 2,\ldots,k, \tag{3b}$$
$$v_{ij} = \mu_{i+j} - \mu_i\mu_j - i\mu_{i-1}\mu_{j+1} - j\mu_{i+1}\mu_{j-1} + ij\sigma^2\mu_{i-1}\mu_{j-1}, \quad i,j = 2,\ldots,k; \tag{3c}$$

the elements v_{ii} *are also denoted by* v_i^2, $i = 1,\ldots,k$.

The proof of Proposition 1 for the case $k = 4$ is contained in Afendras (2013, in the proof of Theorem 3.1), while the proof for general k is similar to the case $k = 4$. Particular cases of the preceding result are contained in the next corollary.

Corollary 1 *If* $k \ge 2$ *and* $\mathbb{E}|X|^{2k} < \infty$, *then*

$$\sqrt{n}(M_{k,n} - \mu_k) \xrightarrow{\text{d}} N\big(0, v_k^2\big); \tag{4}$$

$$\sqrt{n}\begin{pmatrix}\bar{X}_n - \mu\\ M_{k,n} - \mu_k\end{pmatrix} \xrightarrow{\text{d}} N_2\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\sigma^2 & \mu_{k+1} - k\sigma^2\mu_{k-1}\\ \mu_{k+1} - k\sigma^2\mu_{k-1} & v_k^2\end{pmatrix}\right). \tag{5}$$

Note that, as it is well-known, the weak convergence in (2) does not imply convergence of the corresponding moments; e.g., it is not necessarily true that either $\mathbb{E}[\sqrt{n}(M_{k,n} - \mu_k)] \to 0$ or $\mathsf{Var}[\sqrt{n}(M_{k,n} - \mu_k)] \to v_k^2$. Therefore, it is an interesting fact that (2) correctly suggests the limits for the corresponding expectations, variances and covariances. The following proposition asserts that this moment convergence is indeed satisfied when the minimal (natural) set of assumptions is imposed on the moments of X; detailed proofs are given in Appendix A.

Proposition 2 *Let* $k, r \in \{2,3,\ldots\}$ *be fixed.*
(a) *If* $\mathbb{E}|X|^k < \infty$, *then* $\mathbb{E}(M_{k,n}) = \mu_k + o(1/\sqrt{n})$;
(b) *If* $\mathbb{E}|X|^{k+1} < \infty$, *then* $\mathsf{Cov}(\bar{X}_n, M_{k,n}) = (\mu_{k+1} - k\sigma^2\mu_{k-1})/n + o(1/n)$;
(c) *If* $\mathbb{E}|X|^{k+r} < \infty$, *then* $\mathsf{Cov}(M_{r,n}, M_{k,n}) = v_{rk}/n + o(1/n)$, *and in particular, if* $\mathbb{E}|X|^{2k} < \infty$, *then* $\mathsf{Var}[\sqrt{n}(M_{k,n} - \mu_k)] \to v_k^2$.

In the sequel, we shall make use of the following definition.

Definition 1 Let $k \in \{2,3,\ldots\}$ be fixed.
(a) The sample mean, $\bar{X}_n$, is called *asymptotically independent* of the sample central moment, $M_{k,n}$, if there exist independent random variables W_1 and W_k such that

$$\sqrt{n}\begin{pmatrix}\bar{X}_n - \mu\\ M_{k,n} - \mu_k\end{pmatrix} \xrightarrow{\text{d}} \begin{pmatrix}W_1\\ W_k\end{pmatrix};$$

(b) $\bar{X}_n$ is called *asymptotically independent* of the random vector $\boldsymbol{M}_{k,n}^*$ if there exist random variables $W_1,\ldots,W_k$ such that $W_1, \boldsymbol{W}_k^* = (W_2,\ldots W_k)'$ are independent and

$$\sqrt{n}\begin{pmatrix}\bar{X}_n - \mu\\ \boldsymbol{M}_{k,n}^* - \boldsymbol{\mu}_k^*\end{pmatrix} \xrightarrow{\text{d}} \begin{pmatrix}W_1\\ \boldsymbol{W}_k^*\end{pmatrix};$$

(c) $\bar{X}_n$ and $M_{k,n}$ are called *asymptotically uncorrelated* if

$$\mathsf{Cov}\big(\sqrt{n}\bar{X}_n, \sqrt{n}M_{k,n}\big) \to 0.$$

Remark 1 Assume that $\mathbb{E}|X|^{2k} < \infty$ for some $k \in \{2,3,\ldots\}$. According to (5) and Proposition 2(b) (see, also, (34)), $\bar{X}_n$ and $M_{k,n}$ are asymptotically independent if and only if they are asymptotically uncorrelated. Also, the asymptotic normality of Proposition 1 shows that $\bar{X}_n$ and $\boldsymbol{M}^*_{k,n}$ are asymptotically independent if and only if $\bar{X}_n$ and $M_{r,n}$ are asymptotically uncorrelated for all $r \in \{2,\ldots,k\}$. If we merely assume that $\mathbb{E}|X|^{k+1} < \infty$, then, even for those cases where the limiting distribution of $\sqrt{n}(M_{k,n} - \mu_k)$ does not exist, Proposition 2(b) enables one to decide if $\bar{X}_n$ and $M_{k,n}$ are asymptotically uncorrelated, or not.

Assume now $X \sim N(\mu,\sigma^2)$. Observing the dispersion matrix in (2), it becomes clear that the first column – except of the first element, σ^2 – vanish. This is so because $\mu_k = 0$ for all odd k and $\mu_{2r} = \sigma^{2r}(2r)!/(2^r r!)$; thus, for any $k \in \{2,3,\ldots\}$, $\mu_{k+1} = k\sigma^2\mu_{k-1}$. According to Definition 1, this means that $\bar{X}_n$ is asymptotically independent (uncorrelated) of all $M_{k,n}$. But, this is not a surprising fact for the normal distribution, since it is well-known that for any fixed $n \ge 2$, $\bar{X}_n$ is independent of the vector $\boldsymbol{Z} = (X_1 - \bar{X}_n,\ldots,X_n - \bar{X}_n)'$ (it suffices to observe that $(\bar{X}_n, X_1 - \bar{X}_n,\ldots,X_n - \bar{X}_n)'$ follows a multivariate normal distribution and $\mathsf{Cov}(\bar{X}_n, X_i - \bar{X}_n) = 0$ for all i) and, therefore, $\bar{X}_n$ is stochastically independent (and uncorrelated) of any sequence of the form $\{h_r(\boldsymbol{Z}), r = 2,3,\ldots\}$, where $h_r\colon \mathbb{R}^n \to \mathbb{R}$ are arbitrary Borel functions. Since $M_{r,n} = n^{-1}\sum_{i=1}^n (X_i - \bar{X}_n)^r = h_r(\boldsymbol{Z})$, it follows that $\bar{X}_n$ and $\boldsymbol{M}^*_{k,n}$ are independent (and, thus, $\bar{X}_n$ and $M_{k,n}$ are uncorrelated) for all k and n and, certainly, the same is true for their limiting distributions. The interesting fact is that the converse is also true, i.e., the asymptotic independence of $\bar{X}_n$ and $M_{k,n}$ for all k characterizes normality.

Theorem 1 *Assume that X is non-degenerate and has finite moments of any order. If $\bar{X}_n$ and $M_{k,n}$ are asymptotically independent* (or, merely, asymptotically uncorrelated) *for all $k \in \{2,3,\ldots\}$, then X follows a normal distribution.*

Proof From Proposition 2(b) (cf. (2), (5)) it follows that $\bar{X}_n$ and $M_{k,n}$ are asymptotically uncorrelated if and only if $\mu_{k+1} = k\sigma^2\mu_{k-1}$. Since we have assumed that this relation holds *for all* $k \ge 2$ it follows that $\mu_1 = \mu_3 = \mu_5 = \cdots = 0$ and, similarly, for all $r \in \{1,2,\ldots\}$, $\mu_{2r} = \sigma^{2r}(2r)!/(2^r r!)$. But, these are the moments of $N(0,\sigma^2)$, and since normal distributions are characterized by their moment sequence (see, e.g., Billingsley 1995, Example 30.1, p. 389), we conclude that $X - \mu \sim N(0,\sigma^2)$. □

Compared to the classical characterization of normality via the independence of $\bar{X}_n$ and $S_n^2 = [n/(n-1)]M_{2,n}$, the *asymptotic independence* is a much weaker condition to enable a characterization result. For example, (5) and Proposition 2(b) with $k = 2$ (cf. (34)) shows that $\bar{X}_n$ and S_n^2 are asymptotically independent if and only if $\mu_3 = 0$, provided $\mathbb{E}|X|^3 < \infty$. Clearly, the relation $\mathbb{E}(X-\mu)^3 = 0$ is satisfied by any symmetric distribution with finite third moment and by many others. On the other hand, the requirement that the asymptotic independence has to be fulfilled for all $k \ge 2$ may be regarded as too restricted. However, the following result shows that any finite number of k's will not work.

Theorem 2 *For any fixed $k \geq 2$, there exist* (infinitely many) *non-degenerate non-normal random variables X with finite moments of any order such that $\bar{X}_n$ and $\boldsymbol{M}^*_{k,n}$ are asymptotically independent.*

Proof Let $\phi(x) \propto \exp(-x^2/2)$ be the standard normal density and consider the polynomial $P_m(x) = (\mathrm{d}^m/\mathrm{d}x^m)[x^m(1-x)^m]$; i.e., P_m is the shifted Legendre polynomial of degree m. It is well-known that for all $m \geq k+2$, P_m is orthogonal to $\{1, x, \ldots, x^{k+1}\}$ in the interval $[0,1]$, that is,

$$\int_0^1 x^j P_m(x)\,\mathrm{d}x = 0, \quad j = 0, \ldots, k+1.$$

Since P_m is continuous on $[0,1]$, it follows that $0 < \max_{x\in[0,1]} |P_m(x)| = a_m < \infty$. Also, $\min_{x\in[0,1]} \phi(x) = \phi(1) = (2\pi e)^{-1/2} > 0$. Clearly, we can choose an $\varepsilon_m > 0$ small enough to guarantee that $\phi(x) + \varepsilon_m P_m(x) > 0$ for all $x \in [0,1]$ (e.g., $\varepsilon_m = [2a_m(2\pi e)^{1/2}]^{-1}$ suffices). Now, define a sequence of probability densities $\{f_m,\ m \geq k+2\}$ by

$$f_m(x) = \phi(x) + \varepsilon_m P_m(x) \mathbb{1}_{\{0\leq x\leq 1\}}, \quad x \in \mathbb{R},$$

where $\mathbb{1}$ denotes the indicator function (see Figure 1).

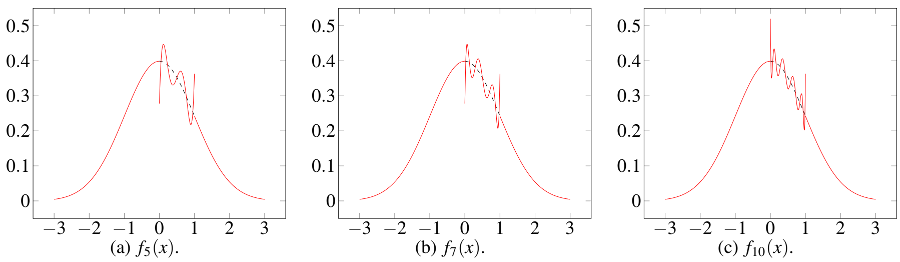

Fig. 1: The densities $f_m(x)$ for $m = 5, 7$ and 10.

