Sankhyā : The Indian Journal of Statistics
2006, Volume 68, Part 2, pp. 326-356

Higher Order Cumulants of Random Vectors and Applications to Statistical Inference and Time Series

S. Rao Jammalamadaka
University of California, Santa Barbara, USA
T. Subba Rao
University of Manchester, Manchester, UK
György Terdik
University of Debrecen, Debrecen, Hungary

Abstract

This paper provides a unified and comprehensive approach for deriving expressions for higher-order cumulants of random vectors. The approach is based on expanding the characteristic functions and cumulant generating functions in terms of the Kronecker products of differential operators. The use of this methodology is then illustrated in three diverse and novel contexts, namely: (i) in obtaining a lower bound (Bhattacharya bound) for the variance-covariance matrix of a vector of unbiased estimators where the density depends on several parameters, (ii) in studying the asymptotic theory of multivariate statistics when the population is not necessarily Gaussian and finally, (iii) in obtaining higher order cumulant spectra in the study of multivariate nonlinear time series models. Our objective here is to derive such expressions for the higher-order cumulants of random vectors using only elementary calculus of several variables and to highlight some important and novel applications in statistics.

AMS (2000) *subject classification.* Primary 62E17, 62E20; secondary 62H10, 60E05.
Keywords and phrases. Cumulants for random vectors, cumulants for likelihood functions, Bhattacharya lower bound, Taylor series expansion, multivariate time series.

1 Introduction and Review

It is well known that cumulants of order greater than two are zero for random variables which are Gaussian. In view of this, higher-order cumulants are often used in testing for Gaussianity as well as to prove classical limit theorems. These are also used in asymptotic theory of statistics, such as in Edgeworth expansions. Consider a scalar random variable X, and let us assume that its moment generating function in a neighbourhood of the origin is finite with moments $\mu_j = \mathsf{E}(X^j),\ j = 1, 2, \ldots$. Let the characteristic function of X be denoted by $\varphi_X(\lambda)$, and then it has the expansion given by

$$\varphi_X(\lambda) = \mathsf{E}(e^{i\lambda X}) = 1 + \sum_{j=1}^{\infty} \mu_j \frac{(i\lambda)^j}{j!},\ \lambda \in \mathbb{R}. \tag{1.1}$$

From (1.1), we observe that $\mu_j = (-i)^j \left[d^j \varphi(\lambda)/d\lambda^j \right]_{\lambda=0}$. In other words, the j^{th} derivative of the Taylor series expansion of $\varphi_X(\lambda)$ evaluated at $\lambda = 0$ gives the j^{th} moment. The "cumulant generating function" $\psi_X(\lambda)$ is defined as (see e.g., Leonov and Shiryaev, 1959)

$$\psi_X(\lambda) = \ln \varphi_X(\lambda) = \sum_{j=1}^{\infty} \kappa_j \frac{(i\lambda)^j}{j!}, \tag{1.2}$$

where κ_j is called as the j^{th} cumulant of the random variable X. As before, we have $\kappa_j = (-i)^j \left[d^j \psi_X(\lambda)/d\lambda^j \right]_{\lambda=0}$. Comparing (1.1) and (1.2), one can write the cumulants in terms of moments and vice versa. For example, $\kappa_1 = \mu_1,\ \kappa_2 = \mu_2 - (\mu_1)^2$ etc.. Now suppose the random variable X is normal with mean μ and variance σ^2. Then we know that $\varphi_X(\lambda) = \exp(i\lambda\mu - \lambda^2\sigma^2/2)$, which implies that $\kappa_j = 0$ for all $j \geq 3$. We now consider generalizing the above results to the case when $\boldsymbol{X}$ is a d-dimensional random vector. The definition of the joint moments and the cumulants of the random vector $\boldsymbol{X}$ requires a Taylor series expansion of a function in several variables and its partial derivatives in these variables, similar to (1.1) and (1.2). Though these expansions may be considered straightforward generalizations, the methodology and the mathematical notation get quite cumbersome when dealing with the derivatives of characteristic functions and the cumulant generating functions of random vectors. However such expansions are essential in studying the asymptotic theory in classical multivariate analysis as well as in multivariate nonlinear time series (see e.g., Subba Rao and Wong, 1999). A unified and streamlined methodology for obtaining such expressions is desirable, and that is what we attempt to do here.

As an example, consider a random sample $(\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n)$ from a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and variance covariance matrix Σ. We know that the sample mean vector $\overline{\boldsymbol{X}}$ has a multivariate normal distribution, the sample variance covariance matrix has a Wishart distribution, and they are independent. However, when the random sample is not from a multivariate normal distribution, one approach for obtaining such distributions is through the multivariate Edgeworth expansion, whose evaluation requires expressions for higher-order cumulants of random vectors. Further applications of these, in the time-series context, can be found in the books of Brillinger (2001), Terdik (1999) and the recent papers of Subba Rao and Wong (1999) and Wong (1997).

Results similar to ours can be found in the works of McCullagh (1987) and Speed (1990), who use Tensor calculus, whereas our methods are simpler requiring only knowledge of calculus of several variables. Also, we believe this to be a more transparent and streamlined approach. Finally, we derive several new results of interest in statistical inference and time series using these methods. We derive Yule-Walker type difference equations in terms of higher-order cumulants for stationary multivariate linear processes. Derived also are the expressions for higher-order cumulant spectra of such processes, which turn out to be useful in constructing statistical tests for linearity and Gaussianity of multivariate time series. The "information inequality" or the Cramer-Rao lower bound for the variance of an unbiased estimator is well known for both single parameter and multiple parameter cases. A more accurate series of bounds for the single parameter case are given by Bhattacharya (1946), and they depend on all higher-order derivatives of the log-likelihood function. Here we give a generalization of this bound for the multiparameter case based on partial derivatives of various orders. We illustrate this with an example, where we find a lower bound for the variance of an unbiased estimator of a nonlinear function of the parameters.

In Section 2, we provide some preparatory materials and define the cumulants of several random vectors in Section 3. In Section 4, we consider various applications of the above methods to statistical inference. These include: (i) properties of the cumulants of the partial derivatives of log-likelihood function of a random sample $(\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n)$ drawn from a distribution $F_{\vartheta}(\boldsymbol{x})$, $\vartheta \in \Omega$, (ii) multivariate measures of skewness and kurtosis, (iii) applications to multiple time series and finally, (iv) Bhattacharya bounds for multiparameter problems based on the partial derivatives and an illustrative example. In the appendix, we derive some properties of commutation matrices and discuss Taylor series expansion in terms of differential operators.

2 Preparatory Materials

2.1. Differential operators. First we introduce the Jacobian matrix and higher-order derivatives. Let $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_d)' \in \mathbb{R}^d$, and let $\boldsymbol{\phi}(\boldsymbol{\lambda}) = [\phi_1(\boldsymbol{\lambda}), \phi_2(\boldsymbol{\lambda}), \ldots, \phi_m(\boldsymbol{\lambda})]'$ be a vector valued function, which is differentiable in all its arguments (here and elsewhere $'$ denotes the transpose). The Jacobian matrix of $\boldsymbol{\phi}$ is defined by

$$D_{\boldsymbol{\lambda}}\boldsymbol{\phi} = \frac{\partial \boldsymbol{\phi}}{\partial \boldsymbol{\lambda}'} = \boldsymbol{\phi}(\boldsymbol{\lambda})\left[\frac{\partial}{\partial \lambda_1}, \frac{\partial}{\partial \lambda_2}, \ldots, \frac{\partial}{\partial \lambda_d}\right] = \begin{bmatrix} \frac{\partial \phi_1}{\partial \lambda_1} & \frac{\partial \phi_1}{\partial \lambda_2} & \cdots & \frac{\partial \phi_1}{\partial \lambda_d} \\ \frac{\partial \phi_2}{\partial \lambda_1} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \frac{\partial \phi_m}{\partial \lambda_1} & \cdots & \cdots & \frac{\partial \phi_m}{\partial \lambda_d} \end{bmatrix}.$$

Here and later on, the differential operator $\partial/\partial\lambda_j$ is acting from right to left keeping the matrix calculus valid. We can write this in a vector form as follows.

DEFINITION 2.1. The operator $D_{\boldsymbol{\lambda}}^{\otimes}$ is defined as

$$D_{\boldsymbol{\lambda}}^{\otimes}\boldsymbol{\phi} = \operatorname{Vec}\left(\frac{\partial \boldsymbol{\phi}}{\partial \boldsymbol{\lambda}'}\right)' = \operatorname{Vec}\begin{bmatrix} \frac{\partial \phi_1}{\partial \lambda_1} & \frac{\partial \phi_1}{\partial \lambda_2} & \cdots & \frac{\partial \phi_1}{\partial \lambda_d} \\ \frac{\partial \phi_2}{\partial \lambda_1} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \frac{\partial \phi_m}{\partial \lambda_1} & \cdots & \cdots & \frac{\partial \phi_m}{\partial \lambda_d} \end{bmatrix}',$$

which is a column vector of order md.

We refer to $D_{\boldsymbol{\lambda}}^{\otimes}$ as K-derivative, and we can also write $D_{\boldsymbol{\lambda}}^{\otimes}$ as a Kronecker product.

$$\begin{aligned} D_{\boldsymbol{\lambda}}^{\otimes}\boldsymbol{\phi} &= \operatorname{Vec}\left(\boldsymbol{\phi}\frac{\partial}{\partial \boldsymbol{\lambda}'}\right)' = \operatorname{Vec}\left(\frac{\partial}{\partial \boldsymbol{\lambda}}\boldsymbol{\phi}'\right) \\ &= [\phi_1(\boldsymbol{\lambda}), \phi_2(\boldsymbol{\lambda}), \ldots, \phi_m(\boldsymbol{\lambda})]' \otimes \left[\frac{\partial}{\partial \lambda_1}, \frac{\partial}{\partial \lambda_2}, \ldots, \frac{\partial}{\partial \lambda_d}\right]'. \end{aligned}$$

If we apply the operator $D_{\boldsymbol{\lambda}}^{\otimes}$ twice, we obtain

$$\begin{aligned} D_{\boldsymbol{\lambda}}^{\otimes 2}\boldsymbol{\phi} = D_{\boldsymbol{\lambda}}^{\otimes}\left(D_{\boldsymbol{\lambda}}^{\otimes}\boldsymbol{\phi}\right) &= \operatorname{Vec}\left[\left(\boldsymbol{\phi} \otimes \frac{\partial}{\partial \boldsymbol{\lambda}}\right)\frac{\partial}{\partial \boldsymbol{\lambda}'}\right]' \\ &= \boldsymbol{\phi} \otimes \left(\frac{\partial}{\partial \boldsymbol{\lambda}}\right)^{\otimes 2} = \boldsymbol{\phi} \otimes \frac{\partial}{\partial \boldsymbol{\lambda}^{\otimes 2}}, \end{aligned}$$

and in general (assuming k times differentiability), the k^{th} K-derivative is given by

$$D_{\boldsymbol{\lambda}}^{\otimes k}\boldsymbol{\phi} = D_{\boldsymbol{\lambda}}^{\otimes}\left(D_{\boldsymbol{\lambda}}^{\otimes k-1}\boldsymbol{\phi}\right)$$
$$= \left[\phi_1(\boldsymbol{\lambda}), \phi_2(\boldsymbol{\lambda}), \ldots, \phi_m(\boldsymbol{\lambda})\right]' \otimes \left[\frac{\partial}{\partial\lambda_1}, \frac{\partial}{\partial\lambda_2}, \ldots, \frac{\partial}{\partial\lambda_d}\right]'^{\otimes k}$$

which is a vector of order md^k containing all possible partial derivatives of entries of $\boldsymbol{\phi}$ according to the Kronecker product $\left(\frac{\partial}{\partial\lambda_1}, \frac{\partial}{\partial\lambda_2}, \ldots, \frac{\partial}{\partial\lambda_d}\right)'^{\otimes k}$.

In the following, we give some additional properties of this operator $D_{\boldsymbol{\lambda}}^{\otimes}$ when applied to products of several functions. Let $\boldsymbol{K}_{3\leftrightarrow 2}(m_1, m_2, d)$ denote the commutation matrix of size $m_1 m_2 d \times m_1 m_2 d$, changing the order in a Kronecker product of three vectors of dimension (m_1, m_2, d) (see the Subsection 5.1 in the Appendix for details), such that the second and the third places are interchanged. For example, if $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$ are vectors of dimension m_1, m_2, d respectively, then we have $\boldsymbol{K}_{3\leftrightarrow 2}(m_1, m_2, d)$ as the matrix defined as

$$\boldsymbol{K}_{3\leftrightarrow 2}(m_1, m_2, d)(\boldsymbol{a}_1 \otimes \boldsymbol{a}_2 \otimes \boldsymbol{a}_3) = \boldsymbol{a}_1 \otimes \boldsymbol{a}_3 \otimes \boldsymbol{a}_2.$$

Property 2.1 (Chain Rule). If $\boldsymbol{\lambda} \in \mathbb{R}^d$, $\boldsymbol{\phi}_1 \in \mathbb{R}^{m_1}$ and $\boldsymbol{\phi}_2 \in \mathbb{R}^{m_2}$, then

$$D_{\boldsymbol{\lambda}}^{\otimes}(\boldsymbol{\phi}_1 \otimes \boldsymbol{\phi}_2) = \boldsymbol{K}_{3\leftrightarrow 2}^{-1}(m_1, m_2, d)\left(\left(D_{\boldsymbol{\lambda}}^{\otimes}\boldsymbol{\phi}_1\right) \otimes \boldsymbol{\phi}_2\right) + \boldsymbol{\phi}_1 \otimes \left(D_{\boldsymbol{\lambda}}^{\otimes}\boldsymbol{\phi}_2\right), \quad (2.1)$$

where $\boldsymbol{K}_{3\leftrightarrow 2}(m_1, m_2, d)$ denotes the commutation matrix. This Chain Rule (2.1) can be extended to products of several functions. If $\boldsymbol{\phi}_k \in \mathbb{R}^{m_k}$, $k = 1, 2, \ldots, M$, then

$$D_{\boldsymbol{\lambda}}^{\otimes}\prod\nolimits_{(1:M)}^{\otimes}\boldsymbol{\phi}_k$$
$$= \sum_{j=1}^{M}\boldsymbol{K}_{\mathfrak{p}_{M+1\to j}}^{-1}(m_{1:M}, d)\left[\prod\nolimits_{(1:j-1)}^{\otimes}\boldsymbol{\phi}_k \otimes \left[D_{\boldsymbol{\lambda}}^{\otimes}\boldsymbol{\phi}_j(\boldsymbol{\lambda})\right] \otimes \prod\nolimits_{(j+1:M)}^{\otimes}\boldsymbol{\phi}_k\right].$$

Here, the commutation matrix $\boldsymbol{K}_{\mathfrak{p}_{M+1\to j}}(m_{1:M}, d)$ permutes the vectors of dimension $(m_{1:M}, d)$ in the Kronecker product according to the permutation $\mathfrak{p}_{M+1\to j}$ of the integers $(1 : M+1) = (1, 2, \ldots, M+1)$.

