Ann. Inst. Statist. Math.
Vol. 50, No. 4, 603–626 (1998)

PARAMETRIC STATISTICAL UNCERTAINTY RELATIONS AND PARAMETRIC STATISTICAL FUNDAMENTAL EQUATIONS

T. MATSUNAWA

The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106-8569, Japan

(Received November 21, 1996; revised October 15, 1997)

Abstract. Multivariate parametric statistical uncertainty relations are proved to specify multivariate basic parametric statistical models. The relations are expressed by inequalities. They generally show that we cannot exactly determine simultaneously both a function of observation objects and a parametric statistical model in a compound parametric statistical system composed of observations and a model. As special cases of the relations, statistical fundamental equations are presented which are obtained as the conditions of attainment of the equality sign in the relations. Making use of the result, a generalized multivariate exponential family is derived as a family of minimum uncertainty distributions. In the final section, several multivariate distributions are derived as basic multivariate parametric statistical models.

Key words and phrases: Parametric statistical uncertainty relations, parametric statistical fundamental equations, specification of basic parametric statistical models, specific intensity of model performance, primary parameter, secondary parameter, minimum statistical uncertainty distributions.

1. Introduction

One of the most important things in statistics is to build a suitable statistical basic model by making effective use of relevant observations based on adequate, real or imaginable experiments. The basic model means an intellectual statistical structure constructed from data with respect to our observational object. It does not mean an approximation to something imagined to be true. Further, it does not mean a population distribution nor a true one, too. These thinkings apparently differ from the usual concepts for statistical models. Nevertheless, this author considers that we should make efforts to construct sound statistical theories of model buildings without assuming true distributions or population distributions, since we cannot consider and verify the objective existence of such distributions. If we succeed in developing such theories, we will be able to get greater recognition among many scientists who challenge the construction of suitable scientific models based on data for new phenomena in various fields.

Incidentally, contemporary statistical fundamental theory which mainly stemmed from R. A. Fisher has not given satisfactory implements for us to handle this important subject. Fisher (1922) stated the problems of specification among the main points of statistical analysis. In his case, this meant a choice of a mathematical form of population distribution with parameters. He regarded this problems as being entirely a matter for the practical statistician. Such a negative understanding towards the problems disappoints us. No idea of resolving the problems was given by him, except for instancing the systems of frequency curves by Karl Pearson (1895, 1916). However, I should remark that Fisher (1936) expected to the possibility that *an even wider type of inductive argument may some day be developed, which shall discuss methods of assigning from the data the functional form of the population*, although he insisted on the existence of population distribution.

Concerning the above subject, Matsunawa (1997) considered specifications of several model distributions for observations by resorting to a modified maximum likelihood method with a specification equation and an estimating equation. In this article, another systematic method to specify parametric statistical model distributions in the multivariate case is presented. Namely, a *parametric statistical uncertainty relation* and *parametric statistical fundamental equations* are presented. With the help of the equations, many standard multivariate distributions can be specified as relevant statistical model distributions. The statistical uncertainty relation in this paper formally resembles the so-called Cramér-Rao variance inequality of an estimator for the unknown parameters of underlying population distribution which is assumed to exist. The inequality for a single parameter was essentially given, among other important things, in the paper by Aitken and Silverstone (1942) as a minimal problem in the calculus of variation. However, the formulation and understanding of our uncertainty relation are essentially different from the background of the existing variance inequalities, although both of them are mathematically nothing but examples of Cauchy-Schwarz inequalities.