If X_m has density f_m, it is clear that, for $j = 0, \ldots, k+1$,

$$\mathbb{E}(X_m^j) = \int_{\mathbb{R}} x^j \phi(x)\,\mathrm{d}x + \varepsilon_m \int_0^1 x^j P_m(x)\,\mathrm{d}x = \int_{\mathbb{R}} x^j \phi(x)\,\mathrm{d}x = \mathbb{E}(Z^j),$$

where $Z \sim N(0,1)$. Obviously, each X_m has finite moments of any order, is non-normal, non-degenerate, and, by Proposition 1, $\bar{X}_n$ and $\boldsymbol{M}^*_{k,n}$ are asymptotically independent. □

5 The singular distributions

First, center the rv X as $U = X - \mu$ with $\mathbb{E}(U^j) = \mu_j$ for all j. Assume that $\mathbb{E}|X|^{2k} < \infty$ for some $k = 2, 3, \ldots$ and consider the random vector

$$\boldsymbol{U}_k = \left(U, U^2, U^3 - 3\sigma^2 U, U^4 - 4\mu_3 U, \ldots, U^k - k\mu_{k-1} U\right)'.$$

It is of some interest to observe that the variance-covariance matrix of the limiting distribution in (2) coincides with the variance-covariance matrix of $\boldsymbol{U}_k$. In particular,

$$v_k^2 = \mathsf{Var}\Big(U^k - k\mu_{k-1}U\Big), \tag{6}$$

$$\mu_{k+1} - k\sigma^2\mu_{k-1} = \mathsf{Cov}\Big(U, U^k - k\mu_{k-1}U\Big),\ v_{rk} = \mathsf{Cov}\Big(U^r - r\mu_{r-1}U, U^k - k\mu_{k-1}U\Big),$$

$r = 2,\ldots,k-1$. Relation (6) evidently shows that $v_k^2 \geq 0$. Of course, the non-negativity of v_k^2 is a consequence of the fact that, by Proposition 2(c), $v_k^2 = \lim_n \mathsf{Var}(\sqrt{n}M_{k,n})$; but, the point here is that we have not to refer to a limit. Moreover, the expression (6) enables us to describe all distributions for which $v_k^2 = 0$. Such distributions will be called *singular*, according to the following definition.

Definition 2 For fixed $k \geq 2$, a non-degenerate random variable X, or its distribution function F, is called singular (of order k) if $\mathbb{E}|X|^{2k} < \infty$ and

$$\sqrt{n}(M_{k,n} - \mu_k) \xrightarrow{\mathrm{p}} 0.$$

The set of all singular random variables of order k will be denoted by $\mathcal{F}_k$; the subset of all standardized (with mean 0 and variance 1) singular random variables of order k will be denoted by $\mathcal{F}_k^0$.

Noting that $Y \in \mathcal{F}_k^0$ if and only if $X = \mu + \sigma Y \in \mathcal{F}_k$ for some $\mu \in \mathbb{R}$ and $\sigma > 0$, it follows that $\mathcal{F}_k$ contains exactly the location-scale family of the random variables that belong to $\mathcal{F}_k^0$. According to (4), (6), and Proposition 2(a),(c) (cf. (32)), $X \in \mathcal{F}_k$ if and only if $v_k^2 = 0$ or, equivalently,

$$(X-\mu)^k = \mu_k + k\mu_{k-1}(X-\mu) \quad \text{with probability one.} \tag{7}$$

We also note that $\mathcal{F}_k$ is non-empty for all $k \geq 2$. Indeed, it is easily seen that the random variable Y with $\mathbb{P}(Y = \pm 1) = 1/2$ belongs to $\mathcal{F}_{2k}^0 \subseteq \mathcal{F}_{2k}$, $k = 1,2,\ldots$, because $\mu = \mathbb{E}(Y) = 0$, $\sigma^2 = \mathbb{E}(Y^2) = 1$, $\mu_{2k} = \mathbb{E}(Y^{2k}) = 1$, $\mu_{2k-1} = \mathbb{E}(Y^{2k-1}) = 0$ and $Y^{2k} = \mu_{2k} + 2k\mu_{2k-1}Y$ with probability one. Similarly, for every $k \in \{1,2,\ldots\}$, the three-valued symmetric random variable Y_{2k+1} with $\mathbb{P}(Y_{2k+1} = \pm\sqrt{2k+1}) = 1/[2(2k+1)]$ and $\mathbb{P}(Y_{2k+1} = 0) = 2k/(2k+1)$ belongs to $\mathcal{F}_{2k+1}^0$. Moreover, we shall show below (Lemma 2) that we can find a unique value of $p = p_{2k+1} \in (1/2, 1)$, for which the two-valued random variable W_{2k+1}, with

$$\mathbb{P}\Big(W_{2k+1} = \sqrt{(1-p)/p}\Big) = p = 1 - \mathbb{P}\Big(W_{2k+1} = -\sqrt{p/(1-p)}\Big),$$

is such that $W_{2k+1} \in \mathcal{F}_{2k+1}^0$.

In general, it is easily seen that the equation $y^k = \alpha + \beta y$ (with $\alpha, \beta \in \mathbb{R}$) has at most two real solutions for even k, and at most three solutions for odd k. Assuming that $X \sim F$ and $X \in \mathcal{F}_k$, it follows from (7) that X takes at most two values (and hence, exactly two values, since X has been assumed to be non-degenerate) if k is even, and two or three values if k is odd. This follows from the fact that the points of increase of F cannot be more than three, if k is odd, and more than two, if k is even. To see this assume, e.g., that k is odd, $X \sim \mathcal{F}_k$, $\mathbb{E}(X) = \mu$, $\mathbb{E}(X-\mu)^k = \mu_k$

and $\mathbb{E}(X-\mu)^{k-1}=\mu_{k-1}$. Let $x_1<x_2<x_3<x_4$ be four distinct points of increase of F. Then, there exists at least one x_i for which $(x_i-\mu)^k-\mu_k-k\mu_{k-1}(x_i-\mu)\neq 0$, and thus, we can find a small $\varepsilon>0$ such that $(x-\mu)^k\neq\mu_k+k\mu_{k-1}(x-\mu)$ for all $x\in(x_i-\varepsilon,x_i+\varepsilon]$. Hence, $\mathbb{P}(x_i-\varepsilon<X\le x_i+\varepsilon)\le\mathbb{P}[(X-\mu)^k\neq\mu_k+k\mu_{k-1}(X-\mu)]$. Since, however, x_i is a point of increase of F, we have $0<F(x_i+\varepsilon)-F(x_i-\varepsilon)=\mathbb{P}(x_i-\varepsilon<X\le x_i+\varepsilon)\le\mathbb{P}[(X-\mu)^k\neq\mu_k+k\mu_{k-1}(X-\mu)]$, which contradicts (7). The same arguments apply to the case where k is even.

Therefore, we have the following description.

Proposition 3 *If $k\ge 2$ is even, then $\mathcal{F}_k$ contains only two-valued random variables. If $k\ge 3$ is odd, then $\mathcal{F}_k$ contains only two-valued and three-valued random variables.*

Our purpose is to describe all singular distributions and to obtain a second order non-degenerate distributional limit for $M_{k,n}-\mu_k$. Firstly, we consider the two-valued distributions because they are possible members of $\mathcal{F}_k$.

Lemma 1 *Let $X\sim b(p)$, i.e., $\mathbb{P}(X=1)=p=1-\mathbb{P}(X=0)$ for some $p\in(0,1)$. Then, $X\in\mathcal{F}_2$ if and only if $p=1/2$. Moreover, if $k\ge 4$ is even, then $X\in\mathcal{F}_k$ if and only if $p\in\{1-p_k,1/2,p_k\}$, where p_k is the unique root of the equation*

$$\left(\frac{p}{1-p}\right)^{k-1}=\frac{(k+1)p-1}{k-(k+1)p},\qquad \frac{k-2}{k-1}<p<\frac{k}{k+1};\tag{8}$$

in particular, $p_4=1/2+\sqrt{3}/6$ and $p_6=1/2+\sqrt{15(4\sqrt{10}-5)}/30$.

Proof Since $\mu=p$ and

$$\mu_k=p(1-p)\Big[(1-p)^{k-1}+(-1)^k p^{k-1}\Big],\quad k=1,2,\ldots,\tag{9}$$

(7) shows that $X\in\mathcal{F}_k$ if and only if

$$(x-p)^k=\mu_k+k\mu_{k-1}(x-p)\quad\text{for } x=0 \text{ and } x=1.\tag{10}$$

Using (9) and the fact that $k\ge 2$ is even, both the equations in (10) are reduced to

$$p^{k-1}[k-(k+1)p]=(1-p)^{k-1}[(k+1)p-1],\quad 0<p<1.\tag{11}$$

Since the value of $p=k/(k+1)$ is a root of the lhs of (11) which is not a root of its rhs we conclude that the equalities in (8) and (11) are equivalent if $0<p<1$. Obviously, $p=1/2$ is a root of (11). In order to find all roots of (11), we make the substitution $t=p/(1-p)$, which monotonically maps $p\in(0,1)$ to $t\in(0,\infty)$. Then, we get the equation

$$p_k(t)=t^k-kt^{k-1}+kt-1=0,\quad 0<t<\infty,\tag{12}$$

which has the obvious solution $t=1$ (corresponding to $p=1/2$). If $k=2$, then (12) is written as $t^2-1=0$, and thus $t=1$ (resp. $p=1/2$) is the unique solution of (12) (resp. (11)). Since for any $t>0$ we have $p_k(1/t)=-p_k(t)/t^k$, it follows that $1/t$ is a root of (12) whenever t is; equivalently, $1-p$ is a root of (11) if p is. For even $k\ge 4$ we see that $p_k(0)=-1$, $p_k(1)=0$ and $p_k(\infty)=\lim_{t\to\infty}p_k(t)=\infty$. Also, $p_k'(t)=k[t^{k-1}-(k-1)t^{k-2}+1]=kq_k(t)$, where $q_k(t)=t^{k-1}-(k-1)t^{k-2}+1$

satisfies $q_k(0) = 1 > 0$, $q_k(1) = -(k-3) < 0$ and $q_k(\infty) = \infty$. Moreover, we see that $q_k'(t) = (k-1)t^{k-3}[t-(k-2)]$ is negative for $t \in (0,k-2)$ and positive for $t \in (k-2,\infty)$; thus, $q_k(t)$ decreases in $(0,k-2)$ and increases in $(k-2,\infty)$. Therefore, there exist $\rho_1 < \rho_2$, with $0 < \rho_1 < 1 < k-2 < \rho_2 < \infty$, such that $q_k(t) < 0$ for t in $(0,\rho_1) \cup (\rho_2,\infty)$ and $q_k(t) > 0$ for t in (ρ_1,ρ_2). Relation $q_k(t) = p_k'(t)/k$ shows that $p_k(t)$ is increasing in $(0,\rho_1)$, decreasing in (ρ_1,ρ_2) and increasing in (ρ_2,∞). From $1 \in (\rho_1,\rho_2)$ and $p_k(1) = 0$, we conclude that $p_k(\rho_1) > 0$ and $p_k(\rho_2) < 0$; hence, there exist unique $t_1 \in (0,\rho_1)$ and $t_2 \in (\rho_2,\infty)$ such that $p_k(t_1) = 0 = p_k(t_2)$. Clearly, $t_1 = 1/t_2$, and the set of roots of (12) is $\{1/t_2, 1, t_2\}$; thus, the set of roots of (11) is $\{1-p_k, 1/2, p_k\}$, with $p_k = t_2/(1+t_2)$. Finally, relation $t_2 > \rho_2 > k-2$ shows that $p_k = t_2/(1+t_2) > (k-2)/[1+(k-2)] = (k-2)/(k-1)$, while $p_k < k/(k+1)$ is obvious because for $p \ge k/(k+1)$ the lhs of (11) is non-positive while its rhs is strictly positive. □

From (8), we see that $1/2 < p_4 < p_6 < \cdots$ and $p_{2k} = 1 - 1/(2k) + o(1/k)$ as $k \to \infty$. Lemma 1 completely describes all $\mathcal{F}_k$ for even k.