Consider the special case $\boldsymbol{\phi}(\boldsymbol{\lambda}) = \boldsymbol{\lambda}^{\otimes k}$. By differentiating according to the definition 1, we get

$$D_{\boldsymbol{\lambda}}^{\otimes}\boldsymbol{\lambda}^{\otimes k} = \operatorname{Vec}\left(\frac{\partial \boldsymbol{\lambda}^{\otimes k}}{\partial \boldsymbol{\lambda}'}\right)' = \boldsymbol{\lambda}^{\otimes k} \otimes \left(\frac{\partial}{\partial \lambda_1}, \frac{\partial}{\partial \lambda_2}, \dots, \frac{\partial}{\partial \lambda_d}\right)'$$

$$= \left(\sum_{j=0}^{k-1} \boldsymbol{K}_{j+1\leftrightarrow k}\left(d_{[k]}\right)\right)\left(\boldsymbol{\lambda}^{\otimes(k-1)} \otimes \boldsymbol{I}_d\right), \tag{2.2}$$

where $d_{[k]} = \underbrace{[d, d, \dots, d]}_{k}$.

Now suppose that $\phi(\boldsymbol{\lambda}) = \boldsymbol{x}'^{\otimes k}\boldsymbol{\lambda}^{\otimes k}$ where the vector $\boldsymbol{x}$ is a vector of constants. Here, ϕ is a scalar valued function. By using the Property 1 and after differentiating r times, we obtain

$$D_{\boldsymbol{\lambda}}^{\otimes r}\boldsymbol{x}'^{\otimes k}\boldsymbol{\lambda}^{\otimes k} = k(k-1)\cdots(k-r+1)\left[\left(\boldsymbol{x}'\boldsymbol{\lambda}\right)^{k-r}\boldsymbol{x}^{\otimes r}\right]. \tag{2.3}$$

The reason for (2.3) is that the Kronecker product $\boldsymbol{x}'^{\otimes k}$ is invariant under the permutation of its component vectors $\boldsymbol{x}$, i.e.,

$$\boldsymbol{x}'^{\otimes l}\boldsymbol{K}_{j+1\to l}\left(d_{[l]}\right) = \boldsymbol{x}'^{\otimes l},$$

for any l and j so that

$$\boldsymbol{x}'^{\otimes k}\left(\sum_{j=0}^{k-1} \boldsymbol{K}_{j+1\to k}\left(d_{[k]}\right)\right) = k\boldsymbol{x}'^{\otimes k},$$

and thus we obtain (2.3). In particular, if $r = k$, we have

$$D_{\boldsymbol{\lambda}}^{\otimes k}\boldsymbol{x}'^{\otimes k}\boldsymbol{\lambda}^{\otimes k} = k!\boldsymbol{x}^{\otimes k}.$$

2.2. Taylor series expansion of functions of several variables. Let $\phi(\boldsymbol{\lambda}) = \phi(\lambda_1, \lambda_2, \dots, \lambda_d)$ and assume that ϕ is differentiable several times in each variable. Here our objective is to expand $\phi(\boldsymbol{\lambda})$ in Taylor series, and the expression is given in terms of differential operators given above. We use this expansion later to define the characteristic function and the cumulant functions in terms of the differential operators. Let $\boldsymbol{\lambda} = \lambda_{(1:d)} = (\lambda_1, \lambda_2, \dots, \lambda_d)' \in \mathbb{R}^d$. It is well known that the Taylor series of $\phi(\boldsymbol{\lambda})$ is

$$\phi(\boldsymbol{\lambda}) = \sum_{k_1, k_2, \dots, k_d = 0}^{\infty} \frac{1}{\boldsymbol{k}!} c(\boldsymbol{k})\, \boldsymbol{\lambda}_{\boldsymbol{k}}, \tag{2.4}$$

where the coefficients are

$$c(\boldsymbol{k}) = \left.\frac{\partial^{\Sigma \boldsymbol{k}} \phi(\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}^{\boldsymbol{k}}}\right|_{\boldsymbol{\lambda}=0}.$$

Here we use the notation

$$\boldsymbol{k} = (k_1, k_2, \ldots, k_d), \ \boldsymbol{k}! = k_1! k_2! \ldots k_d!,$$
$$\boldsymbol{\lambda}_{\boldsymbol{k}} = \prod_{j=1}^{d} \lambda_j^{k_j}, \ \partial \boldsymbol{\lambda}_{\boldsymbol{k}} = \partial \lambda_1^{k_1} \partial \lambda_2^{k_2} \cdots \partial \lambda_d^{k_d}.$$

The Taylor series (2.4) can be written in a more informative form for our purposes, namely

$$\phi(\boldsymbol{\lambda}) = \sum_{m=0}^{\infty} \frac{1}{m!} \boldsymbol{c}(m,d)' \boldsymbol{\lambda}^{\otimes m},$$

where $\boldsymbol{c}(m,d)$ is a column vector, which is the derivative (K-derivative) of the function ϕ given by

$$\boldsymbol{c}(m,d) = \left(D_{\boldsymbol{\lambda}}^{\otimes m} \phi(\boldsymbol{\lambda})\right)\big|_{\boldsymbol{\lambda}=\mathbf{0}}$$

(see Subsection 5.2 for details).

3 Moments and Cumulants of Random Vectors

3.1. Characteristic function and moments of random vectors. Let $\boldsymbol{X}$ be a d-dimensional random vector and let $\boldsymbol{X} = [\boldsymbol{X}_1', \boldsymbol{X}_2']'$, where $\boldsymbol{X}_1$ is of dimension d_1, and $\boldsymbol{X}_2$ is of dimension d_2 such that $d = d_1 + d_2$. Let $\boldsymbol{\lambda} = [\boldsymbol{\lambda}_1', \boldsymbol{\lambda}_2']'$. The characteristic function of $\boldsymbol{X}$ is given by

$$\begin{aligned}
\varphi(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) &= \mathsf{E} \exp\left[i\left(\boldsymbol{X}_1' \boldsymbol{\lambda}_1 + \boldsymbol{X}_2' \boldsymbol{\lambda}_2\right)\right] \\
&= \sum_{k,l=0}^{\infty} \frac{i^{k+l}}{k! l!} \mathsf{E}\left(\boldsymbol{X}_1' \boldsymbol{\lambda}_1\right)^k \left(\boldsymbol{X}_2' \boldsymbol{\lambda}_2\right)^l \\
&= \sum_{k,l=0}^{\infty} \frac{i^{k+l}}{k! l!} \mathsf{E}\left(\boldsymbol{X}_1^{\otimes k\prime} \otimes \boldsymbol{X}_2^{\otimes l\prime}\right) \left(\boldsymbol{\lambda}_1^{\otimes k} \otimes \boldsymbol{\lambda}_2^{\otimes l}\right).
\end{aligned}$$

Here the coefficients of $\boldsymbol{\lambda}_1^{\otimes k} \otimes \boldsymbol{\lambda}_2^{\otimes l}$ can be obtained by using the K-derivative and the formula (2.3). Consider the second K-derivative of φ

$$\begin{aligned}
&D_{\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2}^{\otimes(1,1)} \varphi(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) \\
&= D_{\boldsymbol{\lambda}_2}^{\otimes} \left(D_{\boldsymbol{\lambda}_1}^{\otimes} \varphi(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) \right) \\
&= \varphi(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) \left(\frac{\partial}{\partial \boldsymbol{\lambda}_1} \right) \otimes \left(\frac{\partial}{\partial \boldsymbol{\lambda}_2} \right) \\
&= \sum_{k,l=1}^{\infty} \frac{i^{k+l-2}}{(k-1)!\,(l-1)!} \mathsf{E} \boldsymbol{X}_1^{\otimes k-1\prime} \boldsymbol{\lambda}_1^{\otimes k-1} \boldsymbol{X}_2^{\otimes l-1\prime} \boldsymbol{\lambda}_2^{\otimes l-1} \left(\boldsymbol{X}_1 \otimes \boldsymbol{X}_2 \right).
\end{aligned}$$

Now, by evaluating the derivative $D_{\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2}^{\otimes(1,1)} \varphi(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) \Big|_{\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2=\mathbf{0}}$, we obtain $\mathsf{E}\boldsymbol{X}_1 \otimes X_2$. Other moments can be obtained similarly from higher order derivatives. Therefore, the Taylor series expansion of $\varphi(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2)$ can be written in terms of derivatives and is given by

$$\varphi(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) = \sum_{k,l=0}^{\infty} \frac{i^{k+l}}{k!l!} \left(D_{\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2}^{\otimes(k,l)} \varphi(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) \right)' \Big|_{\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2=\mathbf{0}} \boldsymbol{\lambda}_1^{\otimes k} \otimes \boldsymbol{\lambda}_2^{\otimes l}.$$

We note that in general

$$\begin{aligned}
\left(D_{\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2}^{\otimes(k,l)} \varphi(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) \right) \Big|_{\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2=\mathbf{0}} &= i^{k+l} \mathsf{E} \boldsymbol{X}_1^{\otimes k} \otimes \boldsymbol{X}_2^{\otimes l} \\
&\neq i^{k+l} \mathsf{E} \boldsymbol{X}_2^{\otimes l} \otimes \boldsymbol{X}_1^{\otimes k} \\
&= \left(D_{\boldsymbol{\lambda}_2,\boldsymbol{\lambda}_1}^{\otimes(l,k)} \varphi(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) \right) \Big|_{\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2=\mathbf{0}},
\end{aligned}$$

which shows that the partial derivatives in this case are not symmetric.

Consider a set of vectors $\lambda_{(1:n)} = \left[\boldsymbol{\lambda}_1', \boldsymbol{\lambda}_2', \ldots, \boldsymbol{\lambda}_n' \right]'$ with dimensions $[d_1, d_2, \ldots, d_n]$. We can define the operator $D_{\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2}^{\otimes(1,1)}$ given in the Appendix for the partitioned set of vectors $\lambda_{(1:n)}$. This is achieved recursively. Recall that the K-derivative with respect to$\boldsymbol{\lambda}_j$ is

$$D_{\boldsymbol{\lambda}_j}^{\otimes} \varphi = \operatorname{Vec} \left(\varphi \frac{\partial}{\partial \boldsymbol{\lambda}_j'} \right)'.$$

DEFINITION 3.1. The n^{th} derivative $D_{\boldsymbol{\lambda}_{(1:n)}}^{\otimes n}$ is defined recursively by

$$D_{\boldsymbol{\lambda}_{(1:n)}}^{\otimes n} \varphi = D_{\boldsymbol{\lambda}_n}^{\otimes} \left(D_{\boldsymbol{\lambda}_{(1:n-1)}}^{\otimes n-1} \varphi \right),$$

where $D^{\otimes n}_{\boldsymbol{\lambda}_{(1:n)}}\varphi$ is a column vector of the partial differential operator of order n.

We see that this is the first order derivative of the function, which is already a $(n-1)^{th}$ order partial derivative. The dimension of $D^{\otimes n}_{\boldsymbol{\lambda}_{(1:n)}}$ is $d_{1:n}^{\mathbf{1}_{[n]}} = \prod_{j=1}^{n} d_j$, where $\mathbf{1}_{[n]}$ denotes a row vector having all ones as its entries, i.e., $\mathbf{1}_{[n]} = [1,1,...,1]$ with dimension n. The order of the vectors in $\lambda_{(1:n)}$ is important.

The following definition generalizes a similar well-known result for scalar valued random variables, to the multivariate case. Here we assume that the partial derivatives exist.

Definition 3.2. Suppose that $\mathbf{X}_{(1:n)} = (\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n)$ is a collection of random (column) vectors with dimensions $[d_1, d_2, \ldots, d_n]$. The Kronecker moment is defined by the following K-derivative:

$$\begin{aligned}&\mathsf{E}\left(\boldsymbol{X}_1 \otimes \boldsymbol{X}_2 \cdots \otimes \boldsymbol{X}_n\right)\\ &= \mathsf{E}\prod_{j=1}^{\otimes n} \boldsymbol{X}_j = (-i)^n D^{\otimes n}_{\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \ldots, \boldsymbol{\lambda}_n} \varphi_{\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n}\left(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \ldots, \boldsymbol{\lambda}_n\right)\Big|_{\boldsymbol{\lambda}_{(1:n)}=0}.\end{aligned}$$

We note that the order of the product in the expectations and the derivatives are important, since the Kronecker moment is not symmetric if the variables $\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_n$ are different.

3.2. Cumulant function and cumulants of random vectors. We obtain the cumulant $\operatorname{Cum}_n(\boldsymbol{X})$ as the derivative of the logarithm of the characteristic function $\varphi_{\boldsymbol{X}}(\boldsymbol{\lambda})$ of $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)'$ and then evaluate the function at zero to obtain:

$$(-i)^n \left.\frac{\partial^n \ln \varphi_{\boldsymbol{X}}(\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}^{\mathbf{1}_{[n]}}}\right|_{\boldsymbol{\lambda}_{(1:n)}=0} = \operatorname{Cum}_n(\boldsymbol{X}) = \operatorname{Cum}_n(X_1, X_2, \ldots, X_n),$$

where $\partial \boldsymbol{\lambda}^{\mathbf{1}_{[n]}} = \partial\lambda_1 \partial\lambda_2 \cdots \partial\lambda_n$. See Terdik (1999) for details.

Now consider the collection of random vectors

$$\mathbf{X}_{(1:n)} = (\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n),$$

where each $\boldsymbol{X}_i$ is of order d_i. The corresponding characteristic function of $\operatorname{Vec}\mathbf{X}_{(1:n)}$ is

$$\varphi_{\mathbf{X}_{(1:n)}}\left(\lambda_{(1:n)}\right) = \varphi_{\operatorname{Vec}\mathbf{X}_{(1:n)}}\left(\operatorname{Vec}\lambda_{(1:n)}\right) = \mathsf{E}\exp\left(i\left(\operatorname{Vec}\lambda_{(1:n)}, \operatorname{Vec}\mathbf{X}_{(1:n)}\right)\right),$$

where $\lambda_{(1:n)} = (\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \dots, \boldsymbol{\lambda}_n)$ and $d_{(1:n)} = (d_1, d_2, \dots, d_n)$. We call the logarithm of the characteristic function $\varphi_{\mathrm{Vec}\,\mathbf{X}_{(1:n)}}\left(\mathrm{Vec}\,\lambda_{(1:n)}\right)$ as the cumulant function and denote it by

$$\psi_{\mathrm{Vec}\,\mathbf{X}_{(1:n)}}\left(\mathrm{Vec}\,\lambda_{(1:n)}\right) = \ln \varphi_{\mathrm{Vec}\,\mathbf{X}_{(1:n)}}\left(\mathrm{Vec}\,\lambda_{(1:n)}\right).$$

We write $\psi_{\mathbf{X}_{(1:n)}}(\lambda_{(1:n)})$ for $\psi_{\mathrm{Vec}\,\mathbf{X}_{(1:n)}}(\mathrm{Vec}\,\boldsymbol{\lambda}_{(1:n)})$. The first order K-derivative of the cumulant function $\psi_{\mathbf{X}_{(1:n)}}\left(\lambda_{(1:n)}\right)$ with respect to $\lambda_{(1:n)}$ is defined as the cumulant of $\mathbf{X}_{(1:n)}$. Now we use the operator $D^{\otimes n}_{\boldsymbol{\lambda}_{(1:n)}}\psi = D^{\otimes}_{\boldsymbol{\lambda}_n}\left(D^{\otimes n-1}_{\boldsymbol{\lambda}_{(1:n-1)}}\psi\right)$ recursively and the result is a column vector of the partial differentials of order n, which is first order in each variable $\boldsymbol{\lambda}_j$. The dimension of $D^{\otimes n}_{\boldsymbol{\lambda}_{(1:n)}}$ is $d^{\mathbf{1}_{[n]}}_{1:n} = \prod_{j=1}^{n} d_j$. Now we define the n^{th} order cumulant of vectors $\mathbf{X}_{(1:n)}$ as follows.