As mentioned above, the existence of population distributions or true distributions for observations is not assumed in this paper. Instead, observation devices are considered, as mentioned later. It should be remarked that the main concern of this paper is not to estimate the unknown parameters as in the usual estimation theory, but to specify the functional forms of the parametric statistical fundamental models. Thus, the aspect of the statistical models in this paper is very different from the usual statistical fundamental theory. As a result, the relation is understood to describe the uncertainties of our observational objects and observation devices (= our models). In other words, it is shown that *we cannot determine correctly and simultaneously both of a function of observation objects depending on system parameters and a parametric statistical model within a compound parametric statistical system.* Besides, there is another remarkable structural difference between the Cramér-Rao inequality and our uncertainty inequality (relation) in respect to the number of random elements related to each inequality. The former inequality is *generally* composed of a random sample having *plural* units of random elements, because it is used for the evaluation of the variance of estimator of a unknown parameter. On the other hand, the latter inequality is *in principle* made up of only *one* unit random element as in the uncertainty relation of quantum mechanics, which seems to reflect the sole basic model behind the so-called i.i.d. random observations. These facts also show that the parameter estimation and the model specification are basically different problems from the one Fisher (1922) stated. It should be also noted in advance that the equation obtained when the equality sign holds in the Cramér-Rao inequality shows the relation between an unbiased and sufficient estimator for an unknown parameter of pre-assigned distribution and the marginal distribution of the estimator. On the other hand, the object of the parametric statistical fundamental equation in this paper is the equation to obtain the functional form of an unknown statistical basic model. Namely, we know that the roles of the two equation are completely different.

Needless to say, the uncertainty relation was first presented by Heisenberg (1927) in quantum mechanics between a coordinate and its conjugate momentum in a completely different concept from our statistical one discussed later. His uncertainty relation is chiefly concerned with the microscopic physical world (cf. also, Bohr (1928)). On the contrary, ours describes the relation between measurement observation and its statistical measurement device (= model) in the macroscopic world. Moreover, the notions of physical quantity are completely different and the basic probability concepts are quite different between the two field. Nevertheless, the two uncertainty relations are closely connected in many respects. So, in the future, it is expected that the relations between the two worlds will become clear, if they exists.

In the following section some notations for treating random matrices and necessary assumptions for models are introduced. In Section 3, parametric statistical uncertainty relations in the multivariate case are presented. New definitions are presented for some statistical terminologies such as likelihood and Fisher's information to make clear the concept of statistical uncertainty. A statistical fundamental equation are also given. Making use of the result, a generalized multivariate exponential family is derived as a family of minimum uncertainty distributions. In the final section many standard multivariate distributions are derived.

2. Notations and assumptions

In this paper, we chiefly consider multivariate parametric statistical models. To this end, we introduce some definitions and notations. Further, we need to prepare a few assumptions on the distributions of the models.

Let $\boldsymbol{G} = (\boldsymbol{g}_1, \boldsymbol{g}_2, \ldots, \boldsymbol{g}_n)$ be an $m \times n$ matrix with column vectors $\boldsymbol{g}_i = (g_{1i}, g_{2i}, \ldots, g_{mi})^t$ $(i = 1, 2, \ldots, n)$. Pick up m_i elements from the vector $\boldsymbol{g}_i$ and make a sub-vector $\boldsymbol{g}_i^* = (g_{1i}^*, g_{2i}^*, \ldots, g_{mi}^*)^t$, for each i. Let us now define the *vectorization* of the matrix $\boldsymbol{G}$ by stacking the elements of the sub-vectors:

$$|\boldsymbol{G}\rangle = (g_{11}^*, g_{21}^*, \ldots, g_{m_11}^*, g_{12}^*, g_{22}^*, \ldots, g_{m_22}^*, \ldots, g_{1n}^*, g_{2n}^*, \ldots, g_{m_nn}^*)^t,$$

where $\sum_{i=1}^n m_i = k$ is the dimension of this column vector. We denote its transpose row vector by $\langle\boldsymbol{G}|$ $[\equiv (|\boldsymbol{G}\rangle)^t]$. In multivariate statistical analysis, the following three types of vectorizations of matrices are of importance: (i) Vectorization by picking up the $k = mn$ full elements of an $m \times n$ matrix $\boldsymbol{X}$ and then stacking them. We denote this vectorization as

$$|\,\text{fls}\,\boldsymbol{X}\rangle = (x_{11}, x_{21}, \ldots, x_{m1}, x_{12}, x_{22}, \ldots, x_{m2}, \ldots, x_{1n}, x_{2n}, \ldots, x_{mn})^t \qquad [\equiv (\langle \text{fls}\,\boldsymbol{X}\,|)^t].$$