Corollary 2 *If $k \ge 2$ is even, then $X \in \mathcal{F}_k^0$ if and only if either $\mathbb{P}(X = \pm 1) = 1/2$, or*

$$\mathbb{P}\Big(X = -\sqrt{p_k/(1-p_k)}\Big) = 1-p_k, \qquad \mathbb{P}\Big(X = \sqrt{(1-p_k)/p_k}\Big) = p_k$$

and $k \in \{4,6,\ldots\}$, or

$$\mathbb{P}\Big(X = -\sqrt{(1-p_k)/p_k}\Big) = p_k, \qquad \mathbb{P}\Big(X = \sqrt{p_k/(1-p_k)}\Big) = 1-p_k$$

and $k \in \{4,6,\ldots\}$, where $p_k \in ((k-2)/(k-1), k/(k+1))$ is given by (8).

Corollary 2 says that $\mathcal{F}_2^0$ is singleton and that for every $k \in \{4,6,\ldots\}$, $\mathcal{F}_k^0$ contains exactly three two-valued distributions. When k is odd the nature of $\mathcal{F}_k$ is quite different, because it contains both two-valued and three-valued distributions. First we examine the two-valued case.

Lemma 2 *Let $X \sim b(p)$ for some $p \in (0,1)$. If $k \ge 3$ is odd and $X \in \mathcal{F}_k$, then $p \in \{1-p_k, p_k\}$ where p_k is the unique root of the equation*

$$\left(\frac{p}{1-p}\right)^{k-1} = \frac{(k+1)p-1}{(k+1)p-k}, \qquad \frac{k}{k+1} < p < 1; \tag{13}$$

in particular, $p_3 = 1/2 + \sqrt{3}/6$ and $p_5 = 1/2 + \sqrt{5\sqrt{5}}/10$. Conversely, if $k \ge 3$ is odd and either $X \sim b(p_k)$ or $X \sim b(1-p_k)$, with p_k as above, then $X \in \mathcal{F}_k$.

Proof Assume that $k \ge 3$ is odd, $X \sim b(p)$ and $X \in \mathcal{F}_k$. This means that (10) is satisfied. Using (9) and the fact that $k \ge 3$ is odd, both equations in (10) are reduced to

$$p^{k-1}[(k+1)p-k] = (1-p)^{k-1}[(k+1)p-1], \quad 0 < p < 1. \tag{14}$$

Since the value of $p = k/(k+1)$ is a root of the lhs of (14) which is not a root of its rhs, we conclude that (13) and (14) are equivalent. As in Lemma 1, in order to find all roots of (14) we make the substitution $t = p/(1-p)$, which monotonically maps $p \in (0,1)$ to $t \in (0,\infty)$. Then, we get the equation

$$p_k(t) = t^k - kt^{k-1} - kt + 1 = 0, \quad 0 < t < \infty. \tag{15}$$

Since for any $t > 0$ we have $p_k(1/t) = p_k(t)/t^k$, it follows that $1/t$ is a root of (15) whenever t is; equivalently, $1-p$ is a root of (14) if p is. For odd $k \geq 3$, we see that $p_k(0) = 1 > 0$, $p_k(1) = -2(k-1) < 0$ and $p_k(\infty) = \infty$. Thus, (15) has at least one root in $(0,1)$ and at least one root in $(1,\infty)$. Also, $p_k'(t) = k[t^{k-1} - (k-1)t^{k-2} - 1] = kq_k(t)$, where $q_k(t) = t^{k-1} - (k-1)t^{k-2} - 1$ satisfies $q_k(0) = -1 < 0$ and $q_k(\infty) = \infty$. Moreover, we see that $q_k'(t) = (k-1)t^{k-3}[t-(k-2)]$ is negative for $t \in (0,k-2)$ and positive for $t \in (k-2,\infty)$; thus, $q_k(t)$ decreases in $(0,k-2)$ and increases in $(k-2,\infty)$. Therefore, there exists a unique $\rho > k-2 \geq 1$ such that $q_k(t) < 0$ for t in $(0,\rho)$ and $q_k(t) > 0$ for t in (ρ,∞). Relation $q_k(t) = p_k'(t)/k$ shows that $p_k(t)$ decreases in $(0,\rho)$ and increases in (ρ,∞). From $p_k(0) > 0$, $p_k(1) < 0$ and $p_k(\infty) > 0$ we conclude that there exist unique $t_1 \in (0,1)$ and $t_2 \in (\rho,\infty)$ such that $p_k(t_1) = 0 = p_k(t_2)$. Clearly, $t_1 = 1/t_2$, and the set of roots of (15) is $\{1/t_2, t_2\}$; thus, the set of roots of (14) is $\{1-p_k, p_k\}$, with $p_k = t_2/(1+t_2)$. Finally, relation $t_2 > \rho > k-2$ shows that $p_k = t_2/(1+t_2) > (k-2)/[1+(k-2)] = (k-2)/(k-1)$. However, the root p_k cannot lie in $((k-2)/(k-1), k/(k+1)]$ because for all p in this interval the lhs of (14) is non-positive, while its rhs is strictly positive ($p > (k-2)/(k-1)$ implies $(k+1)p - 1 > (k+1)(k-2)/(k-1) - 1 = [(k-3)(k+1)+2]/(k-1) > 0$, since $k \geq 3$). This verifies that $p_k > k/(k+1)$. Finally, if either $X \sim b(p_k)$ or $X \sim b(1-p_k)$, then (9) and (14) show that (10) and (7) are satisfied and, thus, $X \in \mathcal{F}_k$. □

Corollary 3 *If $k \geq 3$ is odd, then the unique two-valued random variables contained in $\mathcal{F}_k^0$ are the following:*

$$\mathbb{P}\Big(X = -\sqrt{p_k/(1-p_k)}\Big) = 1 - p_k, \qquad \mathbb{P}\Big(X = \sqrt{(1-p_k)/p_k}\Big) = p_k$$

and

$$\mathbb{P}\Big(X = -\sqrt{(1-p_k)/p_k}\Big) = p_k, \qquad \mathbb{P}\Big(X = \sqrt{p_k/(1-p_k)}\Big) = 1 - p_k,$$

where $p_k \in (k/(k+1), 1)$ is given by (13).

Corollary 3 describes all two-valued random variables of $\mathcal{F}_k^0$ when $k \geq 3$ is odd; however, we have already seen that $\mathcal{F}_k^0$ contains also some three-valued random variables. Among them, exactly one is symmetric.

Lemma 3 *If $k \geq 3$ is odd, the unique symmetric random variable of $\mathcal{F}_k^0$ is given by*

$$\mathbb{P}(X = \pm\sqrt{k}) = 1/(2k), \qquad \mathbb{P}(X = 0) = 1 - 1/k.$$

More generally, this is the unique random variable of $\mathcal{F}_k^0$ with $\mu_k = 0$.

Proof Since $X \in \mathcal{F}_k^0$, we have $\mathbb{E}(X)=0$ and $\mathbb{E}(X^2)=1$. Therefore, in view of the assumption $\mu_k=0$, (7) simplifies to

$$X\left(X^{k-1}-k\mu_{k-1}\right)=0 \quad \text{with probability one.}$$

It follows that the support of X is a subset of $A=\{-(k\mu_{k-1})^{1/(k-1)},0,(k\mu_{k-1})^{1/(k-1)}\}$, where $\mu_{k-1}=\mathbb{E}(X^{k-1})>0$, because X is non-degenerate and $k-1$ is even. Let $p=\mathbb{P}(X=0)$, $p_1=\mathbb{P}(X=-(k\mu_{k-1})^{1/(k-1)})$ and $p_2=\mathbb{P}(X=(k\mu_{k-1})^{1/(k-1)})$; p, p_1, p_2 are non-negative and $p+p_1+p_2=1$ because $\mathbb{P}(X\in A)=1$. Assumption $\mathbb{E}(X)=0$ shows that $p_1=p_2$. Thus, $p_1=p_2=(1-p)/2$ and, so, $\mathbb{P}(X=\pm(k\mu_{k-1})^{1/(k-1)})=(1-p)/2$. Calculating $\mu_{k-1}=\mathbb{E}(X^{k-1})=(1-p)k\mu_{k-1}$, we see that $p=1-1/k$ and thus, $\mathbb{P}(X=\pm a)=1/(2k)$ where $a=(k\mu_{k-1})^{1/(k-1)}>0$. Finally, from $1=\mathbb{E}(X^2)=a^2/k$, we conclude that $a=\sqrt{k}$. On the other hand, it is easily seen that for this value of $a=\sqrt{k}$, $\mu_k=0$ and $\mu_{k-1}=k^{(k-3)/2}$ so that $k\mu_{k-1}=k^{(k-1)/2}=(\pm\sqrt{k})^{k-1}$; hence, $A=\{-\sqrt{k},0,\sqrt{k}\}$ and $x(x^{k-1}-k\mu_{k-1})=x[x^{k-1}-(\pm\sqrt{k})^{k-1}]\equiv 0$ for all $x\in A$. □

The following lemma presents a complete description of all tree-valued distributions of $\mathcal{F}_3^0$ and gives a picture of the nature of $\mathcal{F}_k^0$ when $k\geq 3$ is odd.

Lemma 4 *For each $\mu_3\in[-\sqrt{2},\sqrt{2}]$ there exists a unique random variable $X\in\mathcal{F}_3^0$ such that $\mathbb{E}(X^3)=\mu_3$. Cases $\mu_3=\pm\sqrt{2}$ correspond to the two-valued distributions described in Corollary 3 for $k=3$. Any other value of $\mu_3\in(-\sqrt{2},\sqrt{2})$ uniquely determines a three-valued distribution, and in particular, $\mu_3=0$ corresponds to the symmetric distribution of Lemma 3 for $k=3$. Moreover, there do not exist other random variables in $\mathcal{F}_3^0$. Therefore, $\mathcal{F}_3^0$ admits the parametrization*

$$\mathcal{F}_3^0=\{X_\theta,\ -\sqrt{2}\leq\theta\leq\sqrt{2}\},$$

where X_θ is characterized by

$$\mathbb{E}(X_\theta)=0,\quad \mathbb{E}\left(X_\theta^2\right)=1,\quad \mathbb{E}\left(X_\theta^3\right)=\theta \quad \textit{and} \quad \mathbb{P}\left[X_\theta\left(X_\theta^2-3\right)=\theta\right]=1.$$

Proof Let $X\in\mathcal{F}_3^0$ and assume that $\mu_3=\mathbb{E}(X^3)=\theta$. Then, $\mu=\mathbb{E}(X)=0$, $\sigma^2=\mu_2=\mathbb{E}(X^2)=1$ and, according to (7), $X(X^2-3)=\theta$ with probability one. Therefore, since X is non-degenerate, the support of X must contains at least two points which are included in the set of zeros of $y(y^2-3)=\theta$. This shows that $|\theta|\leq 2$ because, otherwise, the set $\{y\in\mathbb{R}\colon y(y^2-3)=\theta\}$ is a singleton. Observe that $\theta=0$ leads, uniquely, to the symmetric random variable of Lemma 3 with $k=3$. Thus, from now on assume that $\theta\neq 0$. The values $\theta=\pm 2$ are impossible because the equations $y(y^2-3)=\pm 2$ have exactly two real solutions, say α,β, with $|\alpha|=1$ and $|\beta|=2$, so that $\mathbb{E}(X)=0$ and $\mathbb{E}(X^2)=1$ are impossible.