DEFINITION 3.3.

$$\mathrm{Cum}_n\left(\mathbf{X}_{(1:n)}\right) = (-i)^n D^{\otimes n}_{\boldsymbol{\lambda}_{(1:n)}}\psi_{\mathbf{X}_{(1:n)}}\left(\lambda_{(1:n)}\right)\Big|_{\boldsymbol{\lambda}_{(1:n)}=0}. \tag{3.1}$$

Therefore, $\mathsf{Cum}_n\left(\mathbf{X}_{(1:n)}\right)$ is a vector of dimension $d^{\mathbf{1}_{[n]}}_{1:n}$ containing all possible cumulants of the elements formed from the vectors $\boldsymbol{X}_1, \boldsymbol{X}_2, \dots, \boldsymbol{X}_n$, and the order is defined by the Kronecker products defined earlier (see also Terdik, 2002). This definition also includes the evaluation of the cumulants, where all the random vectors $\boldsymbol{X}_1, \boldsymbol{X}_2, \dots, \boldsymbol{X}_n$ need not to be distinct. In this case, the characteristic function depends on the sum of the corresponding variables of $\lambda_{(1:n)}$, and we use still the definition (3.1) to obtain the cumulant.

For example, when $n = 1$, we have

$$\mathsf{Cum}_1(\boldsymbol{X}) = \mathsf{E}\boldsymbol{X},$$

and when $n = 2$,

$$\begin{aligned}\mathsf{Cum}_2(\boldsymbol{X}_1, \boldsymbol{X}_2) &= \mathsf{E}\left[(\boldsymbol{X}_1 - \mathsf{E}\boldsymbol{X}_1) \otimes (\boldsymbol{X}_2 - \mathsf{E}\boldsymbol{X}_2)\right] \\ &= \mathrm{Vec}\,\mathrm{Cov}(\boldsymbol{X}_1, \boldsymbol{X}_2),\end{aligned} \tag{3.2}$$

where $\mathrm{Cov}(\boldsymbol{X}_1, \boldsymbol{X}_2)$ denotes the covariance matrix of the vectors $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$. To illustrate the above formulae, let us consider an example.

EXAMPLE 3.1. Let $\mathbf{X}_{(1,2)} = \left(\boldsymbol{X}_1', \boldsymbol{X}_2'\right)'$ and assume that $\mathbf{X}_{(1,2)}$ has a joint normal distribution with moment $(\boldsymbol{\mu}_1', \boldsymbol{\mu}_2')'$, and the variance covariance matrix $\mathbf{C}_{(\boldsymbol{X}_1,\boldsymbol{X}_2)}$ is given by

$$\mathbf{C}_{(\boldsymbol{X}_1',\boldsymbol{X}_2')'} = \begin{bmatrix} \mathbf{C}_{1,1} & \mathbf{C}_{1,2} \\ \mathbf{C}_{2,1} & \mathbf{C}_{2,2} \end{bmatrix}.$$

Then the characteristic function of $\mathbf{X}_{(1,2)}$ is given by

$$\varphi_{\mathbf{X}_{(1,2)}}(\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2) = \exp\Big\{ i\left(\boldsymbol{\mu}_1'\boldsymbol{\lambda}_1 + \boldsymbol{\mu}_2'\boldsymbol{\lambda}_2\right) - \frac{1}{2}\left(\boldsymbol{\lambda}_1'\mathbf{C}_{1,1}\boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_1'\mathbf{C}_{1,2}\boldsymbol{\lambda}_2 + \boldsymbol{\lambda}_2'\mathbf{C}_{2,1}\boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_2'\mathbf{C}_{2,2}\boldsymbol{\lambda}_2\right)\Big\},$$

and the cumulant function of $\mathbf{X}_{(1,2)}$ is

$$\begin{aligned}
&\psi_{\mathbf{X}_{(1,2)}}(\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2)\\
&= \ln\varphi_{\mathbf{X}_{(1,2)}}(\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2)\\
&= i(\boldsymbol{\mu}_1'\boldsymbol{\lambda}_1 + \boldsymbol{\mu}_2'\boldsymbol{\lambda}_2) - \frac{1}{2}\left(\boldsymbol{\lambda}_1'\mathbf{C}_{1,1}\boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_1'\mathbf{C}_{1,2}\boldsymbol{\lambda}_2 + \boldsymbol{\lambda}_2'\mathbf{C}_{2,1}\boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_2'\mathbf{C}_{2,2}\boldsymbol{\lambda}_2\right).
\end{aligned}$$

Now, the first order cumulant is

$$\mathsf{Cum}_1(\boldsymbol{X}_j) = -iD^{\otimes}_{\boldsymbol{\lambda}_j}\varphi_{\mathbf{X}_{(1,2)}}(\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2)\Big|_{\boldsymbol{\lambda}_1=\boldsymbol{\lambda}_2=0} = \boldsymbol{\mu}_j,$$

and it is clear that any cumulant of order higher than two is zero. One can easily show that the second order cumulants are the vectors of the covariance matrices, i.e.,

$$\mathsf{Cum}_2(\boldsymbol{X}_j,\boldsymbol{X}_k) = \operatorname{Vec}\mathbf{C}_{k,j}, \quad j,k = 1,2.$$

For instance, if $j = 2$ and $k = 1$,

$$D^{\otimes 2}_{\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2}\boldsymbol{\lambda}_2'\mathbf{C}_{2,1}\boldsymbol{\lambda}_1 = D^{\otimes}_{\boldsymbol{\lambda}_2}\left(D^{\otimes}_{\boldsymbol{\lambda}_1}\boldsymbol{\lambda}_2'\mathbf{C}_{2,1}\boldsymbol{\lambda}_1\right) = \operatorname{Vec}\mathbf{C}_{2,1}.$$

If $j = k = 1$, then,

$$D^{\otimes}_{\boldsymbol{\lambda}_1}\boldsymbol{\lambda}_1'\mathbf{C}_{1,1}\boldsymbol{\lambda}_1 = 2\operatorname{Vec}\left(\boldsymbol{\lambda}_1'\mathbf{C}_{1,1}\right)' = 2\mathbf{C}_{1,1}\boldsymbol{\lambda}_1,$$

and by applying $D^{\otimes}_{\boldsymbol{\lambda}_1}$ repeatedly, we obtain

$$\mathsf{Cum}_2(\boldsymbol{X}_1,\boldsymbol{X}_1) = D_{\boldsymbol{\lambda}_1}\left(D_{\boldsymbol{\lambda}_1}\boldsymbol{\lambda}_1'\mathbf{C}_{1,1}\boldsymbol{\lambda}_1\right)/2 = \operatorname{Vec}\mathbf{C}_{1,1}.$$

3.3. Basic properties of the cumulants. For convenience of notation, we set the dimensions of $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ to be equal to d. The cumulants are symmetric in scalar valued case, but not for vectors. For example, $\mathsf{Cum}_2(\boldsymbol{X}_1, \boldsymbol{X}_2) \neq \mathsf{Cum}_2(\boldsymbol{X}_2, \boldsymbol{X}_1)$. Here we have to use permutation matrices (see Appendix for details) as will be shown below.

PROPOSITION 3.1. *Let $\mathfrak{p}$ be a permutation of integers $(1:n)$, and let the function $\boldsymbol{f}\left(\lambda_{(1:n)}\right) \in \mathbb{R}^d$ be continuously differentiable n times in all its arguments. Then*

$$D^{\otimes n}_{\boldsymbol{\lambda}_{\mathfrak{p}(1:n)}} \boldsymbol{f} = \left(\boldsymbol{I}_d \otimes \boldsymbol{K}_{\mathfrak{p}(1:n)}(d_{1:n})\right) D^{\otimes n}_{\boldsymbol{\lambda}_{(1:n)}} \boldsymbol{f}.$$

(1) Symmetry. If $d > 1$, then the cumulants are not symmetric but satisfy the relation

$$\mathsf{Cum}_n(\mathbf{X}_{(1:n)}) = \boldsymbol{K}^{-1}_{\mathfrak{p}(1:n)}\left(d_{[n]}\right) \mathsf{Cum}_n(\mathbf{X}_{\mathfrak{p}(1:n)}),$$

where $\mathfrak{p}(1:n) = (\mathfrak{p}(1), \mathfrak{p}(2), \ldots, \mathfrak{p}(n))$ belongs to the set of all possible permutations $\mathfrak{P}_n$ of the numbers $(1:n)$, $d_{[n]} = \underbrace{(d, d, ..., d)}_{n}$, and $\boldsymbol{K}_{\mathfrak{p}(1:n)}\left(d_{[n]}\right)$ is the permutation matrix (see Appendix, equation A.1).

For constant matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ and random vectors $\boldsymbol{Y}_1$, $\boldsymbol{Y}_2$,

$$\mathsf{Cum}_{n+1}(\boldsymbol{A}\boldsymbol{Y}_1 + \boldsymbol{B}\boldsymbol{Y}_2, \mathbf{X}_{(1:n)}) = (\boldsymbol{A} \otimes \boldsymbol{I}_{d^n}) \mathsf{Cum}_{n+1}(\boldsymbol{Y}_1, \mathbf{X}_{(1:n)}) + (\boldsymbol{B} \otimes \boldsymbol{I}_{d^n}) \mathsf{Cum}_{n+1}(\boldsymbol{Y}_2, \mathbf{X}_{(1:n)}).$$

Also

$$\mathsf{Cum}_{n+1}(\boldsymbol{A}\boldsymbol{Y}_1, \boldsymbol{B}\boldsymbol{Y}_2, \mathbf{X}_{(1:n)}) = (\boldsymbol{a} \otimes \boldsymbol{B} \otimes \boldsymbol{I}_{d^n}) \mathsf{Cum}_{n+1}(\boldsymbol{Y}_1, \boldsymbol{Y}_2, \mathbf{X}_{(1:n)}),$$

assuming that the appropriate matrix operations are valid.

For any pair of constant vectors $\boldsymbol{a}$ and $\boldsymbol{b}$,

$$\mathsf{Cum}_{n+1}(\boldsymbol{a} \otimes \boldsymbol{Y}_1 + \boldsymbol{b} \otimes \boldsymbol{Y}_2, \mathbf{X}_{(1:n)}) = \boldsymbol{a} \otimes \mathsf{Cum}_{n+1}(\boldsymbol{Y}_1, \mathbf{X}_{(1:n)}) + \boldsymbol{b} \otimes \mathsf{Cum}_{n+1}(\boldsymbol{Y}_2, \mathbf{X}_{(1:n)}).$$

(2) Independence. If $\mathbf{X}_{(1:n)}$ is independent of $\mathbf{Y}_{(1:m)}$, where $n, m > 0$, then

$$\mathsf{Cum}_{n+m}(\mathbf{X}_{(1:n)}, \mathbf{Y}_{(1:m)}) = 0.$$

In particular, if the dimensions are same, we have

$$\mathsf{Cum}_n(\mathbf{X}_{(1:n)} + \mathbf{Y}_{(1:n)}) = \mathsf{Cum}_n(\mathbf{X}_{(1:n)}) + \mathsf{Cum}_n(\mathbf{Y}_{(1:n)}).$$

(3) Gaussianity. The random vector $\mathbf{X}_{(1:n)}$ is Gaussian if and only if for all subsets $k_{(1:m)}$ of $(1:n)$,

$$\mathsf{Cum}_m(\mathbf{X}_{k_{(1:m)}}) = 0, \ m > 2.$$

For further properties of the cumulants, we need the following Lemma which makes it easier to understand the relations between the moments and the cumulants (see Barndorff-Nielsen and Cox, 1989, p. 140).

REMARK 3.1. Let $\mathcal{P}_n$ be the set of all partitions $\mathcal{K}$ of the integers $(1:n)$. If $\mathcal{K} = \{\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_m\}$, where each $\boldsymbol{b}_j \subset (1:n)$, then $|\mathcal{K}| = m$ denotes the *size* of $\mathcal{K}$. We introduce an ordering among the blocks in $\mathcal{K}$. For $\boldsymbol{b}_j, \boldsymbol{b}_k \in \mathcal{K}$, we write $\boldsymbol{b}_j \leq \boldsymbol{b}_k$ if

$$\sum_{l \in \boldsymbol{b}_j} 2^{-l} \leq \sum_{l \in \boldsymbol{b}_k} 2^{-l}, \tag{3.3}$$

and equality in (3.3) is possible if and only if $j = k$. The partition $\mathcal{K}$ will be considered as ordered if both, the elements of a block are ordered inside the block, and the blocks are ordered by the above relation $\boldsymbol{b}_j \leq \boldsymbol{b}_k$ also. We suppose that all partitions $\mathcal{K}$ in $\mathcal{P}_n$ are ordered. Denote $\boldsymbol{\lambda} = \lambda_{(1:M)} = \left[\boldsymbol{\lambda}_1', \boldsymbol{\lambda}_2', \ldots, \boldsymbol{\lambda}_M'\right]' \in \mathbb{R}^N$,where $\boldsymbol{\lambda}_j \in \mathbb{R}^{d_j}$ and $N = d_{1:n}^{\mathbf{1}[n]}$. In this case, the differential operator $D_{\boldsymbol{\lambda}_{\boldsymbol{b}}}^{\otimes|\boldsymbol{b}|}$ is well defined because the vector $\lambda_{\boldsymbol{b}} = \left[\boldsymbol{\lambda}_j', j \in \boldsymbol{b}\right]$ denotes an ordered subset of vectors $\left[\boldsymbol{\lambda}_1', \boldsymbol{\lambda}_2', \ldots, \boldsymbol{\lambda}_M'\right]$ corresponding to the order in $\boldsymbol{b}$. The permutation $\mathfrak{p}(\mathcal{K})$ of the numbers $(1:n)$ corresponds to the ordered partition $\mathcal{K}$. See Andrews (1976) for more details on partitions.

We can rewrite the formula of Faà di Bruno given for implicit functions (see Lukács, 1955) as follows.