(ii) Vectorization by picking up the $k = m(m+1)/2$ triangular elements of an $m \times m$ symmetric matrix $\boldsymbol{Y}$ and stacking them as

$$|\,\text{trs}\;\boldsymbol{Y}\rangle = (y_{11}, y_{21}, \ldots, y_{m1}, y_{22}, \ldots, y_{m2}, \ldots, y_{m-1,m-1}, y_{m-1,m}, y_{m,m})^t \qquad [\equiv (\langle \text{trs}\;\boldsymbol{Y}\,|)^t].$$

(iii) Vectorization by picking up the $k = m(m-1)/2$ sub-diagonal elements of an $m \times m$ skew-symmetric matrix $\boldsymbol{Z}$ and then stacking them as follows:

$$|\,\text{sds}\,\boldsymbol{Z}\rangle = (z_{21}, z_{31}, \ldots, z_{m1}, z_{32}, \ldots, z_{m2}, \ldots, z_{m-1,m-2}, z_{m,m-2}, z_{m,m-1})^t \qquad [\equiv \langle \text{sds}\,\boldsymbol{Z}\,|)^t].$$

Corresponding to the above notations, we will use the expressions $\boldsymbol{X} \in \mathbf{G}$, $\boldsymbol{Y} \in \mathbf{S}$ and $\boldsymbol{Z} \in \mathbf{K}$ in Section 4, where $\mathbf{G}$, $\mathbf{S}$ and $\mathbf{K}$ denote the family of general rectangle matrices, the family of symmetric matrices, and the family of skew symmetric matrices, respectively.

Now, let $|\boldsymbol{A}\rangle$ be a k-dimensional random vector by vectorizing an $m \times n$ random matrix $\boldsymbol{A}$. Assume that $|\boldsymbol{A}\rangle$ is defined on the measure space $(R^k, \mathbb{B}^k, \mu)$, where R^k denotes the k-dimensional real space (or, respectively, the set of all k-dimensional nonnegative integer points), $\mathbb{B}^k$ stands for the σ-field of subsets of R^k and μ is the Lebesgue (or, respectively, counting) measure on the measurable space $(R^k, \mathbb{B}^k)$. Here it should be noted that if there is no confusion, we will sometimes use the matrix notation $\boldsymbol{A}$ instead of $|\boldsymbol{A}\rangle$ in the following discussion. We will also use the notation $\boldsymbol{A}$ to express the realization of the random matrix $\boldsymbol{A}$, as usual in multivariate analysis.

Let us consider a family of *parametric model* distributions consisting of the probability measures defined on the measure space. Suppose that the family is parametrized as

$$\mathbb{P} = \{P^{\boldsymbol{A}}_{M,\boldsymbol{\Theta}}; \boldsymbol{\Theta} \in \mathcal{S}_{\boldsymbol{\Theta}}, \boldsymbol{\Lambda} \in \mathcal{S}_{\boldsymbol{\Lambda}}\},$$

where $\boldsymbol{\Theta}$ is a $u \times v$ primarily interested parameter matrix and belongs to a parameter space $\mathcal{S}_{\boldsymbol{\Theta}}$ $(\subset R^k)$. $\boldsymbol{\Lambda}$ is a secondary concerned parameter matrix which is a member of another parameter space $\mathcal{S}_{\boldsymbol{\Lambda}}$. We shall call $\boldsymbol{\Theta}$ a *primary parameter* matrix and $\boldsymbol{\Lambda}$ a *secondary* (or a *hidden*) *parameter* matrix (if it exists). Assume that the family $\mathbb{P}$ is dominated by the measure μ so that

$$P^{\boldsymbol{A}}_{\boldsymbol{\Theta},\boldsymbol{\Lambda}}(d\boldsymbol{A}) = p(\boldsymbol{A}; \boldsymbol{\Theta}, \boldsymbol{\Lambda})\mu(d\boldsymbol{A}) \qquad (\boldsymbol{\Theta} \in \mathcal{S}_{\boldsymbol{\Theta}}, \boldsymbol{\Lambda} \in \mathcal{S}_{\boldsymbol{\Lambda}}),$$