Consider now the case where $-2<\theta<0$. Then, $\{y\colon y(y^2-3)=\theta\}=\{-\alpha,\beta,\gamma\}$ where $0<\beta<1<\gamma<\sqrt{3}<\alpha$ and, by definition, the numbers α, β, γ satisfy the relation

$$-\alpha\left(\alpha^2-3\right)=\beta\left(\beta^2-3\right)=\gamma\left(\gamma^2-3\right)=\theta. \tag{16}$$

From (16), we see that $\theta=\theta(\beta)=\beta(\beta^2-3)$ is a strictly decreasing and continuous function of β which maps $\beta\in(0,1)$ to $\theta\in(-2,0)$; thus, its inverse function,

$\beta(\theta):(-2,0) \to (0,1)$, is well-defined, continuous and strictly decreasing in θ with $\beta(-2_+) = 1$ and $\beta(0_-) = 0$. Also, from (16) we get the equation $3(\alpha+\beta) = \alpha^3+\beta^3 = (\alpha+\beta)(\alpha^2-\alpha\beta+\beta^2)$ which shows that $\alpha^2-\alpha\beta+\beta^2 = 3$ and, since $\alpha > \beta/2$, we have

$$\alpha = \alpha(\beta) = \frac{1}{2}(\beta+\delta), \quad \text{where } \delta = \delta(\beta) = \sqrt{3\big(4-\beta^2\big)}. \tag{17}$$

Similarly, (16) yields the equation $3(\gamma-\beta) = \gamma^3-\beta^3 = (\gamma-\beta)(\gamma^2+\gamma\beta+\beta^2)$ which shows, in view of $\beta < \gamma$, that $\gamma^2+\gamma\beta+\beta^2 = 3$. Since $\gamma > 0$ it follows that

$$\gamma = \gamma(\beta) = \frac{1}{2}(-\beta+\delta), \quad \text{where } \delta = \delta(\beta) = \sqrt{3\big(4-\beta^2\big)}. \tag{18}$$

From (17) and (18) we conclude that $\alpha = \beta+\gamma$. Set now $p_1 = \mathbb{P}(X = -\alpha)$, $p_2 = \mathbb{P}(X = \beta)$ and $p_3 = \mathbb{P}(X = \gamma)$. Since $\mathbb{P}(X \in \{-\alpha,\beta,\gamma\}) = 1$ and $\mathbb{E}(X) = 0$, $\mathbb{E}(X^2) = 1$, we get the system of equations (in p_1, p_2, p_3)

$$p_1+p_2+p_3 = 1, \quad -\alpha p_1+\beta p_2+\gamma p_3 = 0, \quad \alpha^2 p_1+\beta^2 p_2+\gamma^2 p_3 = 1,$$

which, in view of $\alpha = \beta+\gamma$, has the unique solution

$$p_1 = \frac{1+\beta\gamma}{(\beta+2\gamma)(2\beta+\gamma)}, \quad p_2 = \frac{\gamma(\beta+\gamma)-1}{(\gamma-\beta)(2\beta+\gamma)}, \quad p_3 = \frac{1-\beta(\beta+\gamma)}{(\gamma-\beta)(\beta+2\gamma)}.$$

Now, since $\gamma^2+\gamma\beta+\beta^2 = 3$, we have $\gamma(\beta+\gamma) = 3-\beta^2$ and $\beta(\beta+\gamma) = 3-\gamma^2$; substituting these values in the numerators of p_2 and p_3 we get

$$p_1 = \frac{1+\beta\gamma}{(\beta+2\gamma)(2\beta+\gamma)}, \quad p_2 = \frac{2-\beta^2}{(\gamma-\beta)(2\beta+\gamma)}, \quad p_3 = \frac{\gamma^2-2}{(\gamma-\beta)(\beta+2\gamma)}. \tag{19}$$

It is clear that $p_1 > 0$ and $p_2 > 0$ for all values of β and γ with (see (18))

$$0 < \beta < 1 < \gamma = \frac{-\beta+\sqrt{3\big(4-\beta^2\big)}}{2} < \sqrt{3}.$$

However, this is not the case for p_3, since $p_3 \ge 0$ requires $\gamma^2 \ge 2$, i.e., $\gamma \ge \sqrt{2}$ (since $\gamma > 0$). Now, from $\mu_3 = \theta = \gamma(\gamma^2-3)$ and the fact that $\gamma \in [\sqrt{2},\sqrt{3})$, we conclude that all possible values of θ (with $\theta < 0$) are included in the interval $[-\sqrt{2},0)$. Using (18) and the fact that $\beta > 0$, it follows that $\gamma \ge \sqrt{2}$ if and only if $0 < \beta \le (\sqrt{6}-\sqrt{2})/2 = \sqrt{2-\sqrt{3}}$. Now, observe that $\gamma = \sqrt{2}$ corresponds to a standardized two-valued random variable with $\mu_3 = \theta = \gamma(\gamma^2-3) = -\sqrt{2}$, taking the values $-\alpha = -\beta-\gamma = -(\sqrt{6}+\sqrt{2})/2 = -\sqrt{2+\sqrt{3}}$ and $\beta = \sqrt{2-\sqrt{3}} = (\sqrt{6}-\sqrt{2})/2$ with respective probabilities $1-p$ and p, where

$$p = \frac{2-\beta^2}{(\gamma-\beta)(2\beta+\gamma)} = \frac{1}{2}+\frac{\sqrt{3}}{6};$$

this is the first two-valued random variable of Corollary 3 when $k = 3$. On the other hand, each value of $\gamma \in (\sqrt{2},\sqrt{3})$ corresponds to a unique value of $\mu_3 = \theta = \gamma(\gamma^2-3) \in (-\sqrt{2},0)$, which, in turn, uniquely determines $\beta = \beta(\theta) \in (0,(\sqrt{6}-\sqrt{2})/2)$ through $\beta = [-\gamma+\sqrt{3(4-\gamma^2)}]/2$ (cf. (18)), and $\alpha = \alpha(\theta)$ through $\alpha = \beta+\gamma$. Finally, these uniquely determined values of α, β and γ specify the (strictly positive) probabilities p_1, p_2 and p_3, through (19), which shows that each $X_\theta \in \mathcal{F}_3^0$ is uniquely determined by $\mathbb{E}(X_\theta^3) = \theta$, $-\sqrt{2} < \theta < 0$.

It remains to examine the cases where $0 < \theta < 2$. However, if $X \in \mathcal{F}_3^0$ and $\mathbb{E}(X^3) = \theta > 0$, then it is easily verified that $-X \in \mathcal{F}_3^0$ and $\mathbb{E}(-X)^3 = -\theta < 0$. By the previous arguments it follows that, necessarily, $-\sqrt{2} \le -\theta < 0$, that $-X$ is determined by the value of $-\theta$, and that $-X$ is a two-valued random variable, if $-\theta = -\sqrt{2}$, and a three-valued random variable otherwise; thus, the same is true for X, and the proof is complete. □

We was not able to completely describe $\mathcal{F}_k^0$ for odd $k \ge 5$. However, the situation seems to be similar to the case $k = 3$, i.e., each $X_\theta \in \mathcal{F}_k^0$ is characterized by its central moment, $\theta = \mathbb{E}(X^k) = \mu_k$, and the possible values of θ form a symmetric interval of the form $[-\alpha_k, \alpha_k]$, where the boundary values $\theta = \pm\alpha_k$ correspond to the two-valued distributions of Corollary 3, while every $\theta \in (-\alpha_k, \alpha_k)$ determines uniquely a three-valued random variable.

6 Limiting distribution under singularness

If the random sample comes from a singular distribution of order $k \ge 2$, then the asymptotic normality of (4) reduces to $\sqrt{n}(M_{k,n} - \mu_k) \xrightarrow{\text{p}} 0$ (see Definition 2). This shows that the order of convergence of $M_{k,n}$ to μ_k is faster than $o(1/\sqrt{n})$, and a second order approximation applies, according to the following lemma.

Lemma 5 *Assume that $\boldsymbol{X}_n$ is a sequence of k-variate random vectors such that*

$$\sqrt{n}(\boldsymbol{X}_n - \boldsymbol{\mu}) \xrightarrow{\text{d}} \boldsymbol{W}, \tag{20}$$

where $\boldsymbol{\mu} \in \mathbb{R}^k$ and $\boldsymbol{W}$ is a k-variate random vector. Suppose that the Borel function $g : \mathbb{R}^k \to \mathbb{R}$ is twice continuously differentiable at a neighborhood of $\boldsymbol{\mu}$ and define

$$\nabla g(\boldsymbol{\mu}) = \left(\frac{\partial g(\boldsymbol{x})}{\partial x_i} \right) \bigg|_{\boldsymbol{x}=\boldsymbol{\mu}} \in \mathbb{R}^k \quad \text{and} \quad \mathbf{H}_k(\boldsymbol{\mu}) = \left(\frac{\partial^2 g(\boldsymbol{x})}{\partial x_i \partial x_j} \right) \bigg|_{\boldsymbol{x}=\boldsymbol{\mu}} \in \mathbb{R}^{k\times k}.$$

If

$$n[\nabla g(\boldsymbol{\mu})]'(\boldsymbol{X}_n - \boldsymbol{\mu}) \xrightarrow{\text{p}} 0, \tag{21}$$

then

$$n[g(\boldsymbol{X}_n) - g(\boldsymbol{\mu})] \xrightarrow{\text{d}} \frac{1}{2}\boldsymbol{W}'\mathbf{H}_k(\boldsymbol{\mu})\boldsymbol{W}. \tag{22}$$

Proof By (20), we see that $\boldsymbol{X}_n \xrightarrow{\text{p}} \boldsymbol{\mu}$. Therefore, the Taylor expansion suggests the approximation

$$n[g(\boldsymbol{X}_n) - g(\boldsymbol{\mu})] = n[\nabla g(\boldsymbol{\mu})]'(\boldsymbol{X}_n - \boldsymbol{\mu}) + \frac{1}{2}[\sqrt{n}(\boldsymbol{X}_n - \boldsymbol{\mu})]'\mathbf{H}_k(\boldsymbol{\mu})[\sqrt{n}(\boldsymbol{X}_n - \boldsymbol{\mu})] + o_p(1)$$

and, by (21), the rhs of the above equals to

$$\frac{1}{2}[\sqrt{n}(\boldsymbol{X}_n-\boldsymbol{\mu})]'\mathbf{H}_k(\boldsymbol{\mu})[\sqrt{n}(\boldsymbol{X}_n-\boldsymbol{\mu})]+o_p(1).$$

By Slutsky's Theorem and in view of (20), we conclude that the above quantity tends in distribution to $\frac{1}{2}\boldsymbol{W}'\mathbf{H}_k(\boldsymbol{\mu})\boldsymbol{W}$. □

Lemma 5 immediately applies to $M_{k,n}$ whenever the random sample arizes from a singular distribution. This result is stated in the following theorem; for the proof see Appendix A.

Theorem 3 *If $M_{k,n}$ is the sample central moment of a singular distribution of order $k\geq 2$, then*

$$n(M_{k,n}-\mu_k)\xrightarrow{\text{d}}\frac{1}{2}k(k-1)\mu_{k-2}W_1^2-kW_1W_{k-1}, \tag{23}$$

where

$$\begin{pmatrix}W_1\\W_{k-1}\end{pmatrix}\sim N_2\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}\sigma^2 & \mu_k\\ \mu_k & \mu_{2k-2}-\mu_{k-1}^2\end{pmatrix}\right). \tag{24}$$

The limiting distribution in (23) can be expressed in terms of two independent and identically distributed standard normal random variables, Z_1, Z_2. Indeed, observing that $\sigma^2(\mu_{2k-2}-\mu_{k-1}^2)-\mu_k^2=\mathsf{Var}[\sigma(X-\mu)^{k-1}-\mu_k(X-\mu)/\sigma]\geq 0$, it is easily seen that

$$\begin{pmatrix}W_1\\W_{k-1}\end{pmatrix}\stackrel{\text{d}}{=}\begin{pmatrix}\sigma Z_1\\ \frac{\mu_k}{\sigma}Z_1+\frac{\gamma_k}{\sigma}Z_2\end{pmatrix},\quad\text{where}\quad \gamma_k=\sqrt{\sigma^2(\mu_{2k-2}-\mu_{k-1}^2)-\mu_k^2}.$$

Therefore, (23) can be rewritten as

$$n(M_{k,n}-\mu_k)\xrightarrow{\text{d}}\left(\frac{1}{2}k(k-1)\sigma^2\mu_{k-2}-k\mu_k\right)Z_1^2-k\gamma_kZ_1Z_2. \tag{25}$$

In order to obtain a further simplification, we shall make use of the following proposition.