LEMMA 3.1. *Consider the implicit function* $f(g(\boldsymbol{\lambda}))$, $\boldsymbol{\lambda} \in \mathbb{R}^d$, *where* f *and* g *are scalar valued functions and are differentiable* M *times. Suppose that* $\boldsymbol{\lambda} = \lambda_{(1:M)} = \left[\boldsymbol{\lambda}_1', \boldsymbol{\lambda}_2', \ldots, \boldsymbol{\lambda}_M'\right]'$ *with dimensions* $[d_1, d_2, \ldots, d_M]$. *Then, for* $n \leq M$,

$$D_{\boldsymbol{\lambda}_{(1:n)}}^{\otimes n} f(g(\boldsymbol{\lambda})) = \sum_{r=1}^{n} f^{(r)}(g(\boldsymbol{\lambda})) \sum_{\substack{\mathcal{K} \in \mathcal{P}_n \\ |\mathcal{K}| = r}} \boldsymbol{K}_{\mathfrak{p}(\mathcal{K})}^{-1}(d_{1:n}) \prod\nolimits_{\boldsymbol{b} \in \mathcal{K}}^{\otimes} \left(D_{\boldsymbol{\lambda}_{\boldsymbol{b}}}^{\otimes|\boldsymbol{b}|} g(\boldsymbol{\lambda})\right), \tag{3.4}$$

where $\mathfrak{p}(\mathcal{K})$ *is a permutation of* $(1:n)$ *defined by the partition* $\mathcal{K}$, *see Remark 3.1.*

We consider particular cases of Equation (3.4) which are useful for proving some properties of cumulants.

3.4. Cumulants in terms of moments. The results obtained here are generalizations of the well known results given for scalar random variables by Leonov and Shiryaev(1959) (see also Brillinger and Rosenblatt, 1967, Terdik, 1999). In order to obtain the cumulants in terms of moments, let us consider the function $f(x) = \ln x$ and $g(\boldsymbol{\lambda}) = \varphi_{\mathbf{X}_{(1:n)}}\left(\lambda_{(1:n)}\right)$. The r^{th} derivative of $f(x) = \ln x$ is

$$f^{(r)}(x) = (-1)^{r-1}(r-1)!x^{-r}.$$

So, the left hand side of Equation (3.4) is the cumulant of $\mathbf{X}_{(1:n)}$. Hence, we obtain

$$\begin{aligned}&\mathsf{Cum}_n(\mathbf{X}_{(1:n)})\\&=\sum_{m=1}^{n}(-1)^{m-1}(m-1)!\sum_{\substack{\mathcal{L}\in\mathcal{P}_{(1:n)}\\|\mathcal{L}|=m}}\boldsymbol{K}^{-1}_{\mathfrak{p}(\mathcal{L})}\left(d_{(1:n)}\right)\prod\nolimits^{\otimes}_{j=1:m}\mathsf{E}\prod\nolimits^{\otimes}_{k\in\boldsymbol{b}_j}\boldsymbol{X}_k,\end{aligned}\quad(3.5)$$

where the second summation is taken over all possible ordered partition $\mathcal{L}\in\mathcal{P}_{(1:n)}$ with $|\mathcal{L}| = m$, see Remark 3.1 for details. The expectation operator E is defined in a way such that $\mathsf{E}(X_1, X_2) = (\mathsf{E}X_1, \mathsf{E}X_2)$.

3.5. Moments in terms of cumulants. Let $f(x) = \exp x$ and $g(\boldsymbol{\lambda}) = \psi_{\mathbf{X}_{(1:n)}}\left(\lambda_{(1:n)}\right)$. Hence, all the derivatives of $f(x) = \exp x$ *are* equal to $\exp x$, and therefore, we have

$$\frac{\partial^n \exp\left(g(\boldsymbol{\lambda})\right)}{\partial\lambda_1\partial\lambda_2\ldots\partial\lambda_n} = \exp\left(g(\boldsymbol{\lambda})\right)\sum_{\mathcal{K}\in\mathcal{P}_n}\boldsymbol{K}^{-1}_{\mathfrak{p}(\mathcal{K})}\left(d_{1:n}\right)\prod\nolimits^{\otimes}_{\boldsymbol{b}\in\mathcal{K}}\left(D^{\otimes|\boldsymbol{b}|}_{\boldsymbol{\lambda}_{\boldsymbol{b}}}g(\boldsymbol{\lambda})\right).\quad(3.6)$$

The expression for the moment $\mathsf{E}\mathbf{X}^{\otimes\mathbf{1}_{[n]}}_{(1:n)}$ is quite general. For example, the moment $\mathsf{E}\mathbf{Y}^{\otimes k_{(1:m)}}_{(1:m)}$ can be obtained from $\mathsf{E}\mathbf{X}^{\otimes\mathbf{1}_{[n]}}_{(1:n)}$, where

$$(\boldsymbol{Y}_{1_{[k_1]}},\ldots,\boldsymbol{Y}_{m_{[k_m]}}) = (\underbrace{\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_1}_{k_1},\ldots,\underbrace{\boldsymbol{Y}_m,\ldots,\boldsymbol{Y}_m}_{k_m}) = \mathbf{X}_{(1:n)},\ \text{say},$$

i.e., the elements in the product $\mathbf{Y}^{\otimes k_{(1:m)}}_{(1:m)}$ are treated as distinct.

$$\mathsf{E}\mathbf{X}^{\otimes\mathbf{1}_{[n]}}_{(1:n)} = \sum_{\mathcal{L}\in\mathcal{P}_{(1:n)}}\boldsymbol{K}^{-1}_{\mathfrak{p}(\mathcal{L})}\left(d_{(1:n)}\right)\prod\nolimits^{\otimes}_{\boldsymbol{b}\in\mathcal{L}}\mathsf{Cum}_{|\boldsymbol{b}|}(\boldsymbol{X}_{\boldsymbol{b}}),\quad(3.7)$$

where the summation is over all ordered partitions $\mathcal{L} = \{\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_k\}$ of $(1:n)$.

3.6. Cumulant of products via products of cumulants. Let $\mathbf{X}_{\mathcal{K}}$ denote the vector where the entries are obtained from the partition $\mathcal{K}$, i.e., if $\mathcal{K} = \{\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_m\}$, then $\mathbf{X}_{\mathcal{K}} = \left(\prod^{\otimes} \mathbf{X}_{\boldsymbol{b}_1}, \prod^{\otimes} \mathbf{X}_{\boldsymbol{b}_2}, \ldots, \prod^{\otimes} \mathbf{X}_{\boldsymbol{b}_m}\right)$. The order of the elements of the subsets $\boldsymbol{b} \in \mathcal{K}$ and the order of the subsets in $\mathcal{K}$ are fixed. Now the cumulant of the products can be expressed by the cumulants of the individual set of variables $\mathbf{X}_{\boldsymbol{b}} = (\boldsymbol{X}_j, j \in \boldsymbol{b})$, $\boldsymbol{b} \in \mathcal{L}$ such that $\mathcal{K} \cup \mathcal{L} = \mathcal{O}$, where $\mathcal{O}$ denotes the coarsest partition with one subset $\{(1:n)\}$ only. Such partitions $\mathcal{L}$ and $\mathcal{K}$ are called indecomposable (see Brillinger, 2001, Terdik, 1999).

$$\begin{aligned}
&\mathsf{Cum}_k \left(\left(\prod^{\otimes} \mathbf{X}_{\boldsymbol{b}_1}, \prod^{\otimes} \mathbf{X}_{\boldsymbol{b}_2}, \ldots, \prod^{\otimes} \mathbf{X}_{\boldsymbol{b}_m}\right)\right) \\
&= \sum_{\mathcal{K} \cup \mathcal{L} = \mathcal{O}} \boldsymbol{K}^{-1}_{\mathfrak{p}(\mathcal{L})} \left(d_{(1:n)}\right) \prod\nolimits^{\otimes}_{\boldsymbol{b} \in \mathcal{L}} \mathsf{Cum}_{|\boldsymbol{b}|}(\mathbf{X}_{\boldsymbol{b}}),
\end{aligned} \tag{3.8}$$

where $\mathbf{X}_{\boldsymbol{b}}$ denotes the set of vectors from $\boldsymbol{X}_s$, $s \in \boldsymbol{b}$.

EXAMPLE 3.2. Let $\boldsymbol{X}$ be a Gaussian random vector with $\mathsf{E}\boldsymbol{X} = 0$, $\boldsymbol{A}$ and $\boldsymbol{B}$ be matrices with appropriate dimensions and $\operatorname{Cov}(\boldsymbol{X}, \boldsymbol{X}) = \boldsymbol{\Sigma}$. Then,

$$\operatorname{Cum}\left(\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}, \boldsymbol{X}'\boldsymbol{B}\boldsymbol{X}\right) = 2 \operatorname{Tr} \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{B}'\boldsymbol{\Sigma} \tag{3.9}$$

(see Taniguchi, 1991). We can use (3.8) to obtain (3.9) as follows:

$$\begin{aligned}
&\operatorname{Cum}\left(\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}, \boldsymbol{X}'\boldsymbol{B}\boldsymbol{X}\right) \\
&= \mathsf{Cum}_2 \left((\operatorname{Vec} \boldsymbol{A})' \boldsymbol{X} \otimes \boldsymbol{X}, (\operatorname{Vec} \boldsymbol{B})' \boldsymbol{X} \otimes \boldsymbol{X}\right] \\
&= \left[(\operatorname{Vec} \boldsymbol{A})' \otimes (\operatorname{Vec} \boldsymbol{B})'\right] \mathsf{Cum}_2 \left(\boldsymbol{X} \otimes \boldsymbol{X}, \boldsymbol{X} \otimes \boldsymbol{X}\right) \\
&= \left[(\operatorname{Vec} \boldsymbol{A})' \otimes (\operatorname{Vec} \boldsymbol{B})'\right] \left[\boldsymbol{K}^{-1}_{2\leftrightarrow 3} + \boldsymbol{K}^{-1}_{2\leftrightarrow 4}\right] \left[\operatorname{Vec} \boldsymbol{\Sigma} \otimes \operatorname{Vec} \boldsymbol{\Sigma}\right] \\
&= 2 \left(\operatorname{Vec} \boldsymbol{\Sigma}\right)' \left(\boldsymbol{a} \otimes \boldsymbol{B}\right) \operatorname{Vec} \boldsymbol{\Sigma} = 2 \operatorname{Tr} \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{B}'\boldsymbol{\Sigma}.
\end{aligned}$$

4 Applications to Statistical Inference

4.1. Cumulants of the log-likelihood function. The above results can be used to obtain the cumulants of the partial derivatives of the log-likelihood function, see Skovgaard (1986). These expressions are useful in the study of the asymptotic theory of statistics.

Consider a random sample $(X_1, X_2, \ldots, X_N) = \boldsymbol{X} \in \mathbb{R}^N$ with the likelihood function $L(\boldsymbol{\vartheta}, X)$, and let $l(\boldsymbol{\vartheta})$ denote the log–likelihood function, i.e.,

$$l(\boldsymbol{\vartheta}) = \ln L(\boldsymbol{\vartheta}, X), \ \boldsymbol{\vartheta} \in \mathbb{R}^d.$$

It is well known that under the regularity conditions, we have

$$\mathsf{E}\frac{\partial l(\boldsymbol{\vartheta})}{\partial \vartheta_1}\frac{\partial l(\boldsymbol{\vartheta})}{\partial \vartheta_2} = -\mathsf{E}\frac{\partial^2 l(\boldsymbol{\vartheta})}{\partial \vartheta_1 \partial \vartheta_2}. \tag{4.1}$$

The result (4.1) can be extended to products of several partial derivatives (for $d = 4$, McCullagh and Cox (1986) use these expressions in the evaluation of Bartlett's correction). We can arrive at the result (4.1) from (3.4) by observing $L(\boldsymbol{\vartheta}, X) = e^{l(\boldsymbol{\vartheta})}$. For$d = 2$, we have

$$\frac{\partial^2 e^{l(\boldsymbol{\vartheta})}}{\partial \vartheta_1 \partial \vartheta_2} = \frac{\partial^2 L(\boldsymbol{\vartheta}, X)}{\partial \vartheta_1 \partial \vartheta_2},$$

and from (3.4), we have

$$\frac{\partial^2 e^{l(\boldsymbol{\vartheta})}}{\partial \vartheta_1 \partial \vartheta_2} = e^{l(\boldsymbol{\vartheta})}\left[\frac{\partial l(\boldsymbol{\vartheta})}{\partial \vartheta_1}\frac{\partial l(\boldsymbol{\vartheta})}{\partial \vartheta_2} + \frac{\partial^2 l(\boldsymbol{\vartheta})}{\partial \vartheta_1 \partial \vartheta_2}\right].$$

Equating the above two expressions, we get

$$\frac{1}{L(\boldsymbol{\vartheta}, X)}\frac{\partial^2 L(\boldsymbol{\vartheta}, X)}{\partial \vartheta_1 \partial \vartheta_2} = \frac{\partial l(\boldsymbol{\vartheta})}{\partial \vartheta_1}\frac{\partial l(\boldsymbol{\vartheta})}{\partial \vartheta_2} + \frac{\partial^2 l(\boldsymbol{\vartheta})}{\partial \vartheta_1 \partial \vartheta_2}.$$

The expected value of the left hand side of the above expression is zero, as interchange of the order of the derivative and the integral is allowed. This gives the result (4.1). The same argument leads, more generally, to a result involving several partial derivatives:

$$\sum_{r=1}^{d} \sum_{\substack{\mathcal{K} \in \mathcal{P}_d \\ |\mathcal{K}| = r}} \mathsf{E} \prod_{\boldsymbol{b} \in \mathcal{K}} \left[\frac{\partial^{|\boldsymbol{b}|}}{\prod_{j \in \boldsymbol{b}} \partial \vartheta_j} l(\boldsymbol{\vartheta})\right] = 0. \tag{4.2}$$

This result is a consequence of (3.6). Proceeding in a similar fashion, assuming the regularity conditions in higher-order and using (3.7), we obtain the cumulant analogue of the above as

$$\sum_{r=1}^{d} \sum_{\substack{\mathcal{K} \in \mathcal{P}_d \\ |\mathcal{K}| = r}} \operatorname{Cum}\left(\frac{\partial^{|\boldsymbol{b}|}}{\prod_{j \in \boldsymbol{b}} \partial \vartheta_j} l(\boldsymbol{\vartheta}), \boldsymbol{b} \in \mathcal{K}\right) = 0. \tag{4.3}$$

The equation (4.2) is in terms of the expected values of the derivatives of the log–likelihood function, whereas (4.3) is in terms of the cumulants. For example, suppose that we have a single parameter ϑ, and let us denote

$$\mu_4(m_1, m_2, m_3, m_4) = \mathsf{E}\left[\frac{\partial}{\partial \vartheta} l(\vartheta)\right]^{m_1}\left[\frac{\partial^2}{\partial \vartheta^2} l(\vartheta)\right]^{m_2}\left[\frac{\partial^3}{\partial \vartheta^3} l(\vartheta)\right]^{m_3}\left[\frac{\partial^4}{\partial \vartheta^4} l(\vartheta)\right]^{m_4},$$

then from the formula (4.2) we obtain

$$\mu_4(0,0,0,1)+4\mu_4(1,0,1,0)+6\mu_4(2,1,0,0)+3\mu_4(0,2,0,0)+\mu_4(4,0,0,0)=0. \tag{4.4}$$

To obtain (4.4), we proceed as follows. Consider the partitions $\mathcal{K}\in\mathcal{P}_4$. If $|\mathcal{K}|=1$, we have only one partition $(1,2,3,4)$, if $|\mathcal{K}|=2$, we have 4 terms of type $\{(1,2,3),(4)\}$ and 3 terms of type $\{(1,2),(4,3)\}$, if $|\mathcal{K}|=3$, we have 6 terms of the type $\{(1),(2),(4,3)\}$. Now if $\vartheta_1=\vartheta_2=\vartheta_3=\vartheta_4=\vartheta$, then m_1,m_2,m_3,m_4 show the numbers of the elements of the subsets in a partition. For instance, $(m_1,m_2,m_3,m_4)=(2,1,0,0)$ corresponds to the partitions of the type $\{(1),(2),(4,3)\}$ and so on. Hence the result (4.4). McCullagh and Cox (1986, eqn. (10), p. 142) obtained a similar result for cumulants:

$$\begin{aligned}\operatorname{Cum}\left(\frac{\partial^4}{\partial\vartheta^4}l(\vartheta)\right)&+4\operatorname{Cum}\left(\frac{\partial}{\partial\vartheta}l(\vartheta),\frac{\partial^2}{\partial\vartheta^2}l(\vartheta)\right)\\&+6\operatorname{Cum}\left(\frac{\partial}{\partial\vartheta}l(\vartheta),\frac{\partial}{\partial\vartheta}l(\vartheta),\frac{\partial^2}{\partial\vartheta^2}l(\vartheta)\right)\\&+3\operatorname{Cum}\left(\frac{\partial^2}{\partial\vartheta^2}l(\vartheta),\frac{\partial^2}{\partial\vartheta^2}l(\vartheta)\right)\\&+\operatorname{Cum}\left(\frac{\partial}{\partial\vartheta}l(\vartheta),\frac{\partial}{\partial\vartheta}l(\vartheta),\frac{\partial}{\partial\vartheta}l(\vartheta),\frac{\partial}{\partial\vartheta}l(\vartheta)\right)\\&=0,\end{aligned} \tag{4.5}$$

which is a special case of (4.3).