where $p(\boldsymbol{A}; \boldsymbol{\Theta}, \boldsymbol{\Lambda})$ is an imaginary a Radon-Nikodym derivative of $P^{\boldsymbol{A}}_{\boldsymbol{\Theta},\boldsymbol{\Lambda}}$ with respect to the measure μ. The reason why I said imaginary is that we may not find a suitable solution for the parametric statistical fundamental equation proposed later for $p(\boldsymbol{A}; \boldsymbol{\Theta}, \boldsymbol{\Lambda})$. In such a case, our operation ends in failure and eventually we are not able to construct a parametric basic statistical model. Such unfortunate situations may happen when we are not able to give an adequate expression of the observation error $\boldsymbol{\Delta}$, an appropriate measurement scale Ξ and/or a normalized constant c to the candidate function for $p(\boldsymbol{A}; \boldsymbol{\Theta}, \boldsymbol{\Lambda})$. Such being the case, we may compare $p(\boldsymbol{A}; \boldsymbol{\Theta}, \boldsymbol{\Lambda})$ to an imaginary *observation device* when we want to measure observation objects. (The device is understood as a *scaffolding distribution* in Matsunawa (1997), where univariate parametric basic statistical model buildings are discussed by a sort of modified maximum likelihood method.) Well, even if we prepare an observation device in advance, it may happen that we cannot suitably measure the observation objects, in spite of our initial expectation. If, however, we accomplish our task with the help of the device, as we expected, i.e., $p(\boldsymbol{A}; \boldsymbol{\Theta}, \boldsymbol{\Lambda})$ is found as a proper pdf, we say that we succeed in constructing the parametric basic statistical model and we represent it by the same notation as that of the observation device $p(\boldsymbol{A}; \boldsymbol{\Theta}, \boldsymbol{\Lambda})$.

We now make the following assumptions:

(A.1) $\boldsymbol{\Theta}$ and $\boldsymbol{\Lambda}$ are functionally independent, and $\boldsymbol{A}$ is not dependent on these parameters,

(A.2) $\mathcal{S}_{\boldsymbol{\Theta}}$ is either k-dimensional Euclidian space or a rectangle in it,

(A.3) $\partial p(\boldsymbol{A}; \boldsymbol{\Theta}, \boldsymbol{\Lambda})/\partial \boldsymbol{\Theta}$ exists with all finite elements $\mathbb{P}$-a.s. for all $\boldsymbol{\Theta} \in \mathcal{S}_{\boldsymbol{\Theta}}$.

Let $\boldsymbol{\Phi}(\boldsymbol{A}; \boldsymbol{\Theta}, \boldsymbol{\Lambda})$ be a $u \times v$ matrix-valued function of $\boldsymbol{A}$, $\boldsymbol{\Theta}$ and $\boldsymbol{\Lambda}$ (if it exists), which is our *observation object function*. We further assume that the function is partially differentiable with respect to $\boldsymbol{\Theta}$. Corresponding to the function, let us consider a matrix-valued *approximate function* $\boldsymbol{\Psi}(\boldsymbol{A}; \boldsymbol{\Theta}, \boldsymbol{\Lambda})$ which has the same order as that of $\boldsymbol{\Phi}(\boldsymbol{A}; \boldsymbol{\Theta}, \boldsymbol{\Lambda})$. The approximate function is also assumed to be differentiable with respect to $\boldsymbol{\Theta}$. There is a special case such that $\boldsymbol{\Phi}(\boldsymbol{A}; \boldsymbol{\Theta}, \boldsymbol{\Lambda}) = \boldsymbol{A}$ and $\boldsymbol{\Psi}(\boldsymbol{A}; \boldsymbol{\Theta}, \boldsymbol{\Lambda}) = \boldsymbol{\Theta}$, which is an important typical case. Now, put

$$\boldsymbol{\Delta} \doteq \boldsymbol{\Delta}(\boldsymbol{A}; \boldsymbol{\Theta}, \boldsymbol{\Lambda}) = \boldsymbol{\Phi}(\boldsymbol{A}; \boldsymbol{\Theta}, \boldsymbol{\Lambda}) - \boldsymbol{\Psi}(\boldsymbol{A}; \boldsymbol{\Theta}, \boldsymbol{\Lambda}),$$

which means the *measurement error function* when the observation object function $\boldsymbol{\Phi}$ is approximated by the approximate function $\boldsymbol{\Psi}$.