Proposition 4 *If Z_1, Z_2 are independent and identically distributed standard normal random variables, then, for arbitrary constants $\alpha,\beta\in\mathbb{R}$,*

$$\alpha Z_1^2+\beta Z_1Z_2\stackrel{\text{d}}{=}\frac{1}{2}\left(\sqrt{\alpha^2+\beta^2}+\alpha\right)Z_1^2-\frac{1}{2}\left(\sqrt{\alpha^2+\beta^2}-\alpha\right)Z_2^2. \tag{26}$$

Proof The assertion is obvious if $\beta=0$. Assume that $\beta\neq 0$ and set $\rho=\sqrt{\alpha^2+\beta^2}>0$. It is easily seen that the moment generating function of the rhs of (26) is given by

$$M_2(t)=\frac{1}{\sqrt{1-2\alpha t-\beta^2t^2}}$$

and it is finite in the interval $\{t\in\mathbb{R}\colon 1-2\alpha t-\beta^2t^2>0\}=(-(\rho-\alpha)^{-1},(\rho+\alpha)^{-1})$, which contains the origin because $\rho+\alpha>0$ and $\rho-\alpha>0$. Also, the moment generating function of the lhs of (26) is

$$M_1(t)=\mathbb{E}\left[\exp\left(\alpha tZ_1^2+\beta tZ_1Z_2\right)\right]=\frac{1}{2\pi}\iint_{\mathbb{R}^2}e^{-\frac{1}{2}\gamma(x,y)}\,\mathrm{d}y\,\mathrm{d}x,$$

where

$$\gamma(x,y) = x^2 + y^2 - 2\alpha t x^2 - 2\beta t x y = \big(1 - 2\alpha t - \beta^2 t^2\big)x^2 + (y - \beta t x)^2.$$

Therefore, for $t \in (-(\rho-\alpha)^{-1}, (\rho+\alpha)^{-1})$,

$$\begin{aligned} M_1(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}(1-2\alpha t-\beta^2 t^2)} \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(y-\beta t x)^2}\, \mathrm{d}y \right) \mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}(1-2\alpha t-\beta^2 t^2)}\, \mathrm{d}x = M_2(t), \end{aligned}$$

and the proof is complete. □

Corollary 4 *If $(X_1,X_2)'$ follows a bivariate normal distribution with $\mathbb{E}(X_1) = \mu_1$, $\mathbb{E}(X_2) = \mu_2$, $\mathsf{Var}(X_1) = \sigma_1^2$, $\mathsf{Var}(X_2) = \sigma_2^2$ and $\mathsf{Cov}(X_1,X_2) = \rho\sigma_1\sigma_2$, where $\mu_1, \mu_2 \in \mathbb{R}$ and $\sigma_1 \geq 0$, $\sigma_2 \geq 0$ and $-1 \leq \rho \leq 1$ are arbitrary constants, then*

$$(X_1-\mu_1)(X_2-\mu_2) \stackrel{\mathrm{d}}{=} \sigma_1\sigma_2 \left[\frac{1}{2}(1+\rho)Z_1^2 - \frac{1}{2}(1-\rho)Z_2^2 \right].$$

Proof Since $(X_1-\mu_1, X_2-\mu_2)' \stackrel{\mathrm{d}}{=} (\sigma_1 Z_1, \sigma_2(\rho Z_1 + \sqrt{1-\rho^2} Z_2))'$, we have that $(X_1 - \mu_1)(X_2-\mu_2) \stackrel{\mathrm{d}}{=} \sigma_1\sigma_2(\rho Z_1^2 + \sqrt{1-\rho^2} Z_1 Z_2)$, and the assertion follows from (26) with $\alpha = \rho$ and $\beta = \sqrt{1-\rho^2}$. □

The main result is contained in the following theorem; its proof, being an immediate consequence of (25) and Proposition 4, is omitted.

Theorem 4 *If $M_{k,n}$ is the sample central moment of a singular distribution of order $k \geq 2$, then*

$$n(\mu_k - M_{k,n}) \xrightarrow{\mathrm{d}} \frac{k}{2}(\sigma\sqrt{\theta_k} + \alpha_k)Z_1^2 - \frac{k}{2}(\sigma\sqrt{\theta_k} - \alpha_k)Z_2^2,$$

where Z_1, Z_2 are independent and identically distributed standard normal random variables and

$$\begin{aligned} \alpha_k &= \mu_k - \frac{1}{2}(k-1)\sigma^2\mu_{k-2}, \\ \theta_k &= \mu_{2k-2} - \mu_{k-1}^2 - (k-1)\mu_{k-2}\left[\mu_k - \frac{1}{4}(k-1)\sigma^2\mu_{k-2}\right]. \end{aligned} \tag{27}$$

Corollary 5 *If $M_{k,n}$ is the sample central moment of a singular distribution of order $k \geq 2$, then there exists a constant $\lambda_k \in \mathbb{R}$ such that*

$$n(\mu_k - M_{k,n}) \xrightarrow{\mathrm{d}} \lambda_k \chi_1^2$$

if and only if

$$\mu_k^2 = \sigma^2\big(\mu_{2k-2} - \mu_{k-1}^2\big). \tag{28}$$

If (28) *holds,* $\lambda_k = k(\mu_k - (1/2)(k-1)\sigma^2\mu_{k-2})$ *and, thus,*

$$n(\mu_k - M_{k,n}) \xrightarrow{\text{d}} k\left[\mu_k - \frac{1}{2}(k-1)\sigma^2\mu_{k-2}\right]\chi_1^2. \tag{29}$$

If (28) *does not hold,*

$$n(\mu_k - M_{k,n}) \xrightarrow{\text{d}} \lambda_k\chi_1^2 - \tilde{\lambda}_k\tilde{\chi}_1^2 \tag{30}$$

with $\lambda_k = (k/2)(\sigma\sqrt{\theta_k} + \alpha_k) > 0$, $\tilde{\lambda}_k = (k/2)(\sigma\sqrt{\theta_k} - \alpha_k) > 0$, α_k *and* θ_k *as in* (27)*, and where* χ_1^2 *and* $\tilde{\chi}_1^2$ *are independent and identically distributed random variables from the chi-square distribution with one degree of freedom.*

After some algebra it follows that (28) is satisfied by all two-valued distributions of Corollaries 2 and 3. In particular, from (29) we can show that

$$n(\mu_k - M_{k,n}) \xrightarrow{\text{d}} \frac{k(k-1)}{2}p_k^{k-1}\frac{(k+1)p_k^2 - (k+1)p_k + 1}{(k+1)p_k - 1}\chi_1^2, \quad k = 3,4,\ldots\ .$$

For example, the two-valued standardized distribution of Corollary 2 with $p_6 = 1/2 + \sqrt{15(4\sqrt{10}-5)}/30$ has sixth central moment equal to $\mu_6 = (50 - 13\sqrt{10})/45$ and

$$n\left(\frac{4\sqrt{10}-5}{135} - M_{6,n}\right) \xrightarrow{\text{d}} \frac{50 - 13\sqrt{10}}{45}\chi_1^2.$$

Finally, for the symmetric distributions of Lemma 3 one finds that (28) is not satisfied and that $\lambda_k = \tilde{\lambda}_k = (\sqrt{k-1}/2)k^{(k-1)/2}$. Hence, since $\mu_k = 0$, we conclude from (30) the limit

$$nM_{k,n} \xrightarrow{\text{d}} \frac{\sqrt{k-1}}{2}k^{(k-1)/2}(\chi_1^2 - \tilde{\chi}_1^2), \quad k = 3,5,7,\ldots\ .$$

Appendix A Proofs

We shall make use of the following Lemmas. For the proof of Lemma 6 see, e.g., Gut (1988, p. 18); for more general results, see Afendras and Markatou (2016).

Lemma 6 *If* $X, X_1, \ldots, X_n$ *are independent and identically distributed with* $\mathbb{E}(X) = \mu$, $\mathsf{Var}(X) = \sigma^2$ *and* $\mathbb{E}|X|^\delta < \infty$ *for some* $\delta \geq 2$*, then, for any* $\alpha \in (0, \delta]$,

$$\mathbb{E}|\sqrt{n}(\bar{X}_n - \mu)|^\alpha \to \sigma^\alpha\,\mathbb{E}|Z|^\alpha,$$

where $Z \sim N(0,1)$ *and* $\bar{X}_n = (X_1 + \cdots + X_n)/n$.

Lemma 7 *If* $X, X_1, \ldots, X_n$ *are independent and identically distributed with* $\mathbb{E}(X) = \mu$ *and* $\mathbb{E}|X|^\nu < \infty$ *for some* $\nu \in \{2,3,\ldots\}$*, then, for any* $j \in \{2,\ldots,\nu\}$,

$$\mathbb{E}|m_{j,n}|^{\nu/j} \leq \mathbb{E}|X - \mu|^\nu, \tag{31}$$

where $m_{j,n} = n^{-1}\sum_{i=1}^n (X_i - \mu)^j$.

Proof If $j = \nu$, then (31) follows by taking expectations to the obvious inequality $|m_{j,n}| \le \frac{1}{n}\sum_{i=1}^{n}|X_i-\mu|^j = \frac{1}{n}\sum_{i=1}^{n}|X_i-\mu|^\nu$. If $j < \nu$ (and thus, $\nu \ge 3$), we apply the inequality

$$\left|\sum_{i=1}^{n} x_i\right|^p \le \left(\sum_{i=1}^{n}|x_i|\right)^p \le n^{p-1}\sum_{i=1}^{n}|x_i|^p, \quad p>1,$$

(the last inequality is a by-product of Hölder's inequality) for $p = \nu/j$ and $x_i = (X_i-\mu)^j$. Then, we have

$$\begin{aligned}\mathbb{E}|m_{j,n}|^{\nu/j} &= \frac{1}{n^{\nu/j}}\,\mathbb{E}\left|\sum_{i=1}^{n}(X_i-\mu)^j\right|^{\nu/j} \le \frac{1}{n^{\nu/j}}\,\mathbb{E}\left(\sum_{i=1}^{n}|X_i-\mu|^j\right)^{\nu/j}\\ &\le \frac{1}{n^{\nu/j}}\,\mathbb{E}\left(n^{\nu/j-1}\sum_{i=1}^{n}|X_i-\mu|^\nu\right) = \frac{1}{n}\,\mathbb{E}\left(\sum_{i=1}^{n}|X_i-\mu|^\nu\right) = \mathbb{E}|X-\mu|^\nu. \qquad \square\end{aligned}$$

Proof of Proposition 2 (a) Observe that the statement in Proposition 2(a) is equivalent to

$$\mathbb{E}[\sqrt{n}(M_{k,n}-\mu_k)] \to 0. \tag{32}$$

Writing

$$M_{k,n}-\mu_k = (m_{k,n}-\mu_k) + (-1)^{k-1}(k-1)m_{1,n}^k + \sum_{j=2}^{k-1}(-1)^{k-j}\binom{k}{j}m_{1,n}^{k-j}m_{j,n}, \tag{33}$$

it suffices to verify that

(i) $\sqrt{n}\,\mathbb{E}(m_{k,n}-\mu_k) = 0$,
(ii) $\sqrt{n}\,\mathbb{E}(m_{1,n}^k) \to 0$,
(iii) $\sqrt{n}\,\mathbb{E}(m_{1,n}^{k-j}m_{j,n}) \to 0$, $j = 2,\ldots,k-2$ (provided $k \ge 4$), and
(iv) $\sqrt{n}\,\mathbb{E}(m_{1,n}m_{k-1,n}) \to 0$ (provided $k \ge 3$).