4.2. Cumulants of the log-likelihood function, the multiple parameter case. The multivariate extension (i.e., when the elements of the parameter vector are vectors as well) of the formula (4.2) can easily be obtained using Lemma 1. If we partition the vector parameters into n subsets, $\boldsymbol{\vartheta}=\vartheta_{(1:n)}=[\boldsymbol{\vartheta}_1',\boldsymbol{\vartheta}_2',\ldots,\boldsymbol{\vartheta}_n']'$ with dimensions $[d_1,d_2,\ldots,d_n]$ respectively, then it follows that

$$\sum_{r=1}^{d}\sum_{\substack{\mathcal{K}\in\mathcal{P}_d\\|\mathcal{K}|=r}}\boldsymbol{K}^{-1}_{\mathfrak{p}(\mathcal{K})}(d_{1:n})\,\mathsf{E}\prod\nolimits^{\otimes}_{\boldsymbol{b}\in\mathcal{K}}\left(D^{\otimes|\boldsymbol{b}|}_{\vartheta_{\boldsymbol{b}}}l(\boldsymbol{\vartheta})\right)=0, \tag{4.6}$$

where $\vartheta_{\boldsymbol{b}}$ denotes the subset of vectors $[\boldsymbol{\vartheta}_j,j\in\boldsymbol{b}]$. In particular, if $n=2$ and $\boldsymbol{\vartheta}_1=\boldsymbol{\vartheta}_2=\boldsymbol{\vartheta}$, then (4.6) gives the well known result

$$\operatorname{Cov}\left(D_{\boldsymbol{\vartheta}}l(\boldsymbol{\vartheta}),D_{\boldsymbol{\vartheta}}l(\boldsymbol{\vartheta})\right)=-\mathsf{E}\left(D_{\boldsymbol{\vartheta}}D^{\otimes}_{\boldsymbol{\vartheta}}l(\boldsymbol{\vartheta})\right).$$

In vectorized form, the same can be written as

$$\mathsf{E}\left(D_{\boldsymbol{\vartheta}}^{\otimes} l\left(\boldsymbol{\vartheta}\right) \otimes D_{\boldsymbol{\vartheta}}^{\otimes} l\left(\boldsymbol{\vartheta}\right)\right) = -\mathsf{E}\left(D_{\boldsymbol{\vartheta}}^{\otimes 2} l\left(\boldsymbol{\vartheta}\right)\right).$$

In the case $n = 4$, say and $\boldsymbol{\vartheta}_1 = \boldsymbol{\vartheta}_2 = \boldsymbol{\vartheta}_3 = \boldsymbol{\vartheta}_4 = \boldsymbol{\vartheta}$, we have

$$\boldsymbol{\mu}_4\left(0,0,0,1\right)+4\boldsymbol{\mu}_4\left(1,0,1,0\right)+6\boldsymbol{\mu}_4\left(2,1,0,0\right)+3\boldsymbol{\mu}_4\left(0,2,0,0\right)+\boldsymbol{\mu}_4\left(4,0,0,0\right) = 0,$$

where

$$\begin{aligned}&\boldsymbol{\mu}_4\left(m_1,m_2,m_3,m_4\right)\\ &= \mathsf{E}\left[D_{\boldsymbol{\vartheta}}^{\otimes} l\left(\boldsymbol{\vartheta}\right)\right]^{\otimes m_1} \otimes \left[D_{\boldsymbol{\vartheta}}^{\otimes 2} l\left(\boldsymbol{\vartheta}\right)\right]^{\otimes m_2} \otimes \left[D_{\boldsymbol{\vartheta}}^{\otimes 2} l\left(\boldsymbol{\vartheta}\right)\right]^{\otimes m_3} \otimes \left[D_{\boldsymbol{\vartheta}}^{\otimes 2} l\left(\boldsymbol{\vartheta}\right)\right]^{\otimes m_4}.\end{aligned}$$

We can obtain a similar expression for the cumulants, and it is given by

$$\sum_{r=1}^{d} \sum_{\substack{\mathcal{K}\in\mathcal{P}_d\\ |\mathcal{K}|=r}} \boldsymbol{K}_{\mathfrak{p}(\mathcal{K})}^{-1}\left(d_{1:n}\right) \mathsf{Cum}_r\left(D_{\boldsymbol{\vartheta}_{\boldsymbol{b}}}^{\otimes|\boldsymbol{b}|} l\left(\boldsymbol{\vartheta}\right), \boldsymbol{b}\in\mathcal{K}\right) = 0.$$

4.3. Multivariate measures of skewness and kurtosis for random vectors. In this section, we define what we consider to be natural measures of multivariate skewness and kurtosis and show their relation to the measures defined by Mardia (1970). Let $\boldsymbol{X}$ be a d-dimensional random vector whose first four moments exist. Let $\boldsymbol{\Sigma}$ denote the positive-definite variance covariance matrix. The *"skewness vector"* of $\boldsymbol{X}$ is defined by

$$\begin{aligned}\boldsymbol{\zeta}_{\boldsymbol{X}} &= \mathsf{Cum}_3\left(\boldsymbol{\Sigma}^{-1/2}\boldsymbol{X}, \boldsymbol{\Sigma}^{-1/2}\boldsymbol{X}, \boldsymbol{\Sigma}^{-1/2}\boldsymbol{X}\right)\\ &= \left(\boldsymbol{\Sigma}^{-1/2}\right)^{\otimes 3} \mathsf{Cum}_3\left(\boldsymbol{X},\boldsymbol{X},\boldsymbol{X}\right),\end{aligned}$$

and the *"total skewness"* is

$$\zeta_{\boldsymbol{X}} = \left\|\boldsymbol{\zeta}_{\boldsymbol{X}}\right\|^2.$$

The *"kurtosis vector"* of $\boldsymbol{X}$ is defined by

$$\begin{aligned}\boldsymbol{\kappa}_{\boldsymbol{X}} &= \mathsf{Cum}_4\left(\boldsymbol{\Sigma}^{-1/2}\boldsymbol{X}, \boldsymbol{\Sigma}^{-1/2}\boldsymbol{X}, \boldsymbol{\Sigma}^{-1/2}\boldsymbol{X}, \boldsymbol{\Sigma}^{-1/2}\boldsymbol{X}\right)\\ &= \left(\boldsymbol{\Sigma}^{-1/2}\right)^{\otimes 4} \mathsf{Cum}_4\left(\boldsymbol{X},\boldsymbol{X},\boldsymbol{X},\boldsymbol{X}\right),\end{aligned}$$

and the *"total kurtosis"* is

$$\kappa_{\boldsymbol{X}} = \operatorname{Tr}\left(\operatorname{Vec}^{-1}\boldsymbol{\kappa}_{\boldsymbol{X}}\right),$$

where $\operatorname{Vec}^{-1}\boldsymbol{\kappa}_{\boldsymbol{X}}$ is the matrix $\mathbf{M}$ such that $\operatorname{Vec}\mathbf{M} = \boldsymbol{\kappa}_{\boldsymbol{X}}$. The skewness and the kurtosis for a multivariate Gaussian vector $\boldsymbol{X}$ is zero. $\boldsymbol{\zeta}_{\boldsymbol{X}}$ is also zero for any distribution which is symmetric. The skewness and the kurtosis are expressed in terms of the moments. Suppose that $\mathsf{E}\boldsymbol{X} = 0$. Then

$$\boldsymbol{\zeta}_{\boldsymbol{X}} = \left(\boldsymbol{\Sigma}^{-1/2}\right)^{\otimes 3} \mathsf{E}\boldsymbol{X}^{\otimes 3}. \tag{4.7}$$

The total skewness $\zeta_{\boldsymbol{X}}$, which is just the norm square of the skewness vector $\boldsymbol{\zeta}_{\boldsymbol{X}}$, coincides with the measure of skewness $\beta_{1,d}$ defined by Mardia (1970). For any set of random vectors, we have

$$\begin{aligned}\mathsf{Cum}_4\left(\boldsymbol{X}_{1:4}\right) = \mathsf{E}\prod\nolimits^{\otimes} \boldsymbol{X}_{1:4} &- \mathsf{Cum}_2\left(\boldsymbol{X}_1, \boldsymbol{X}_2\right) \otimes \mathsf{Cum}_2\left(\boldsymbol{X}_3, \boldsymbol{X}_4\right) \\ &- \boldsymbol{K}^{-1}_{\mathfrak{p}_{2\leftrightarrow 3}}\left(d_{[4]}\right) \mathsf{Cum}_2\left(\boldsymbol{X}_1, \boldsymbol{X}_3\right) \otimes \mathsf{Cum}_2\left(\boldsymbol{X}_2, \boldsymbol{X}_4\right) \\ &- \boldsymbol{K}^{-1}_{\mathfrak{p}_{4\leftrightarrow 2}}\left(d_{[4]}\right) \mathsf{Cum}_2\left(\boldsymbol{X}_1, \boldsymbol{X}_4\right) \otimes \mathsf{Cum}_2\left(X_2, \boldsymbol{X}_3\right),\end{aligned} \tag{4.8}$$

and therefore the kurtosis vector of $\boldsymbol{X}$ can be expressed in terms of the fourth order moments. By putting $\boldsymbol{X}_1 = \boldsymbol{X}_2 = \boldsymbol{X}_3 = \boldsymbol{X}_4 = \boldsymbol{X}$ in the above,

$$\begin{aligned}\boldsymbol{\kappa}_{\boldsymbol{X}} &= \left(\boldsymbol{\Sigma}^{-1/2}\right)^{\otimes 4} \mathsf{Cum}_4\left(\boldsymbol{X}, \boldsymbol{X}, \boldsymbol{X}, \boldsymbol{X}\right) \\ &= \left(\boldsymbol{\Sigma}^{-1/2}\right)^{\otimes 4} \mathsf{E}\boldsymbol{X}^{\otimes 4} - \left(\boldsymbol{I} + \boldsymbol{K}^{-1}_{\mathfrak{p}_{2\leftrightarrow 3}}\left(d_{[4]}\right) + \boldsymbol{K}^{-1}_{\mathfrak{p}_{4\leftrightarrow 2}}\left(d_{[4]}\right)\right) \\ &\quad \times \left(\boldsymbol{\Sigma}^{-1/2}\right)^{\otimes 4} \mathsf{Cum}_2\left(\boldsymbol{X}, \boldsymbol{X}\right) \otimes \mathsf{Cum}_2\left(\boldsymbol{X}, \boldsymbol{X}\right) \\ &= \left(\boldsymbol{\Sigma}^{-1/2}\right)^{\otimes 4} \mathsf{E}\boldsymbol{X}^{\otimes 4} - \left(\boldsymbol{I} + \boldsymbol{K}^{-1}_{\mathfrak{p}_{2\leftrightarrow 3}}\left(d_{[4]}\right) + \boldsymbol{K}^{-1}_{\mathfrak{p}_{4\leftrightarrow 2}}\left(d_{[4]}\right)\right) \left[\operatorname{Vec}\boldsymbol{I}_d \otimes \operatorname{Vec}\boldsymbol{I}_d\right].\end{aligned} \tag{4.9}$$

Mardia (1970) defined the measure of kurtosis as

$$\beta_{2,d} = \mathsf{E}\left(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)^2,$$

and this is related to our total kurtosis measure $\kappa_{\boldsymbol{X}}$ as follows

$$\beta_{2,d} = \kappa_{\boldsymbol{X}} + d(d+2) = \operatorname{Tr}\left(\operatorname{Vec}^{-1}\boldsymbol{\kappa}_{\boldsymbol{X}}\right) + d(d+2).$$

Indeed

$$\begin{aligned}&\operatorname{Tr}\left(\operatorname{Vec}^{-1}\left[\left(\boldsymbol{\Sigma}^{-1/2}\right)^{\otimes 4} \mathsf{E}\boldsymbol{X}^{\otimes 4}\right]\right) \\ &\quad = \mathsf{E}\operatorname{Tr}\left(\left[\boldsymbol{\Sigma}^{-1/2}\boldsymbol{X}\right]^{\otimes 2}\left[\boldsymbol{\Sigma}^{-1/2}\boldsymbol{X}\right]^{\prime\otimes 2}\right) \\ &\quad = \mathsf{E}\operatorname{Tr}\left(\left[\left(\boldsymbol{\Sigma}^{-1/2}\boldsymbol{X}\right)'\left(\boldsymbol{\Sigma}^{-1/2}\boldsymbol{X}\right)\right]^{\otimes 2}\right) = \mathsf{E}\left(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)^2.\end{aligned}$$

We note that if $\boldsymbol{X}$ is Gaussian, then $\boldsymbol{\kappa}_{\boldsymbol{X}} = 0$, and hence $\beta_{2,d} = d(d+2)$.