Let us consider the ability of the parametric model (= our observation device) to measure $\boldsymbol{\Delta}$. In order to do that, we regard the so-called likelihood $p(\cdot; \boldsymbol{\Theta}, \cdot)$ as a *performance of data description* by the model. It should be remarked that, according to the above interpretation, the usual log likelihood function

$$\ln p(\boldsymbol{A}; \boldsymbol{\Theta}, \boldsymbol{\Lambda}) \qquad [= \ln(P^{\boldsymbol{A}}_{\boldsymbol{\Theta}, \boldsymbol{\Lambda}}(d\boldsymbol{A})/\mu(d\boldsymbol{A}))]$$

can be considered as the *information specific intensity*, because it yields generalized Kullback-Leibler information per an observation matrix by taking expectation with respect to the model distribution. This aspect naturally leads to the statistical concept of entropy, but we will not enter the topic in this article (cf. Matsunawa (1995)).

We further need to grasp the variation of the performance of data description $p(\cdot; \boldsymbol{\Theta}, \cdot)$ when the primary parameter matrix $\boldsymbol{\Theta}$ has small changes. In order to this we consider a *performance specific intensity* per an observation matrix with respect to $\boldsymbol{\Theta}$:

$$\mathcal{P} = \mathcal{P}(\boldsymbol{A}; \boldsymbol{\Theta}, \boldsymbol{\Lambda}) = \partial \ln p(\boldsymbol{A}; \boldsymbol{\Theta}, \boldsymbol{\Lambda})/\partial \boldsymbol{\Theta} \qquad [= (\partial \ln p/\partial \Theta_{ij})].$$

Of course, this is well known as the score function introduced by R. A. Fisher, but our newly introduced terminologies are useful to relate statistical fundamental theories to other closely related sciences such as physics and thermodynamics (cf. Kapur (1989)).

Throughout the paper, we assume that there exist some nonsingular matrices defined by the following expectations by the Radon-Nikodym derivative of $p(\boldsymbol{A}; \boldsymbol{\Theta}, \boldsymbol{\Lambda})$ with respect to measure μ:

$\boldsymbol{\Sigma} := E[|\boldsymbol{\Delta}\rangle\langle\boldsymbol{\Delta}|]$: (the mean fluctuation of the observation object =: the *uncertainty* on the measurement of the object),

$\boldsymbol{I} := E[|\mathcal{P}\rangle\langle\mathcal{P}|]$: (the mean fluctuation of specific intensity of the model-performance = a standard scale unit for measuring the error by the model =: the *uncertainty* on the model-performance),

$\boldsymbol{J}^t := E[|\boldsymbol{\Delta}\rangle\langle\mathcal{P}|]$: (the mean fluctuation by interaction from the object to the model),

$\boldsymbol{J} := E[|\mathcal{P}\rangle\langle\boldsymbol{\Delta}|]$: (the mean fluctuation by interaction from the model to the object),

$\boldsymbol{K} := \boldsymbol{I}(\boldsymbol{J}^t)^{-1}$: (an adjusted scale unit for measuring the error by the model).

The meanings explained in the parentheses are there to help in our later statistical discussion. For example, our understanding of the quantity $\boldsymbol{I}$ becomes more suitable than the usual one of Fisher's information matrix.