Now, (i) is obvious (since $\mathbb{E}(m_{k,n}) = \mu_k$), (iv) follows from $\mathbb{E}(m_{1,n}m_{k-1,n}) = \mu_k/n$ and (ii) can be seen by using Lemma 6 with $\alpha = \delta = k$, which shows that

$$\left|n^{k/2}\,\mathbb{E}\left(m_{1,n}^k\right)\right| \le n^{k/2}\,\mathbb{E}|m_{1,n}|^k = \mathbb{E}|\sqrt{n}(\bar{X}_n-\mu)|^k \to \sigma^k\,\mathbb{E}|Z|^k < \infty,$$

and thus, $|\sqrt{n}\,\mathbb{E}(m_{1,n}^k)| \le n^{-(k-1)/2}\,\mathbb{E}|\sqrt{n}(\bar{X}_n-\mu)|^k \to 0$. To show (iii), we assume that $k \ge 4$ and $2 \le j \le k-2$, and we use Hölder's inequality with $p = k/(k-j) > 1$, Lemma 7 with $\nu = k$ and Lemma 6 with $\alpha = \delta = k$ to obtain

$$\begin{aligned}&\left|\sqrt{n}\,\mathbb{E}\left(m_{1,n}^{k-j}m_{j,n}\right)\right|\\ &\le \sqrt{n}\,\mathbb{E}\left(|m_{1,n}|^{k-j}|m_{j,n}|\right) \le \sqrt{n}\left(\mathbb{E}|m_{1,n}|^k\right)^{(k-j)/k}\left(\mathbb{E}|m_{j,n}|^{k/j}\right)^{j/k}\\ &\le \sqrt{n}\left[n^{-k/2}\,\mathbb{E}\,|\sqrt{n}\,(\bar{X}_n-\mu)|^k\right]^{(k-j)/k}\left(\mathbb{E}|X-\mu|^k\right)^{j/k}\\ &= n^{-(k-1-j)/2}\left[\mathbb{E}\,|\sqrt{n}\,(\bar{X}_n-\mu)|^k\right]^{(k-j)/k}\left(\mathbb{E}|X-\mu|^k\right)^{j/k}\\ &= n^{-(k-1-j)/2}O(1) \to 0,\end{aligned}$$

because $\mathbb{E}|\sqrt{n}(\bar{X}_n-\mu)|^k \to \sigma^k\,\mathbb{E}|Z|^k < \infty$.

(b) Observe that the statement in Proposition 2(b) is equivalent to

$$\mathsf{Cov}\left[\sqrt{n}(\bar{X}_n-\mu), \sqrt{n}(M_{k,n}-\mu_k)\right] \to \mu_{k+1} - k\sigma^2\mu_{k-1}, \tag{34}$$

and since $\mathbb{E}(\bar{X}_n-\mu)=0$, it suffices to verify that

$$n\,\mathbb{E}\left[(\bar{X}_n-\mu)(M_{k,n}-\mu_k)\right] = n\,\mathbb{E}[m_{1,n}(M_{k,n}-\mu_k)] \to \mu_{k+1} - k\sigma^2\mu_{k-1}. \tag{35}$$

If $k=2$, then $n\,\mathbb{E}[m_{1,n}(M_{2,n}-\mu_2)] = n\,\mathbb{E}[(\bar{X}_n-\mu)(m_{2,n}-\mu_2)] - n\,\mathbb{E}(\bar{X}_n-\mu)^3 = \mu_3 - n\,\mathbb{E}(\bar{X}_n-\mu)^3$, and it easily seen, by Lemma 6 with $\alpha=\delta=3$, that $|n\,\mathbb{E}(\bar{X}_n-\mu)^3| \le n^{-1/2}\,\mathbb{E}|\sqrt{n}(\bar{X}_n-\mu)|^3 \to 0$; thus, $n\,\mathbb{E}[m_{1,n}(M_{2,n}-\mu_2)] \to \mu_3$. Since $\mu_1=0$, (35) is satisfied for $k=2$.

If $k=3$, $n\,\mathbb{E}[m_{1,n}(M_{3,n}-\mu_3)] = n\,\mathbb{E}[(\bar{X}_n-\mu)(m_{3,n}-\mu_3)] + 2n\,\mathbb{E}(\bar{X}_n-\mu)^4 - 3n\,\mathbb{E}[m_{2,n}(\bar{X}_n-\mu)^2]$, and it is easy to see that $n\,\mathbb{E}[(\bar{X}_n-\mu)(m_{3,n}-\mu_3)] = \mu_4$. Also, by Lemma 6 with $\alpha=\delta=4$, $2n\,\mathbb{E}(\bar{X}_n-\mu)^4 \to 0$. Finally, $-3n\,\mathbb{E}[m_{2,n}(\bar{X}_n-\mu)^2] = -3[\mu_4+(n-1)\mu_2^2]/n \to -3\mu_2^2 = -3\sigma^4$, which verifies (35) for $k=3$.

In the general case when $k\ge 4$, we write $M_{k,n}-\mu_k$ as in (33) and we observe that for (35) to hold it suffices to verify that

(i) $n\,\mathbb{E}[m_{1,n}(m_{k,n}-\mu_k)] = \mu_{k+1}$,

(ii) $n\,\mathbb{E}(m_{1,n}^{k+1}) \to 0$,

(iii) $n\,\mathbb{E}(m_{1,n}^{k+1-j} m_{j,n}) \to 0$, $j=2,\ldots,k-2$, and

(iv) $n\,\mathbb{E}(m_{1,n}^2 m_{k-1,n}) \to \sigma^2\mu_{k-1}$.

Calculating $\mathbb{E}[m_{1,n}(m_{k,n}-\mu_k)] = \mathbb{E}[(\bar{X}_n-\mu)(m_{k,n}-\mu_k)] = \mathbb{E}[(\bar{X}_n-\mu)m_{k,n}] = n^{-2}\sum_{i_1=1}^n\sum_{i_2=1}^n \mathbb{E}[(X_{i_1}-\mu)(X_{i_2}-\mu)^k] = \mu_{k+1}/n$, we conclude (i), while (ii) follows by using Lemma 6 with $\alpha=\delta=k+1$. Also,

$$\begin{aligned} n\,\mathbb{E}\left(m_{1,n}^2 m_{k-1,n}\right) &= \frac{1}{n^2}\sum_{i_1=1}^n\sum_{i_2=1}^n\sum_{i_3=1}^n \mathbb{E}\left[(X_{i_1}-\mu)(X_{i_2}-\mu)\left(X_{i_3}-\mu\right)^{k-1}\right] \\ &= \frac{1}{n^2}\left[n\mu_{k+1} + n(n-1)\sigma^2\mu_{k-1}\right] \to \sigma^2\mu_{k-1}, \end{aligned}$$

which shows that (iv) is satisfied, and it remains to verify (iii). To this end, we use Hölder's inequality with $p=(k+1)/(k+1-j)>1$ and Lemma 7 with $\nu=k+1$ to obtain

$$\begin{aligned} &\left| n\,\mathbb{E}\left(m_{1,n}^{k+1-j} m_{j,n}\right)\right| \\ &\le n\,\mathbb{E}|m_{1,n}|^{k+1-j}|m_{j,n}| \le n\left(\mathbb{E}|m_{1,n}|^{k+1}\right)^{(k+1-j)/(k+1)}\left(\mathbb{E}|m_{j,n}|^{(k+1)/j}\right)^{j/(k+1)} \\ &\le n\left[n^{-(k+1)/2}\,\mathbb{E}\left|\sqrt{n}\left(\bar{X}_n-\mu\right)\right|^{k+1}\right]^{(k+1-j)/(k+1)}\left(\mathbb{E}|X-\mu|^{k+1}\right)^{j/(k+1)} \\ &= n^{-(k-1-j)/2}\left[\mathbb{E}\left|\sqrt{n}\left(\bar{X}_n-\mu\right)\right|^{k+1}\right]^{(k+1-j)/(k+1)}\left(\mathbb{E}|X-\mu|^{k+1}\right)^{j/(k+1)} \to 0, \end{aligned}$$

because $n^{-(k-1-j)/2}\to 0$; and, by Lemma 6 with $\alpha=\delta=k+1$, $\mathbb{E}|\sqrt{n}(\bar{X}_n-\mu)|^{k+1} \to \sigma^{k+1}\,\mathbb{E}|Z|^{k+1} < \infty$.

(c) Without loss of generality assume that $2 \le r \le k$ and observe that the first statement of Proposition 2(c) is equivalent to

$$\mathsf{Cov}[\sqrt{n}(M_{r,n}-\mu_r),\sqrt{n}(M_{k,n}-\mu_k)] \to v_{rk}. \tag{36}$$

Since $\mathbb{E}|X|^{r+k} < \infty$, (32) shows that $\mathbb{E}[\sqrt{n}(M_{k,n}-\mu_k)] \to 0$ and $\mathbb{E}[\sqrt{n}(\bar{X}_n-\mu)] \to 0$, and it suffices to verify that

$$\begin{aligned} n\,\mathbb{E}[(M_{r,n}-\mu_r)(M_{k,n}-\mu_k)] \to v_{rk} = \mu_{r+k} - \mu_r\mu_k - r\mu_{r-1}\mu_{k+1} \\ - k\mu_{r+1}\mu_{k-1} + rk\sigma^2\mu_{r-1}\mu_{k-1}. \end{aligned} \tag{37}$$

The proof can be deduced by showing that (37) holds for each one of the cases $r=k=2$; $r=2$, $k=3$; $r=k=3$; $r=2$, $k\ge 4$; $r=3$, $k\ge 4$; $4\le r\le k$. In the following we shall present the details only for the case where $4\le r\le k$; the other cases can be treated using similar (and simpler) arguments.

Assume now that $4\le r\le k$. From (33), we have

$$\begin{aligned} M_{r,n}-\mu_r = (m_{r,n}-\mu_r) - r m_{1,n}m_{r-1,n} \\ + \sum_{j_1=2}^{r-2} (-1)^{r-j_1}\binom{r}{j_1} m_{1,n}^{r-j_1} m_{j_1,n} + (-1)^{r-1}(r-1)m_{1,n}^r, \end{aligned} \tag{38}$$

$$\begin{aligned} M_{k,n}-\mu_k = (m_{k,n}-\mu_k) - k m_{1,n}m_{k-1,n} \\ + \sum_{j_2=2}^{k-2} (-1)^{k-j_2}\binom{k}{j_2} m_{1,n}^{k-j_2} m_{j_2,n} + (-1)^{k-1}(k-1)m_{1,n}^k. \end{aligned} \tag{39}$$

We shall show that the asymptotic covariance in (36) can be determined by using only the first two terms in (38) and (39). Indeed, it is easily seen that (37) holds true if it can be shown that

(i) $n\,\mathbb{E}[(m_{r,n}-\mu_r)(m_{k,n}-\mu_k)] = \mu_{r+k}-\mu_r\mu_k$,
(ii) $n\,\mathbb{E}[m_{1,n}m_{k-1,n}(m_{r,n}-\mu_r)] \to \mu_{r+1}\mu_{k-1}$,
(iii) $n\,\mathbb{E}[m_{1,n}m_{r-1,n}(m_{k,n}-\mu_k)] \to \mu_{r-1}\mu_{k+1}$,
(iv) $n\,\mathbb{E}(m_{1,n}^2 m_{r-1,n}m_{k-1,n}) \to \sigma^2\mu_{r-1}\mu_{k-1}$,
(v) $n\,\mathbb{E}[m_{1,n}^{k-j_2}m_{j_2,n}(m_{r,n}-\mu_r)] \to 0$, $j_2=2,\ldots,k-2$,
(vi) $n\,\mathbb{E}[m_{1,n}^k(m_{r,n}-\mu_r)] \to 0$,
(vii) $n\,\mathbb{E}(m_{1,n}^{k+1-j_2}m_{j_2,n}m_{r-1,n}) \to 0$, $j_2=2,\ldots,k-2$,
(viii) $n\,\mathbb{E}(m_{1,n}^{k+1}m_{r-1,n}) \to 0$,
(ix) $n\,\mathbb{E}[m_{1,n}^{r-j_1}m_{j_1,n}(m_{k,n}-\mu_k)] \to 0$, $j_1=2,\ldots,r-2$,
(x) $n\,\mathbb{E}(m_{1,n}^{r+1-j_1}m_{j_1,n}m_{k-1,n}) \to 0$, $j_1=2,\ldots,r-2$,
(xi) $n\,\mathbb{E}(m_{1,n}^{r+k-j_1-j_2}m_{j_1,n}m_{j_2,n}) \to 0$, $j_1=2,\ldots,r-2$, $j_2=2,\ldots,k-2$,
(xii) $n\,\mathbb{E}(m_{1,n}^{r+k-j_1}m_{j_1,n}) \to 0$, $j_1=2,\ldots,r-2$,

(xiii) $n\,\mathbb{E}[m_{1,n}^{r}(m_{k,n}-\mu_k)]\to 0$,
(xiv) $n\,\mathbb{E}(m_{1,n}^{r+1}m_{k-1,n})\to 0$,
(xv) $n\,\mathbb{E}(m_{1,n}^{r+k-j_2}m_{j_2,n})\to 0,\ j_2=2,\ldots,k-2$, and
(xvi) $n\,\mathbb{E}(m_{1,n}^{r+k})\to 0$.