4.4. Multiple linear time series. Let $\boldsymbol{X}_t$ be a d dimensional discrete time stationary time series. Let $\boldsymbol{X}_t$ satisfy the linear representation (see Hannan, 1970, p. 208)

$$\boldsymbol{X}_t = \sum_{k=0}^{\infty} \boldsymbol{H}(k)\, \boldsymbol{e}_{t-k}, \tag{4.10}$$

where $\boldsymbol{H}(0)$ is identity, $\sum \|\boldsymbol{H}(k)\| < \infty$, the $\boldsymbol{e}_t$'s are independent and identically distributed random vectors with $\mathrm{E}\boldsymbol{e}_t = 0$, $\mathrm{E}\boldsymbol{e}_t\boldsymbol{e}_t' = \boldsymbol{\Sigma}$. Let $\boldsymbol{\kappa}_{m+1}(\boldsymbol{e}) = \mathsf{Cum}_{m+1}(\boldsymbol{e}_t, \boldsymbol{e}_t, \ldots, \boldsymbol{e}_t)$ be the vector $d^{m+1} \times 1$. We note that $\boldsymbol{\kappa}_2(\boldsymbol{e}) = \operatorname{Vec} \boldsymbol{\Sigma}$, and the cumulant of $\boldsymbol{X}_t$ is

$$\begin{aligned}
&\mathsf{Cum}_{m+1}(\boldsymbol{X}_t, \boldsymbol{X}_{t+\tau_1}, \boldsymbol{X}_{t+\tau_2}, \ldots, \boldsymbol{X}_{t+\tau_m}) \\
&\quad = \sum_{k=0}^{\infty} \boldsymbol{H}(k) \otimes \boldsymbol{H}(k+\tau_1) \otimes \cdots \otimes \boldsymbol{H}(k+\tau_m)\, \boldsymbol{\kappa}_{m+1}(\boldsymbol{e}) \qquad (4.11) \\
&\quad = \boldsymbol{C}_{m+1}(\tau_1, \tau_2, \ldots, \tau_m).
\end{aligned}$$

Let $\boldsymbol{X}_t$ satisfy the autoregressive model of order p given by

$$\boldsymbol{X}_t + \boldsymbol{A}_1 \boldsymbol{X}_{t-1} + \boldsymbol{A}_2 \boldsymbol{X}_{t-2} + \cdots + \boldsymbol{A}_p \boldsymbol{X}_{t-p} = \boldsymbol{e}_t,$$

which can be written as

$$\left(\boldsymbol{I} + \boldsymbol{A}_1 B + \boldsymbol{A}_2 B^2 + \cdots + \boldsymbol{A}_p B^p\right) \boldsymbol{X}_t = \boldsymbol{e}_t,$$

where B is the backshift operator. We assume that the coefficients $\{\boldsymbol{A}_j\}$ satisfy the usual stationarity condition (see Hannan, 1970, p. 212) and note that

$$\boldsymbol{X}_t = \left(\boldsymbol{I} + \boldsymbol{A}_1 B + \boldsymbol{A}_2 B^2 + \cdots + \boldsymbol{A}_p B^p\right)^{-1} \boldsymbol{e}_t = \left(\sum_{k=0}^{\infty} \boldsymbol{H}(k)\, B^k\right) \boldsymbol{e}_t. \tag{4.12}$$

From (4.10) and (4.12), we have

$$\left(\boldsymbol{I} + \boldsymbol{A}_1 B + \boldsymbol{A}_2 B^2 + \cdots + \boldsymbol{A}_p B^p\right) \left(\sum_{k=0}^{\infty} \boldsymbol{H}(k)\, B^k\right) = \boldsymbol{I}, \tag{4.13}$$

from which we obtain

$$\begin{aligned}
&\boldsymbol{H}(0) + (\boldsymbol{H}(1) + \boldsymbol{A}_1 \boldsymbol{H}(0))\, B + (\boldsymbol{H}(2) + \boldsymbol{A}_1 \boldsymbol{H}(1) + \boldsymbol{A}_2 \boldsymbol{H}(0))\, B^2 + \cdots \\
&+ (\boldsymbol{H}(p) + \boldsymbol{A}_1 \boldsymbol{H}(p-1) + \boldsymbol{A}_2 \boldsymbol{H}(p-2) + \cdots + \boldsymbol{A}_p \boldsymbol{H}(0))\, B^p + \cdots \\
&+ \boldsymbol{H}(p+1) + \boldsymbol{A}_1 \boldsymbol{H}(p) + \cdots + \\
&\qquad = \boldsymbol{I}.
\end{aligned}$$

Equating powers of B^j for $j \geq 1$, we get

$$\boldsymbol{H}(j) + \boldsymbol{A}_1 \boldsymbol{H}(j-1) + \boldsymbol{A}_2 \boldsymbol{H}(j-2) + \cdots + \boldsymbol{A}_p \boldsymbol{H}(j-p) = 0, \ j \geq 1 \quad (4.14)$$

(here we use the convention $\boldsymbol{H}(j) = 0$, if $j < 0$). A recursive formula for $\boldsymbol{H}(j)$ follows from (4.14). By substituting for $\boldsymbol{H}(k+\tau_1)$ from this formula into (4.11), for $\tau_1 \geq 1$, we get

$$\begin{aligned}
&\boldsymbol{C}_{m+1}(\tau_1, \tau_2, \ldots, \tau_m) \\
&= -\sum_{k=0}^{\infty} \boldsymbol{H}(k) \otimes [\boldsymbol{A}_1 \boldsymbol{H}(k+\tau_1-1) + \boldsymbol{A}_2 \boldsymbol{H}(k+\tau_1-2) + \cdots + \boldsymbol{A}_p \boldsymbol{H}(k+\tau_1-p)] \\
&\qquad\qquad \otimes \cdots \otimes \boldsymbol{H}(k+\tau_m) \boldsymbol{\kappa}_{m+1}(\boldsymbol{e}) \\
&= -\sum_{j=1}^{p} \sum_{k=0}^{\infty} \boldsymbol{H}(k) \otimes \boldsymbol{A}_j \boldsymbol{H}(k+\tau_1-j) \otimes \boldsymbol{H}(k+\tau_2) \otimes \cdots \otimes \boldsymbol{H}(k+\tau_m) \boldsymbol{\kappa}_{m+1}(\boldsymbol{e}) \\
&= -\sum_{j=1}^{p} \sum_{k=0}^{\infty} [\boldsymbol{I}_d \otimes \boldsymbol{A}_j \otimes \boldsymbol{I}_{d^{m-1}}] [\boldsymbol{H}(k) \otimes \boldsymbol{H}(k+\tau_1-j) \otimes \boldsymbol{H}(k+\tau_2) \otimes \boldsymbol{H}(k+\tau_m)] \\
&\qquad\qquad \times \boldsymbol{\kappa}_{m+1}(\boldsymbol{e}) \\
&= \sum_{j=1}^{p} (\boldsymbol{I}_d \otimes \boldsymbol{A}_j \otimes \boldsymbol{I}_{d^{m-1}}) \boldsymbol{C}_{m+1}(\tau_{1-j}, \tau_2, \ldots, \tau_m).
\end{aligned}$$

Thus, we obtain

$$\boldsymbol{C}_{m+1}(\tau_1, \tau_2, \ldots, \tau_m) = -\sum_{j=1}^{p} (\boldsymbol{I}_d \otimes \boldsymbol{A}_j \otimes \boldsymbol{I}_{d^{m-1}}) \boldsymbol{C}_{m+1}(\tau_1 - j, \tau_2, \ldots, \tau_m). \quad (4.15)$$

If we put $m = 1$ in (4.15), we get

$$\boldsymbol{C}_2(\tau_1) = -\sum_{j=1}^{p} (\boldsymbol{I}_d \otimes \boldsymbol{A}_j) \boldsymbol{C}_2(\tau_1 - j),$$

which can be written in matrix form as

$$\boldsymbol{C}_2(\tau_1) = -\sum_{j=1}^{p} \boldsymbol{A}_j \mathbf{C}_2(\tau_1 - j),$$

which is the well known Yule-Walker equation in terms of second order covariances. Therefore, we can consider (4.15) as an extension of Yule-Walker equations in terms of higher-order cumulants for multivariate autoregressive models.

The definition of the higher-order cumulant spectra for stationary time series comes in a natural way. Consider the time series $\boldsymbol{X}_t$ with $(m+1)^{th}$ order cumulant function

$$\mathsf{Cum}_{m+1}\left(\boldsymbol{X}_t, \boldsymbol{X}_{t+\tau_1}, \boldsymbol{X}_{t+\tau_2}, \ldots, \boldsymbol{X}_{t+\tau_m}\right) = \boldsymbol{C}_{m+1}\left(\tau_1, \tau_2, \ldots, \tau_m\right),$$

and define the m^{th} order cumulant spectrum as the Fourier transform of the cumulants

$$\boldsymbol{S}_m\left(\omega_1, \omega_2, \ldots, \omega_m\right) = \sum_{\tau_1, \tau_2, \ldots, \tau_m = -\infty}^{\infty} \boldsymbol{C}_{m+1}\left(\tau_1, \tau_2, \ldots, \tau_m\right) \exp\left(-i \sum_{j=1}^{m} \tau_j \omega_j\right),$$

provided that the infinite sum converges. We note here that the connection between the usual matrix notation for the second order spectrum $\mathbf{S}_2(\omega)$ is given as

$$\boldsymbol{S}_2(\omega) = \operatorname{Vec}\left[\mathbf{S}_2(\omega)\right]',$$

see (3.2).

4.5. Bhattacharya-type lower bound for the multiparameter case. In this section, we obtain a lower bound for the variance covariance matrix of an unbiased vector of statistics, which is a linear function of the first k partial derivatives. This corresponds to the well known Bhattacharya bound (see Bhattacharya, 1946, Linnik, 1970) for the multiparameter case, which does not seem to have been considered anywhere in the literature. Consider a random sample $(\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n) = \boldsymbol{X} \in \mathbb{R}^{nd_0}$ with likelihood function $L(\boldsymbol{\vartheta}, \boldsymbol{X})$, $\boldsymbol{\vartheta} \in \mathbb{R}^d$. Suppose that we have a vector of unbiased estimators, say, $\widehat{\boldsymbol{g}}(\boldsymbol{X})$ of $\boldsymbol{g}(\boldsymbol{\vartheta}) \in \mathbb{R}^{d_1}$. Define the random vectors

$$\begin{aligned}\boldsymbol{\Upsilon}'_{Df} &= \Bigg(\frac{1}{L(\boldsymbol{\vartheta}, \boldsymbol{X})} D_{\boldsymbol{\vartheta}}^{\otimes} L(\boldsymbol{\vartheta}, \boldsymbol{X})', \frac{1}{L(\boldsymbol{\vartheta}, \boldsymbol{X})} D_{\boldsymbol{\vartheta}}^{\otimes 2} L(\boldsymbol{\vartheta}, \boldsymbol{X})', \\ &\qquad \ldots, \frac{1}{L(\boldsymbol{\vartheta}, \boldsymbol{X})} D_{\boldsymbol{\vartheta}}^{\otimes k} L(\boldsymbol{\vartheta}, \boldsymbol{X})'\Bigg), \\ \boldsymbol{\Upsilon}' &= \left(\widehat{\boldsymbol{g}}'(\boldsymbol{X}), \boldsymbol{\Upsilon}'_{Df}\right),\end{aligned}$$

where the dimension of $\boldsymbol{\Upsilon}$ is $d_1 + d + d^2 + \ldots + d^k$. The second order cumulant between $\widehat{\boldsymbol{g}}(\boldsymbol{X})$ and the derivatives $\frac{1}{L(\boldsymbol{\vartheta}, \boldsymbol{X})} D_{\boldsymbol{\vartheta}}^{\otimes j} L(\boldsymbol{\vartheta}, \boldsymbol{X})$, $j = 1, 2, \ldots, k$, is as follows

$$\begin{aligned}\mathsf{Cum}\left(\widehat{\boldsymbol{g}}\left(\boldsymbol{X}\right),\frac{1}{L\left(\boldsymbol{\vartheta},\boldsymbol{X}\right)}D_{\boldsymbol{\vartheta}}^{\otimes j}L\left(\boldsymbol{\vartheta},\boldsymbol{X}\right)\right)\\ &= \operatorname{Vec}\int\left[D_{\boldsymbol{\vartheta}}^{\otimes j}L\left(\boldsymbol{\vartheta},\boldsymbol{X}\right)\right]\widehat{\boldsymbol{g}}\left(\boldsymbol{X}\right)'d\boldsymbol{X}\\ &= \int\widehat{\boldsymbol{g}}\left(\boldsymbol{X}\right)\otimes D_{\boldsymbol{\vartheta}}^{\otimes j}L\left(\boldsymbol{\vartheta},\boldsymbol{X}\right)d\boldsymbol{X} \quad = \quad D_{\boldsymbol{\vartheta}}^{\otimes j}\boldsymbol{g}\left(\boldsymbol{\vartheta}\right).\end{aligned}$$

The covariance matrix between $\widehat{\boldsymbol{g}}\left(\boldsymbol{X}\right)$ and $\frac{1}{L\left(\boldsymbol{\vartheta},\boldsymbol{X}\right)}D_{\boldsymbol{\vartheta}}^{\otimes j}L\left(\boldsymbol{\vartheta},\boldsymbol{X}\right)$ is calculated using (3.2). The variance matrix $\operatorname{Var}\left(\boldsymbol{\Upsilon}_{Df}\right)$ is singular because the elements of the derivatives $D_{\boldsymbol{\vartheta}}^{\otimes j}L\left(\boldsymbol{\vartheta},\boldsymbol{X}\right)$ are not distinct. Therefore, we reduce the vector of derivatives using distinct elements only. To make it precise, we first consider second order derivatives. We define the duplication matrix $\mathfrak{D}_{2,d}$, which reduces the symmetric matrix V_d to the matrix $\boldsymbol{\nu}_2\left(V_d\right)$, which is the vector of lower triangular elements of V_d. We define $\mathfrak{D}_{2,d}$ as follows:

$$\mathfrak{D}_{2,d}\boldsymbol{\nu}_2\left(V_d\right) = \operatorname{Vec}V_d.$$

The dimension of $\boldsymbol{\nu}_2\left(V_d\right)$ is $d\left(d+1\right)/2$, and that of $\mathfrak{D}_{2,d}$ is $d^2\times d\left(d+1\right)/2$. It is easy to see that $\mathfrak{D}_{2,d}'\mathfrak{D}_{2,d}$ is non-singular (the columns of $\mathfrak{D}_{2,d}$ are linearly independent – each row has exactly one nonzero element). Therefore, the Moore-Penrose inverse $\mathfrak{D}_{2,d}^{+}$ of $\mathfrak{D}_{2,d}$ is

$$\mathfrak{D}_{2,d}^{+} = \left(\mathfrak{D}_{2,d}'\mathfrak{D}_{2,d}\right)^{-1}\mathfrak{D}_{2,d}'$$

such that

$$\boldsymbol{\nu}_2\left(V_d\right) = \mathfrak{D}_{2,d}^{+}\operatorname{Vec}V_d$$