We now proceed to verify (i)–(xvi). Since $\mathbb{E}(m_{r,n})=\mu_r$ and $\mathbb{E}(m_{k,n})=\mu_k$, we have

$$\begin{aligned} &n\,\mathbb{E}[(m_{r,n}-\mu_r)(m_{k,n}-\mu_k)] \\ &\quad = n[\mathbb{E}(m_{r,n}m_{k,n})-\mu_r\mu_k] = n\left\{\frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\mathbb{E}\left[(X_{i_1}-\mu)^r(X_{i_2}-\mu)^k\right]-\mu_r\mu_k\right\} \\ &\quad = n\left\{\frac{1}{n^2}[n\mu_{r+k}+n(n-1)\mu_r\mu_k]-\mu_r\mu_k\right\} = \mu_{r+k}-\mu_r\mu_k, \end{aligned}$$

which shows (i). Also, (ii), (iii) and (iv) follow by straightforward computations; e.g., for (ii) we have

$$\begin{aligned} n\,\mathbb{E}[m_{1,n}m_{k-1,n}(m_{r,n}-\mu_r)] &= -\mu_r\mu_k+\frac{\mu_{r+k}+(n-1)(\mu_{r+1}\mu_{k-1}+\mu_r\mu_k)}{n} \\ &\to \mu_{r+1}\mu_{k-1}, \end{aligned}$$

while (iii) is similar to (ii), and (iv) can be deduced from

$$n\,\mathbb{E}\big(m_{1,n}^2m_{r-1,n}m_{k-1,n}\big)=\frac{1}{n^3}\big[n(n-1)(n-2)\sigma^2\mu_{r-1}\mu_{k-1}+o\big(n^3\big)\big]\to\sigma^2\mu_{r-1}\mu_{k-1}.$$

The vanishing limits (vi)–(viii) and (x)–(xvi) are by-products of Lemmas 6 and 7 with $\alpha=\delta=\nu=r+k$, since $\mathbb{E}|X|^{r+k}<\infty$. Indeed, we have $|n\,\mathbb{E}(m_{1,n}^{r+k})|\le n\,\mathbb{E}|m_{1,n}|^{r+k}=n^{-(r+k-2)/2}\,\mathbb{E}|\sqrt{n}(\bar{X}_n-\mu)|^{r+k}\to 0$, which verifies (xvi). Also, using Hölder's inequality with $p=(r+k)/(r+k-j_2)>1$, we obtain (xv) as follows:

$$\begin{aligned} &\left|n\,\mathbb{E}\left(m_{1,n}^{r+k-j_2}m_{j_2,n}\right)\right| \\ &\quad\le n\,\mathbb{E}\left(|m_{1,n}|^{r+k-j_2}|m_{j_2,n}|\right)\le n\left(\mathbb{E}|m_{1,n}|^{r+k}\right)^{\frac{r+k-j_2}{r+k}}\left(\mathbb{E}|m_{j_2,n}|^{\frac{r+k}{j_2}}\right)^{\frac{j_2}{r+k}} \\ &\quad\le n\left[n^{-(r+k)/2}\,\mathbb{E}|\sqrt{n}(\bar{X}_n-\mu)|^{r+k}\right]^{\frac{r+k-j_2}{r+k}}\left(\mathbb{E}|X-\mu|^{r+k}\right)^{\frac{j_2}{r+k}} \\ &\quad= n^{-(r+k-j_2-2)/2}\left[\mathbb{E}|\sqrt{n}(\bar{X}_n-\mu)|^{r+k}\right]^{\frac{r+k-j_2}{r+k}}\left(\mathbb{E}|X-\mu|^{r+k}\right)^{\frac{j_2}{r+k}}\to 0, \end{aligned}$$

because $n^{-(r+k-j_2-2)/2}\to 0$ and $\mathbb{E}|\sqrt{n}(\bar{X}_n-\mu)|^{r+k}\to\sigma^{r+k}\,\mathbb{E}|Z|^{r+k}<\infty$; (xii) is similar to (xv). For the limit (xiv) we have

$$\begin{aligned} &\left|n\,\mathbb{E}\left[m_{1,n}^{r+1}m_{k-1,n}\right]\right| \\ &\quad\le n\,\mathbb{E}\left(|m_{1,n}|^{r+1}|m_{k-1,n}|\right)\le n\left(\mathbb{E}|m_{1,n}|^{r+k}\right)^{\frac{r+1}{r+k}}\left(\mathbb{E}|m_{k-1,n}|^{\frac{r+k}{k-1}}\right)^{\frac{k-1}{r+k}} \\ &\quad\le n^{-(r-1)/2}\left[\mathbb{E}|\sqrt{n}(\bar{X}_n-\mu)|^{r+k}\right]^{\frac{r+1}{r+k}}\left(\mathbb{E}|X-\mu|^{r+k}\right)^{\frac{k-1}{r+k}}\to 0, \end{aligned}$$

and similarly for (viii). In order to prove (xiii), it is sufficient to show that $n\,\mathbb{E}[m_{1,n}^r m_{k,n}] \to 0$ and $n\,\mathbb{E}(m_{1,n}^r) \to 0$. The second limit is obvious since, as for (xvi), one can easily verify that $|n\,\mathbb{E}(m_{1,n}^r)| \le n^{-(r-2)/2}\,\mathbb{E}|\sqrt{n}(\bar{X}_n-\mu)|^r = n^{-(r-2)/2}O(1) \to 0$. For the first limit, we have

$$\begin{aligned}\left|n\,\mathbb{E}\left(m_{1,n}^r m_{k,n}\right)\right| &\le n\,\mathbb{E}\left(|m_{1,n}|^r|m_{k,n}|\right) \le n\left(\mathbb{E}|m_{1,n}|^{r+k}\right)^{\frac{r}{r+k}}\left(\mathbb{E}|m_{k,n}|^{\frac{r+k}{k}}\right)^{\frac{k}{r+k}}\\ &\le n^{-(r-2)/2}\left[\mathbb{E}\left|\sqrt{n}\,(\bar{X}_n-\mu)\right|^{r+k}\right]^{\frac{r}{r+k}}\left(\mathbb{E}|X-\mu|^{r+k}\right)^{\frac{k}{r+k}} \to 0.\end{aligned}$$

Limit (vi) is similar to (xiii) and its proof is omitted. Regarding (xi), we have

$$\begin{aligned}&\left|n\,\mathbb{E}\left(m_{1,n}^{r+k-j_1-j_2}m_{j_1,n}m_{j_2,n}\right)\right|\\ &\le n\,\mathbb{E}\left(|m_{1,n}|^{r+k-j_1-j_2}|m_{j_1,n}m_{j_2,n}|\right)\\ &\le n\left(\mathbb{E}|m_{1,n}|^{r+k}\right)^{\frac{r+k-j_1-j_2}{r+k}}\left(\mathbb{E}|m_{j_1,n}m_{j_2,n}|^{\frac{r+k}{j_1+j_2}}\right)^{\frac{j_1+j_2}{r+k}}\\ &\le n\left(\mathbb{E}|m_{1,n}|^{r+k}\right)^{\frac{r+k-j_1-j_2}{r+k}}\left[\left(\mathbb{E}|m_{j_1,n}|^{\frac{r+k}{j_1}}\right)^{\frac{j_1}{j_1+j_2}}\left(\mathbb{E}|m_{j_2,n}|^{\frac{r+k}{j_2}}\right)^{\frac{j_2}{j_1+j_2}}\right]^{\frac{j_1+j_2}{r+k}}\\ &\le n\left(\mathbb{E}|m_{1,n}|^{r+k}\right)^{\frac{r+k-j_1-j_2}{r+k}}\left(\mathbb{E}|X-\mu|^{r+k}\right)^{\frac{j_1+j_2}{r+k}}\\ &= n^{-(r+k-j_1-j_2-2)/2}\left[\mathbb{E}\left|\sqrt{n}\,(\bar{X}_n-\mu)\right|^{r+k}\right]^{\frac{r+k-j_1-j_2}{r+k}}\left(\mathbb{E}|X-\mu|^{r+k}\right)^{\frac{j_1+j_2}{r+k}} \to 0.\end{aligned}$$

Similarly, for (x) we have

$$\begin{aligned}&\left|n\,\mathbb{E}\left(m_{1,n}^{r+1-j_1}m_{j_1,n}m_{k-1,n}\right)\right|\\ &\le n\left(\mathbb{E}|m_{1,n}|^{r+1-j_1}|m_{j_1,n}m_{k-1,n}|\right)\\ &\le n\left(\mathbb{E}|m_{1,n}|^{r+k}\right)^{\frac{r+1-j_1}{r+k}}\left(\mathbb{E}|m_{j_1,n}m_{k-1,n}|^{\frac{r+k}{j_1+k-1}}\right)^{\frac{j_1+k-1}{r+k}}\\ &\le n\left(\mathbb{E}|m_{1,n}|^{r+k}\right)^{\frac{r+1-j_1}{r+k}}\left[\left(\mathbb{E}|m_{j_1,n}|^{\frac{r+k}{j_1}}\right)^{\frac{j_1}{j_1+k-1}}\left(\mathbb{E}|m_{k-1,n}|^{\frac{r+k}{k-1}}\right)^{\frac{k-1}{j_1+k-1}}\right]^{\frac{j_1+k-1}{r+k}}\\ &\le n\left(\mathbb{E}|m_{1,n}|^{r+k}\right)^{\frac{r+1-j_1}{r+k}}\left(\mathbb{E}|X-\mu|^{r+k}\right)^{\frac{j_1+k-1}{r+k}}\\ &= n^{-(r-j_1-1)/2}\left[\mathbb{E}\left|\sqrt{n}\,(\bar{X}_n-\mu)\right|^{r+k}\right]^{\frac{r+1-j_1}{r+k}}\left(\mathbb{E}|X-\mu|^{r+k}\right)^{\frac{j_1+k-1}{r+k}} \to 0,\end{aligned}$$

while (vii) is similar to (x).