(see Magnus and Neudecker, 1999, Ch. 3 Sec. 8, for details). The operator $D_{\boldsymbol{\vartheta}}^{\otimes 2}$ is defined by

$$D_{\boldsymbol{\vartheta}}^{\otimes 2} = \operatorname{Vec}\frac{\partial}{\partial\boldsymbol{\lambda}}\frac{\partial}{\partial\boldsymbol{\lambda}'},$$

which is actually $\left(\frac{\partial}{\partial\boldsymbol{\lambda}}\right)^{\otimes 2}$. The matrix $\frac{\partial}{\partial\boldsymbol{\lambda}}\frac{\partial}{\partial\boldsymbol{\lambda}'}$ is symmetric, and therefore we can use the inverse $\mathfrak{D}_{2,d}^{+}$ of the duplication matrix

$$\mathfrak{D}_{2,d}^{+}D_{\boldsymbol{\vartheta}}^{\otimes 2} = \boldsymbol{\nu}_2\left(D_{\boldsymbol{\vartheta}}^{\otimes 2}\right)$$

to get the necessary elements of the derivatives. We can extend this procedure for higher-order derivatives by defining

$$\mathfrak{D}_{k,d}^{+}D_{\boldsymbol{\vartheta}}^{\otimes k} = \boldsymbol{\nu}_k\left(D_{\boldsymbol{\vartheta}}^{\otimes k}\right),$$

where $\boldsymbol{\nu}_k\left(D_{\boldsymbol{\vartheta}}^{\otimes k}\right)$ is a vector of the distinct elements of $D_{\boldsymbol{\vartheta}}^{\otimes k}$ listed in the original order in $D_{\boldsymbol{\vartheta}}^{\otimes k}$. Now, let

$$\boldsymbol{C}_{g,j} = \operatorname{Cov}\left(\widehat{\boldsymbol{g}}\left(\boldsymbol{X}\right), \frac{1}{L\left(\boldsymbol{\vartheta}, \boldsymbol{X}\right)}\left(\mathfrak{D}_{j,d}^{+} D_{\boldsymbol{\vartheta}}^{\otimes j}\right) L\left(\boldsymbol{\vartheta}, \boldsymbol{X}\right)\right)$$

The elements of $\boldsymbol{C}_{g,j}$ are those of the cumulant (see (3.2))

$$\operatorname{Cum}\left(\widehat{\boldsymbol{g}}\left(\boldsymbol{X}\right), \frac{1}{L\left(\boldsymbol{\vartheta}, \boldsymbol{X}\right)}\left(\mathfrak{D}_{j,d}^{+} D_{\boldsymbol{\vartheta}}^{\otimes j}\right) L\left(\boldsymbol{\vartheta}, \boldsymbol{X}\right)\right).$$

Now, considering the vector of all distinct and nonzero derivatives,

$$\begin{aligned}\boldsymbol{\Upsilon}_{Df}' &= \left(\frac{1}{L\left(\boldsymbol{\vartheta}, \boldsymbol{X}\right)} D_{\boldsymbol{\vartheta}}^{\otimes} L\left(\boldsymbol{\vartheta}, \boldsymbol{X}\right)', \frac{1}{L\left(\boldsymbol{\vartheta}, \boldsymbol{X}\right)} \mathfrak{D}_{2,d}^{+} D_{\boldsymbol{\vartheta}}^{\otimes 2} L\left(\boldsymbol{\vartheta}, \boldsymbol{X}\right)', \right.\\ &\qquad\left. \ldots, \frac{1}{L\left(\boldsymbol{\vartheta}, \boldsymbol{X}\right)} \mathfrak{D}_{k,d}^{+} D_{\boldsymbol{\vartheta}}^{\otimes k} L\left(\boldsymbol{\vartheta}, \boldsymbol{X}\right)'\right),\\ \boldsymbol{\Upsilon}' &= \left(\widehat{\boldsymbol{g}}'\left(\boldsymbol{X}\right), \boldsymbol{\Upsilon}_{Df}'\right),\end{aligned}$$

we obtain the generalized Bhattacharya lower bound in the case of multiple parameters. This is obtained by considering the variance matrix of $\boldsymbol{\Upsilon}'$, whose positive semi-definiteness implies

$$\operatorname{Var}\left(\widehat{\boldsymbol{g}}\left(\boldsymbol{X}\right)\right) - \boldsymbol{C}_{g,Df} \operatorname{Var}\left(\boldsymbol{\Upsilon}_{Df}\right)^{-1} \boldsymbol{C}_{g,Df}' \geq 0 \tag{4.16}$$

with $\boldsymbol{C}_{g,Df} = \left[\boldsymbol{C}_{g,1}, \boldsymbol{C}_{g,2}, \ldots, \boldsymbol{C}_{g,k}\right]$. The Cramer- Rao inequality is obtained by setting $k = 1$, i.e., by considering only the first derivative vector.

Let us now consider an example to illustrate the Bhattacharya bound given by (4.16).

Example 4.1. Let $\left(\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n\right) = \boldsymbol{X} \in \mathbb{R}^{nd_0}$ be a sequence of independent Gaussian random vectors with mean vector $\boldsymbol{\vartheta} \in \mathbb{R}^{d_0}$, and variance matrix $\boldsymbol{I}_{d_0}$. Suppose that we want to estimate the function $g\left(\boldsymbol{\vartheta}\right) = \left\|\boldsymbol{\vartheta}\right\|^2 \in \mathbb{R}$. Here $d = d_0$, $d_1 = 1$. The unbiased estimator for $g\left(\boldsymbol{\vartheta}\right)$ is

$$\widehat{\boldsymbol{g}}\left(\boldsymbol{X}\right) = \sum_{k=1}^{d}\left(\overline{X}_k^2 - \frac{1}{n}\right),$$

where $\overline{X}_k$ is the sample mean computed using the random sample consisting of n observations on the k^{th} random variable of the random vector $\boldsymbol{X}$. The variance of the estimator $\widehat{\boldsymbol{g}}(\boldsymbol{X})$ is

$$\operatorname{Var}(\widehat{\boldsymbol{g}}(\boldsymbol{X}))=\sum_{k=1}^{d}\left(\frac{4\vartheta_k^2}{n}+\frac{2}{n^2}\right)=\frac{4}{n}\|\boldsymbol{\vartheta}\|^2+\frac{2d}{n^2}. \tag{4.17}$$

The Cramer-Rao bound for this estimator is $\frac{4}{n}\|\boldsymbol{\vartheta}\|^2$, which is strictly less than the actual variance. The derivatives $D_{\boldsymbol{\vartheta}}^{\otimes j}L(\boldsymbol{\vartheta},\boldsymbol{X})$ for $j>2$ are zero. For $j=1,2$, we have

$$\begin{aligned}D_{\boldsymbol{\vartheta}}^{\otimes}L(\boldsymbol{\vartheta},\boldsymbol{X})&=n\left(\overline{\boldsymbol{X}}-\boldsymbol{\vartheta}\right)L(\boldsymbol{\vartheta},\boldsymbol{X}),\\ D_{\boldsymbol{\vartheta}}^{\otimes 2}L(\boldsymbol{\vartheta},\boldsymbol{X})&=n^2\left[\left(\overline{\boldsymbol{X}}-\boldsymbol{\vartheta}\right)^{\otimes 2}-\frac{1}{n}\operatorname{Vec}\boldsymbol{I}_d\right]L(\boldsymbol{\vartheta},\boldsymbol{X}).\end{aligned}$$

Therefore, we obtain (using all the elements of second partial derivative matrix)

$$\begin{aligned}\widetilde{\boldsymbol{\Upsilon}}'_{Df}&=\left(\frac{1}{L(\boldsymbol{\vartheta},\boldsymbol{X})}D_{\boldsymbol{\vartheta}}^{\otimes}L(\boldsymbol{\vartheta},\boldsymbol{X})',\frac{1}{L(\boldsymbol{\vartheta},\boldsymbol{X})}D_{\boldsymbol{\vartheta}}^{\otimes 2}L(\boldsymbol{\vartheta},\boldsymbol{X})'\right)\\&=\left(n\left(\overline{\boldsymbol{X}}-\boldsymbol{\vartheta}\right)',n^2\left[\left(\overline{\boldsymbol{X}}-\boldsymbol{\vartheta}\right)^{\otimes 2}-\frac{1}{n}\operatorname{Vec}\boldsymbol{I}_d\right]'\right).\end{aligned}$$

Note that if we consider only the vector of first derivatives, then the second element of above vector will not be included in the lower bound, making the Cramer-Rao bound smaller. If we use the reduced number of elements for $\widetilde{\boldsymbol{\Upsilon}}'_{Df}$, we have

$$\boldsymbol{\Upsilon}'_{Df}=\left(n\left(\overline{\boldsymbol{X}}-\boldsymbol{\vartheta}\right)',n^2\left[\mathfrak{D}_{2,d}^{+}\left(\overline{\boldsymbol{X}}-\boldsymbol{\vartheta}\right)^{\otimes 2}-\frac{1}{n}\mathfrak{D}_{2,d}^{+}\operatorname{Vec}\boldsymbol{I}_d\right]'\right).$$

The variance matrix of $\boldsymbol{\Upsilon}_{Df}$ will contain the following matrix as a diagonal block:

$$\begin{aligned}n^2\boldsymbol{C}_2&=n^2\operatorname{Vec}_{d^2,d^2}^{-1}\left(\mathfrak{D}_{2,d}^{+}\right)^{\otimes 2}\operatorname{Cum}_2\left(\left[\left(\overline{\boldsymbol{X}}-\boldsymbol{\vartheta}\right)^{\otimes 2}-\frac{1}{n}\operatorname{Vec}\boldsymbol{I}_d\right]\right)\\&=\operatorname{Vec}_{d^2,d^2}^{-1}\left(\mathfrak{D}_{2,d}^{+}\right)^{\otimes 2}\left[\left(\boldsymbol{K}_{\mathfrak{p}_{2\leftrightarrow 3}}^{-1}\left(d_{[4]}\right)+\boldsymbol{K}_{\mathfrak{p}_{1\leftrightarrow 3}}^{-1}\left(d_{[4]}\right)\right)\left(\operatorname{Vec}\boldsymbol{I}_d\right)^{\otimes 2}\right]\\&=\mathfrak{D}_{2,d}^{+}\left[\boldsymbol{I}_{d^2}+\boldsymbol{K}_{\mathfrak{p}_{1\leftrightarrow 2}}\left(d_{[4]}\right)\right]\left(\mathfrak{D}_{2,d}^{+}\right)'.\end{aligned}$$

Denote

$$\frac{1}{2}\left[\boldsymbol{I}_{d^2}+\boldsymbol{K}_{\mathfrak{p}_{1\leftrightarrow 2}}\left(d_{[4]}\right)\right]=\mathbf{N}_d,$$

and then the matrices satisfy

$$\mathbf{N}_d = \mathbf{N}_d' = \mathbf{N}_d^2, \\ \text{and } \mathbf{N}_d = \mathfrak{D}_{2,d}\mathfrak{D}_{2,d}^+$$

(see Magnus and Neudecker, 1999, Ch. 3 Sec. 7-8, Theorem 11 and 12). We obtain

$$n^2 \boldsymbol{C}_2 = 2\mathfrak{D}_{2,d}^+ \mathbf{N}_d \left(\mathfrak{D}_{2,d}^+ \mathbf{N}_d\right)' = 2\mathfrak{D}_{2,d}^+ \left(\mathfrak{D}_{2,d}^+\right)' = 2\left(\mathfrak{D}_{2,d}'\mathfrak{D}_{2,d}\right)^{-1},$$

which is invertible. The inverse of the variance matrix of $\boldsymbol{\Upsilon}_{Df}$ is given by

$$[\operatorname{Var}(\boldsymbol{\Upsilon}_{Df})]^{-1} = \begin{bmatrix} \frac{1}{n}\boldsymbol{I}_d & 0 \\ 0 & \frac{1}{2n^2}\mathfrak{D}_{2,d}'\mathfrak{D}_{2,d} \end{bmatrix}.$$

Now, to obtain the matrix $\boldsymbol{C}_{g,Df} = [\boldsymbol{C}_{g,1}, \boldsymbol{C}_{g,2}]$, we need

$$\mathsf{Cum}\left(\widehat{\boldsymbol{g}}(\boldsymbol{X}), \frac{1}{L(\boldsymbol{\vartheta}, \boldsymbol{X})} D_{\boldsymbol{\vartheta}}^{\otimes} L(\boldsymbol{\vartheta}, \boldsymbol{X})\right) = D_{\boldsymbol{\vartheta}}^{\otimes} \boldsymbol{g}(\boldsymbol{\vartheta}) = D_{\boldsymbol{\vartheta}}^{\otimes} \boldsymbol{\vartheta}'\boldsymbol{\vartheta} = 2\boldsymbol{\vartheta}, \\ \boldsymbol{C}_{g,1} = 2\boldsymbol{\vartheta}',$$

and

$$\mathsf{Cum}\left(\widehat{\boldsymbol{g}}(\boldsymbol{X}), \frac{1}{L(\boldsymbol{\vartheta}, \boldsymbol{X})} \boldsymbol{\nu}\left(D_{\boldsymbol{\vartheta}}^{\otimes 2}\right) L(\boldsymbol{\vartheta}, \boldsymbol{X})\right) = 2\mathfrak{D}_{2,d}^+ \operatorname{Vec} \boldsymbol{I}_d \\ \boldsymbol{C}_{g,2}' = 2\mathfrak{D}_{2,d}^+ \operatorname{Vec} \boldsymbol{I}_d.$$

Finally, we obtain

$$\begin{aligned} &\boldsymbol{C}_{g,Df} \operatorname{Var}(\boldsymbol{\Upsilon}_{Df})^{-1} \boldsymbol{C}_{g,Df}' \\ &\quad = \frac{4}{n}\|\boldsymbol{\vartheta}\|^2 + \frac{2}{n^2}(\operatorname{Vec} \boldsymbol{I}_d)' \left(\mathfrak{D}_{2,d}\mathfrak{D}_{2,d}^+\right)' \mathfrak{D}_{2,d}\mathfrak{D}_{2,d}^+ \operatorname{Vec} \boldsymbol{I}_d \\ &\quad = \frac{4}{n}\|\boldsymbol{\vartheta}\|^2 + \frac{1}{n^2}(\operatorname{Vec} \boldsymbol{I}_d)' \mathbf{N}_d \operatorname{Vec} \boldsymbol{I}_d \\ &\quad = \frac{4}{n}\|\boldsymbol{\vartheta}\|^2 + \frac{2d}{n^2}, \end{aligned}$$

which is the Bhattacharya bound and is the same as the variance of the statistic $\widehat{\boldsymbol{g}}(\boldsymbol{X})$, given by (4.17).

Appendix

A.1. Commutation matrices. The Kronecker products have the advantage in the sense that we can commute the elements of the products using linear operators called commutation matrices (see Magnus and Neudecker, 1999, Ch. 3 Sec. 7, for details). We use these operators here in the case of vectors. Let $\boldsymbol{A}$ be a matrix of order $m \times n$, and the vector $\operatorname{Vec} \boldsymbol{A}'$ is a permutation of the vector $\operatorname{Vec}\boldsymbol{A}$. Therefore there exists a permutation matrix $\boldsymbol{K}_{m \cdot n}$ of order $mn \times mn$, called *commutation matrix,* which is defined by the relation

$$\boldsymbol{K}_{m \cdot n} \operatorname{Vec} \boldsymbol{A} = \operatorname{Vec} \boldsymbol{A}'.$$

Now, if $\boldsymbol{a}$ is $m \times 1$ and $\boldsymbol{b}$ is $n \times 1$, then

$$\boldsymbol{K}_{m \cdot n} (\boldsymbol{b} \otimes \boldsymbol{a}) = \boldsymbol{K}_{m \cdot n} \operatorname{Vec} \left(\boldsymbol{a}\boldsymbol{b}'\right) = \operatorname{Vec} \left(\boldsymbol{b}\boldsymbol{a}'\right) = \boldsymbol{a} \otimes \boldsymbol{b}.$$

From now on in the sequel, we shall use a more convenient notation

$$\boldsymbol{K}_{m \cdot n} = \boldsymbol{K} (n, m),$$

which means that we are changing the order in a K-product $\boldsymbol{b} \otimes \boldsymbol{a}$ of vectors $\boldsymbol{b} \in \mathbb{R}^n$ and $\boldsymbol{a} \in \mathbb{R}^m$.