It remains to verify (v) and (ix); but, since they are similar, it suffices to prove (v). If $j_2 \in \{2,\dots,k-3\}$ (and hence, $k \geq 5$ and $j_2 < k-2$), we have

$$\begin{aligned}&\left| n\,\mathbb{E}\left[m_{1,n}^{k-j_2} m_{j_2,n}(m_{r,n}-\mu_r)\right]\right| \\ &\quad \leq n\,\mathbb{E}\left(|m_{1,n}|^{k-j_2}|m_{j_2,n}m_{r,n}|\right) + n|\mu_r|\mathbb{E}\left(|m_{1,n}|^{k-j_2}|m_{j_2,n}|\right),\end{aligned}$$

and it suffices to prove that $n\,\mathbb{E}(|m_{1,n}|^{k-j_2}|m_{j_2,n}m_{r,n}|) \to 0$ and $n\,\mathbb{E}(|m_{1,n}|^{k-j_2}|m_{j_2,n}|) \to 0$. For the first quantity, we have

$$\begin{aligned}&n\,\mathbb{E}\left(|m_{1,n}|^{k-j_2}\left|m_{j_2,n}m_{r,n}\right|\right)\\ &\leq n\left(\mathbb{E}|m_{1,n}|^{r+k}\right)^{\frac{k-j_2}{r+k}}\left(\mathbb{E}|m_{j_2,n}m_{r,n}|^{\frac{r+k}{r+j_2}}\right)^{\frac{r+j_2}{r+k}}\\ &\leq n\left(\mathbb{E}|m_{1,n}|^{r+k}\right)^{\frac{k-j_2}{r+k}}\left[\left(\mathbb{E}|m_{j_2,n}|^{\frac{r+k}{j_2}}\right)^{\frac{j_2}{r+j_2}}\left(\mathbb{E}|m_{r,n}|^{\frac{r+k}{r}}\right)^{\frac{r}{r+j_2}}\right]^{\frac{r+j_2}{r+k}}\\ &\leq n^{-(k-j_2-2)/2}\left[\mathbb{E}\left|\sqrt{n}\,(\bar{X}_n-\mu)\right|^{r+k}\right]^{\frac{k-j_2}{r+k}}\left(\mathbb{E}|X-\mu|^{r+k}\right)^{\frac{r+j_2}{r+k}} \to 0,\end{aligned}$$

because $k-j_2-2>0$. Similarly, for the second quantity we have

$$\begin{aligned}&n\,\mathbb{E}\left(|m_{1,n}|^{k-j_2}\left|m_{j_2,n}\right|\right)\\ &\leq n\left(\mathbb{E}|m_{1,n}|^{r+k}\right)^{\frac{k-j_2}{r+k}}\left(\mathbb{E}|m_{j_2,n}|^{\frac{r+k}{r+j_2}}\right)^{\frac{r+j_2}{r+k}}\\ &\leq n\left(\mathbb{E}|m_{1,n}|^{r+k}\right)^{\frac{k-j_2}{r+k}}\left(\mathbb{E}|m_{j_2,n}|^{\frac{r+k}{j_2}}\right)^{\frac{j_2}{r+k}}\\ &\leq n^{-(k-j_2-2)/2}\left[\mathbb{E}\left|\sqrt{n}\,(\bar{X}_n-\mu)\right|^{r+k}\right]^{\frac{k-j_2}{r+k}}\left(\mathbb{E}|X-\mu|^{r+k}\right)^{\frac{j_2}{r+k}} \to 0,\end{aligned}$$

because $k-j_2-2>0$. Finally, it remains to study the limit (v) when $j_2=k-2$; in this case the above limits do not necessarily vanish. However, since $j_2=k-2$ we have

$$n\,\mathbb{E}\left[m_{1,n}^{k-j_2}m_{j_2,n}(m_{r,n}-\mu_r)\right] = n\,\mathbb{E}\big(m_{1,n}^2 m_{r,n}m_{k-2,n}\big) - n\mu_r\,\mathbb{E}\big(m_{1,n}^2 m_{k-2,n}\big),$$

and direct computations show that

$$n\,\mathbb{E}\big(m_{1,n}^2 m_{r,n}m_{k-2,n}\big) = \frac{1}{n^3}\left[n(n-1)(n-2)\sigma^2\mu_r\mu_{k-2}+o\big(n^3\big)\right] \to \sigma^2\mu_r\mu_{k-2}$$

and

$$n\,\mathbb{E}\big(m_{1,n}^2 m_{k-2,n}\big) = \frac{1}{n^2}\left[n(n-1)\sigma^2\mu_{k-2}+o\big(n^2\big)\right] \to \sigma^2\mu_{k-2}.$$

Hence, when $j_2 = k-2$ we have

$$\begin{aligned} &n\,\mathbb{E}\left[m_{1,n}^{k-j_2} m_{j_2,n}(m_{r,n}-\mu_r)\right] \\ &\quad = n\,\mathbb{E}\left(m_{1,n}^2 m_{r,n} m_{k-2,n}\right) - n\mu_r\,\mathbb{E}\left(m_{1,n}^2 m_{k-2,n}\right) \to \sigma^2\mu_r\mu_{k-2} - \mu_r\sigma^2\mu_{k-2} = 0, \end{aligned}$$

and the proof is complete. □

Proof of Theorem 3 Observe that $M_{k,n} - \mu_k = g_{k,k}(\boldsymbol{m}_{k,n}) - g_{k,k}(\boldsymbol{\mu}_k)$; see in Section 2. Also, $\sqrt{n}(\boldsymbol{m}_n - \boldsymbol{\mu}_k) \xrightarrow{\text{d}} \boldsymbol{W}_k$, where $\boldsymbol{W}_k = (W_1,\ldots,W_k)' \sim N(\mathbf{0}_k, \boldsymbol{\Sigma}_k)$, see (1). Hence, Lemma 5 applies to $\boldsymbol{X}_n = \boldsymbol{m}_{k,n}$, provided (21) is fulfilled for $\boldsymbol{m}_{k,n}$, i.e., provided that $n[\nabla g_{k,k}(\boldsymbol{\mu}_k)]'(\boldsymbol{m}_{k,n} - \boldsymbol{\mu}_k) \xrightarrow{\text{p}} 0$. Because $\nabla g_{k,k}(\boldsymbol{\mu}_k) = (-k\mu_{k-1}, 0, \ldots, 0, 1)'$, we get $[\nabla g_{k,k}(\boldsymbol{\mu}_k)]'(\boldsymbol{m}_{k,n} - \boldsymbol{\mu}_k) = -k\mu_{k-1}m_{1,n} + (m_{k,n} - \mu_k)$. Since $\mathbb{E}(m_{j,n}) = \mu_j$ for all n and j we get $\mathbb{E}[-k\mu_{k-1}m_{1,n} + (m_{k,n} - \mu_k)] = 0$. Also,

$$\begin{aligned} &\mathsf{Var}[-k\mu_{k-1}m_{1,n} + (m_{k,n} - \mu_k)] \\ &\quad = k^2\mu_{k-1}^2\,\mathsf{Var}(m_{1,n}) + \mathsf{Var}(m_{k,n}) - 2k\mu_{k-1}\mathsf{Cov}(m_{1,n}, m_{k,n}) \\ &\quad = k^2\mu_{k-1}^2\frac{\sigma^2}{n} + \frac{\mu_{2k} - \mu_k^2}{n} - 2k\mu_{k-1}\frac{\mu_{k+1}}{n} \\ &\quad = \frac{1}{n}\left(k^2\mu_{k-1}^2\sigma^2 + \mu_{2k} - \mu_k^2 - 2k\mu_{k-1}\mu_{k+1}\right) \\ &\quad = \frac{1}{n}\left[\mu_{2k} - \mu_k^2 + k\mu_{k-1}\left(k\sigma^2\mu_{k-1} - 2\mu_{k+1}\right)\right] = \frac{v_k^2}{n} = 0, \end{aligned}$$

because $v_k^2 = 0$ by the assumed singularness. Therefore, $[\nabla g_{k,k}(\boldsymbol{\mu}_k)]'(\boldsymbol{m}_{k,n} - \boldsymbol{\mu}_k) = 0$ with probability one and, thus, $n[\nabla g_{k,k}(\boldsymbol{\mu}_k)]'(\boldsymbol{m}_{k,n} - \boldsymbol{\mu}_k) \xrightarrow{\text{p}} 0$ in a trivial sense. Now, a simple calculation, since $\nabla g_{k,k}(\boldsymbol{\mu}_k) = (-k\mu_{k-1}, 0, \ldots, 0, 1)'$, shows that

$$\mathbf{H}_k(\boldsymbol{\mu}_k) = \begin{pmatrix} k(k-1)\mu_{k-2} & 0 & \cdots & 0 & -k & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 \\ -k & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 \end{pmatrix},$$

i.e.,

$$\mathbf{H}_2(\boldsymbol{\mu}_2) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}, \quad \mathbf{H}_3(\boldsymbol{\mu}_3) = \begin{pmatrix} 0 & -3 & 0 \\ -3 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{H}_4(\boldsymbol{\mu}_4) = \begin{pmatrix} 12\sigma^2 & 0 & -4 & 0 \\ 0 & 0 & 0 & 0 \\ -4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

e.tc. Applying (22), we see that $n(M_{k,n} - \mu_k)$ converges weakly to the distribution of $\frac{1}{2}\boldsymbol{W}_k'\mathbf{H}_k(\boldsymbol{\mu}_k)\boldsymbol{W}_k = \frac{1}{2}k(k-1)\mu_{k-2}W_1^2 - kW_1W_{k-1}$, while, by (1), the distribution of $(W_1, W_{k-1})'$ is given by (24). □

Acknowledgements We would like to thank H. Papageorgiou for helpful discussions.

References

Afendras, G. (2013). Moment-based inference for Pearson's quadratic q subfamily of distributions. *Comm. Statist. – Theory Methods*, **42**(12), 2271–2280.

Afendras, G.; Markatou, M. (2016). Uniform integrability of the OLS estimators, and the convergence of their moments. *TEST*, **25**(4), 775–784.

Billingsley, P. (1995). *Probability and Measure* (3rd ed.), John Wiley & Sons, New York.

Geary, R.C. (1936). The distribution of "Student's" ratio for non-normal samples. *J. Roy. Statist. Soc.* Ser. B, **3**, 178–184.

Gut, A. (1988). *Stopped Random Walks: Limit Theorems and Applications*. Springer–Verlag, N.Y.

Haug, S; Klüppelberg, C.; Lindner, A.; Zapp, M. (2007). Method of moment estimation in the COGARCH(1,1) model. *Econom. J.*, **10**(2), 320–341.

Kagan, A.M.; Linnik, Y.V.; Rao, C.R. (1973). *Characterization Problems in Mathematical Statistics*. John Wiley, N.Y.

Kourouklis, S. (2012). A new estimator of the variance based on minimizing mean squared error. *The American Statistician*, **66**(4), 234–236.

Laha, R.G.; Lukacs, E.; Newman, M. (1960). On the independence of a sample central moment and the sample mean. *Ann. Math. Statist.*, **31**, 1028–1033.

Lehmann, E.L. (1999). *Elements of Large-Sample Theory*. Springer, N.Y.

Pewsey, A. (2005). The large-sample distribution of the most fundamental of statistical summaries. *J. Statist. Plann. Inference*, **134**, 434–444.

Stefanski, L.A.; Boos, D.D. (2002). The calculus of M-estimation. *Amer. Statist.*, **56**(1), 29–38.

Wu, S.-F.; Liang, M.-C. (2010). A note on the asymptotic distribution of the process capability index C_{pkm}. *J. Statist. Comput. Simul.*, **80**, 227–235.

Yatracos, Y. (2005). Artificially augmented samples, shrinkage, and mean squared error reduction. *J. Amer. Statist. Assoc.*, **100**, 1168–1175.

van der Vaart, A.W. (1998). *Asymptotic Statistics*. Cambridge University Press, N.Y.

Zinger, A.A. (1958). Independence of quasi-polynomial statistics and analytical properties of distributions. *Theory of Probability & Its Applications*, **3**(3), 247-265.