Now, consider a set of vectors $(\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n)$ with dimensions $d_{(1:n)} = (d_1, d_2, \ldots, d_n)$ respectively. Define the matrix

$$\boldsymbol{K}_{j+1 \leftrightarrow j} \left(d_{(1:n)}\right) = \prod\nolimits_{i=1:j-1}^{\otimes} \boldsymbol{I}_{d_i} \otimes \boldsymbol{K} (d_j, d_{j+1}) \otimes \prod\nolimits_{i=j+2:n}^{\otimes} \boldsymbol{I}_{d_i},$$

where $\prod_{i=1:j}^{\otimes}$ stands for the Kronecker product of the matrices indexed by $1 : j = (1, 2, \ldots, j)$. Clearly,

$$\begin{aligned}
&\boldsymbol{K}_{j+1 \leftrightarrow j} (d_{1:n}) \prod\nolimits_{i=1:n}^{\otimes} \boldsymbol{a}_i \\
&= \prod\nolimits_{i=1:j-1}^{\otimes} (\boldsymbol{I}_{d_i} \boldsymbol{a}_i) \otimes (\boldsymbol{K} (d_j, d_{j+1}) (\boldsymbol{a}_j \otimes \boldsymbol{a}_{j+1})) \otimes \prod\nolimits_{i=j+2:n}^{\otimes} (\boldsymbol{I}_{d_i} \boldsymbol{a}_i) \\
&= \prod\nolimits_{i=1:j-1}^{\otimes} \boldsymbol{a}_i \otimes \boldsymbol{a}_{j+1} \otimes \boldsymbol{a}_j \otimes \prod\nolimits_{i=j+2:n}^{\otimes} \boldsymbol{a}_i.
\end{aligned}$$

Therefore, one is able to transpose (interchange) the elements $\boldsymbol{a}_j$ and $\boldsymbol{a}_{j+1}$ in a Kronecker product of vectors by the help of the matrix $\boldsymbol{K}_{j \leftrightarrow j+1} (d_{1:n})$. In general, $\boldsymbol{K}'_{j \leftrightarrow j+1} (d_{1:n}) = \boldsymbol{K}^{-1}_{j \leftrightarrow j+1} (d_{1:n})$ but $\boldsymbol{K}_{j+1 \leftrightarrow j} \neq \boldsymbol{K}_{j \leftrightarrow j+1}$ because the dimensions d_{j+1} and d_j are not necessarily equal. If they are equal, then $\boldsymbol{K}_{j+1 \leftrightarrow j} = \boldsymbol{K}_{j \leftrightarrow j+1} = \boldsymbol{K}^{-1}_{j \leftrightarrow j+1} = \boldsymbol{K}'_{j \leftrightarrow j+1}$. We recall that $\mathfrak{P}_n$ denotes the set of all permutations of the numbers $(1:n) = (1,2,\ldots,n)$. If $\mathfrak{p} \in \mathfrak{P}_n$ then $\mathfrak{p}(1:n) = (\mathfrak{p}(1), \mathfrak{p}(2), \ldots, \mathfrak{p}(n))$. From this, it follows that for each permutation $\mathfrak{p}(1:n) = (\mathfrak{p}(1), \mathfrak{p}(2), \ldots, \mathfrak{p}(n))$, $\mathfrak{p} \in \mathfrak{P}_\mathfrak{n}$, there exists a matrix $\boldsymbol{K}_{\mathfrak{p}(1:n)}(d_{1:n})$ such that

$$\boldsymbol{K}_{\mathfrak{p}(1:n)}\left(d_{(1:n)}\right) \prod\nolimits_{i=1:n}^{\otimes} \boldsymbol{a}_i = \prod\nolimits_{i=1:n}^{\otimes} \boldsymbol{a}_{\mathfrak{p}(i)}, \tag{A.1}$$

just because any permutation $\mathfrak{p}(1:n)$ can be obtained from the product by the transposition of neighbouring elements. Since there is an inverse of the permutation $\mathfrak{p}(1:n)$, there exists an inverse $\boldsymbol{K}^{-1}_{\mathfrak{p}(1:n)}(d_{1:n})$ for $\boldsymbol{K}_{\mathfrak{p}(1:n)}(d_{1:n})$ as well. Note that the entries of $d_{1:n}$ are not necessarily equal – they are the dimensions of the vectors $\boldsymbol{a}_i$, $i = 1,2,\ldots,n$. The following example shows that $\boldsymbol{K}_{\mathfrak{p}(1:n)}(d_{1:n})$ is uniquely defined by the permutation $\mathfrak{p}(1:n)$ and the set $d_{1:n}$. The permutation $\mathfrak{p}_{2\to 4}$ is the product of two interchanges $\mathfrak{p}_{2\longleftrightarrow 3}$ and $\mathfrak{p}_{3\longleftrightarrow 4}$, i.e.,

$$\begin{aligned}\boldsymbol{K}_{\mathfrak{p}_{2\to4}}(d_{1:4}) &= \boldsymbol{K}_{\mathfrak{p}_{3\longleftrightarrow4}}(d_1,d_3,d_2,d_4)\, \boldsymbol{K}_{\mathfrak{p}_{2\longleftrightarrow3}}(d_{1:4}) \\ &= \left(\boldsymbol{I}_{d_1} \otimes \boldsymbol{I}_{d_3} \otimes \boldsymbol{K}_{d_4 \cdot d_2}\right)\left(\boldsymbol{I}_{d_1} \otimes \boldsymbol{K}_{d_3 \cdot d_2} \otimes \boldsymbol{I}_{d_4}\right).\end{aligned}$$

This process can be followed for any permutation $\mathfrak{p}(1:n)$ and for any set $d_{1:n}$ of the dimensions.

Likewise, the matrix with only the pair of elements j and k transposed in the product, may be denoted by $\boldsymbol{K}_{\mathfrak{p}_{j\leftrightarrow k}}(d_{1:n})$. We will use the simplified notations $\boldsymbol{K}_{j\leftrightarrow k}$ and $\boldsymbol{K}_{j\to k}$ for the operators $\boldsymbol{K}_{\mathfrak{p}_{j\leftrightarrow k}}$ and $\boldsymbol{K}_{\mathfrak{p}_{j\to k}}$, respectively. It can be seen that

$$\boldsymbol{K}'_{j\leftrightarrow k} = \boldsymbol{K}^{-1}_{j\leftrightarrow k} = \boldsymbol{K}_{k\leftrightarrow j}. \tag{A.2}$$

Let $\boldsymbol{A}$ be $m \times n$ and $\boldsymbol{B}$ be $p \times q$ matrices, it is well known that

$$\boldsymbol{K}_{1\leftrightarrow 2}(m,p)(\boldsymbol{A} \otimes \boldsymbol{B})\boldsymbol{K}_{1\leftrightarrow 2}(q,n) = \boldsymbol{B} \otimes \boldsymbol{A}.$$

The same argument in the case of vector Kronecker product leads to the technique of permuting matrices in a Kronecker product by the help of commutation matrix $\boldsymbol{K}_\mathfrak{p}$.

Using the above notation, we can write

$$\begin{aligned}\operatorname{Vec}(\boldsymbol{A} \otimes \boldsymbol{B}) &= (\boldsymbol{I}_n \otimes \boldsymbol{K}(m,q) \otimes \boldsymbol{I}_p) \operatorname{Vec} \boldsymbol{A} \otimes \operatorname{Vec} \boldsymbol{B} \\ &= \boldsymbol{K}_{2\leftrightarrow 3}(n,m,q,p) \operatorname{Vec} \boldsymbol{A} \otimes \operatorname{Vec} \boldsymbol{B}.\end{aligned} \tag{A.3}$$

A.2. Taylor series in terms of differential operators. We have

$$\psi(\boldsymbol{\lambda}) = \sum_{k_1,k_2,\ldots,k_d=0}^{\infty} \frac{1}{\boldsymbol{k}!} c(\boldsymbol{k}) \boldsymbol{\lambda}_{\boldsymbol{k}} = \sum_{m=0}^{\infty} \frac{1}{m!} \sum_{\substack{k_1,k_2,\ldots,k_d=0 \\ \Sigma k_j = m}}^{m} \frac{m!}{\boldsymbol{k}!} c(\boldsymbol{k}) \boldsymbol{\lambda}_{\boldsymbol{k}}.$$

This can be re-written in the form

$$\psi(\boldsymbol{\lambda}) = \sum_{m=0}^{\infty} \frac{1}{m!} \boldsymbol{c}(m,d)' \boldsymbol{\lambda}^{\otimes m},$$

where $\boldsymbol{c}(m,d)$ is a column vector

$$\boldsymbol{c}(m,d) = \left(D_{\boldsymbol{\lambda}}^{\otimes m} \psi(\boldsymbol{\lambda}) \right)\Big|_{\boldsymbol{\lambda}=\mathbf{0}}$$

with appropriate entries from the vectors $\{c(\boldsymbol{k}), \Sigma k_j = m\}$. The dimension of $\boldsymbol{c}(m,d)$ is the same as that of $\boldsymbol{\lambda}^{\otimes m}$, i.e., d^m. To obtain the above expansion, we proceed as follows. Let $\boldsymbol{x} \in \mathbb{R}^d$ be a real vector and consider

$$(\boldsymbol{x}'\boldsymbol{\lambda})^m = \left(\sum_{j=1}^{d} x_j \lambda_j \right)^m = \sum_{\substack{k_1,k_2,\ldots,k_d=0 \\ \Sigma k_j = m}}^{m} \frac{m!}{\boldsymbol{k}!} \boldsymbol{x}^{\boldsymbol{k}} \boldsymbol{\lambda}^{\boldsymbol{k}},$$

and we can also write

$$(\boldsymbol{x}'\boldsymbol{\lambda})^m = (\boldsymbol{x}'\boldsymbol{\lambda})^{\otimes m} = \left(\boldsymbol{x}^{\otimes m}\right)' \boldsymbol{\lambda}^{\otimes m}.$$

Therefore

$$\sum_{\substack{k_1,k_2,\ldots,k_d=0 \\ \Sigma k_j = m}}^{m} \frac{m!}{\boldsymbol{k}!} \boldsymbol{x}_{\boldsymbol{k}} \boldsymbol{\lambda}_{\boldsymbol{k}} = \left(\boldsymbol{x}^{\otimes m}\right)' \boldsymbol{\lambda}^{\otimes m}.$$

The entries of the vector $\boldsymbol{c}(m,d)$ correspond to the operator $\dfrac{\partial^{\Sigma \boldsymbol{k}}}{\partial \boldsymbol{\lambda}_{\boldsymbol{k}}}$ having the same symmetry as $\boldsymbol{x}_{\boldsymbol{k}}$. Therefore, if $\boldsymbol{x}^{\otimes m}$ is invariant under some permutation of its factors, then $\boldsymbol{c}(m,d)$ is invariant as well. From Equation (2.3) we obtain that

$$\boldsymbol{c}(m,d) = \left(D_{\boldsymbol{\lambda}}^{\otimes m} \psi(\boldsymbol{\lambda}) \right)\big|_{\boldsymbol{\lambda}=\mathbf{0}}.$$

References

ANDREWS, G.E. (1976). *The Theory of Partitions*, Encyclopedia of Mathematics and its Applications **2**, Addison-Wesley, Reading, Massachussetts.

BARNDORFF-NIELSEN, O.E. and COX, D.R. (1989). *Asymptotic Techniques for Use in Statistics*, Monographs on Statistics and Applied Probability, Chapman and Hall, London.

BHATTACHARYA, A. (1946). On some analogues to the amount of information and their uses in statistical estimation. *Sankhyā*, **8**, 1–14.

BRILLINGER, D.R. (2001). *Time Series: Data Analysis and Theory* (reprint of the 1981 edition), Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

BRILLINGER, D.R. and ROSENBLATT, M. (1967). Asymptotic Theory of k-th order spectra. In *Spectral Analysis of Time Series*, B. Harris, ed., New York, Wiley, 153–188.

HANNAN, E.J. (1970). *Multiple Time Series*, Wiley, New York.

LEONOV, V.P. and SHIRYAEV, A.N. (1959). On a method of calculation semi-invariants. *Theor. Probab. Appl.* **4**, 319–329.

LINNIK, YU.V. (1970). A note on Rao-Cramer and Bhattacharya inequalities. *Sankhyā Ser. A*, **32**, 449–452.

LUKACS, E. (1955). Applications of Faa di Bruno's formula in statistics. *Amer. Math. Monthly*, **62**, 340–348.

MARDIA, K.V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, **57**, 519–530.

MCCULLAGH, P. (1987). *Tensor Methods in Statistics*, Monographs on Statistics and Applied Probability, Chapman and Hall, London.

MCCULLAGH, P. and COX, D.R. (1986). Invariants and likelihood ratio statistics. *Ann. Statist.* **14**, 1419–1430.

MAGNUS, J.R. and NEUDECKER, H. (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics* (revised reprint of the 1988 original), Wiley, Chichester.

SKOVGAARD, IB. (1986). A note on the differentiation of cumulants of log likelihood derivatives. *Internat. Statist. Rev.*, **54**, 29–32.

SPEED, T.P. (1990). Invariant moments and cumulants, In *Coding Theory and Design Theory, Part II*, IMA Volumes in Mathematics and its Applications **21**, D. Ray-Chaudhuri, ed., Springer-Verlag, New York, 319–335.

SUBBA RAO, T. and WONG, W.K. (1999). Some Contributions to Multivariate Nonlinear Time Series and to Bilinear Models. In *Asymptotics, Nonparametrics and Time Series*, S. Ghosh, ed., Marcel Dekker, New York, 1–42.

TANIGUCHI, M. (1991). *Higher Order Asymptotic Theory for Time Series Analysis*, Lecture Notes in Statistics **68**, Springer-Verlag, New York.

Terdik, Gy. (1999). *Bilinear Stochastic Models and Related Problems of Nonlinear Time Series Analysis: A Frequency Domain Approach*, Lecture Notes in Statistics **142**, Springer Verlag, New York.

Terdik, Gy. (2002). Higher order statistics and multivariate vector Hermite polynomials for nonlinear analysis of multidimensional time series. *Teor. Ĭmovīr. Mat. Stat.* **66**, 147–168.

Wong, W.K. (1997). Frequency domain tests of multivariate Gaussianity and linearity. *J. Time Ser. Anal.* **18**, 181–194.

S. Rao Jammalamadaka
Department of Statistics
and Applied Probability
University of California
Santa Barbara, CA 93106-3110, USA
E-mail: jammalam@pstat.uscb.edu

T. Subba Rao
School of Mathematics
University of Manchester
PO Box 88,
Manchester M60 1QD, UK
E-mail: tata.subbarao@manchester.ac.uk

György Terdik
Faculty of Informatics
University of Debrecen
4010 Debrecen, Pf.12, Hungary
E-mail: terdik@delfin.unideb.hu

Paper received May 2005; revised February 2006.