3. Statistical uncertainty relations

Under the set-up described in the preceding section, we have the following statistical uncertainty relation:

THEOREM 3.1. *For any* $\boldsymbol{\Theta} \in \mathcal{S}_{\boldsymbol{\Theta}}$, $\boldsymbol{\Lambda} \in \mathcal{S}_{\boldsymbol{\Lambda}}$ *and* $\boldsymbol{y} \in R^k$ *the following quadratic form inequalities hold*:

$$\text{(3.1)} \qquad \langle \boldsymbol{y}|\boldsymbol{\Sigma}|\boldsymbol{y}\rangle \geq \langle \boldsymbol{J}\boldsymbol{y}|\boldsymbol{I}^{-1}|\boldsymbol{J}\boldsymbol{y}\rangle, \qquad [\textit{resp.}\ \langle \boldsymbol{J}^{-1}\boldsymbol{y}|\boldsymbol{I}^{-1}|\boldsymbol{J}^{-1}\boldsymbol{y}\rangle \geq \langle \boldsymbol{y}|\boldsymbol{\Sigma}^{-1}|\boldsymbol{y}\rangle].$$

We symbolically denote these as

$$\text{(3.2)} \qquad \boldsymbol{\Sigma} - \boldsymbol{J}^t\boldsymbol{I}^{-1}\boldsymbol{J} \gg \boldsymbol{0}, \qquad [\textit{resp.}\ \boldsymbol{J}^{-1}\boldsymbol{I}(\boldsymbol{J}^{-1})^t - \boldsymbol{\Sigma}^{-1} \gg \boldsymbol{0}].$$

The equality signs hold if and only if there exists some measurement precision matrix $\boldsymbol{K} = \boldsymbol{K}(\boldsymbol{\Theta}, \boldsymbol{\Lambda}) = \boldsymbol{I}(\boldsymbol{J}^t)^{-1} \neq \boldsymbol{0}$ [*resp.* $\boldsymbol{L} = \boldsymbol{L}(\Theta, \Lambda) = \boldsymbol{J}\boldsymbol{\Sigma}^{-1} \neq \boldsymbol{0}$] *such that the following parametric statistical fundamental equation holds*:

$$\text{(3.3)} \qquad \left|\frac{\partial \ln p(\boldsymbol{A}; \boldsymbol{\Theta}, \boldsymbol{\Lambda})}{\partial \boldsymbol{\Theta}}\right\rangle = \boldsymbol{K}(\boldsymbol{\Theta}, \boldsymbol{\Lambda})|\boldsymbol{\Delta}(\boldsymbol{A}; \boldsymbol{K}(\boldsymbol{\Theta}, \boldsymbol{\Lambda}))\rangle \qquad (\mu\text{-}a.e.).$$

PROOF. Let us consider a k-dimensional random vectors

$$d \equiv |\Delta\rangle - J^t I^{-1} |\mathcal{P}\rangle, \qquad [\text{resp. } \delta = |\mathcal{P}\rangle - J\Sigma^{-1}|\Delta\rangle]. \tag{3.4}$$

Since $|d\rangle\langle d| \gg 0_{k\times k}$ (μ-a.e.), for any nonzero vector $y \in R^k$, we get $\langle y|E[|d\rangle\langle d|]|y\rangle \geq 0$, where the equality sign holds if and only if $d = 0_{k\times 1}$. Therefore, we may only evaluate $E[|d\rangle\langle d|] \gg 0_{k\times k}$:

$$\begin{aligned} E[|d\rangle\langle d|] &= E[\{|\Delta\rangle - J^t I^{-1}|\mathcal{P}\rangle\}\{\langle\Delta| - \langle\mathcal{P}|(I^{-1})^t J\}] \\ &= E[|\Delta\rangle\langle\Delta|] - E[|\Delta\rangle\langle\mathcal{P}|](I^{-1})^t J - J^t I^{-1} E[|\mathcal{P}\rangle\langle\Delta|] \\ &\quad + J^t I^{-1} E[|\mathcal{P}\rangle\langle\mathcal{P}|](I^{-1})^t J \\ &= \Sigma - J^t (I^{-1})^t J - J^t I^{-1} J + J^t I^{-1} I (I^{-1})^t J = \Sigma - J^t I^{-1} J \gg 0_{k\times k}. \end{aligned}$$

It is clear that $|d\rangle\langle d| = 0_{k\times k}$ if and only if $d = |\Delta\rangle - J^t I^{-1}|\mathcal{P}\rangle = 0_{k\times 1}$ (μ-a.e.). Thus, $|\mathcal{P}\rangle = I(J^t)^{-1}|\Delta\rangle = K|\Delta\rangle$ [resp. $|\mathcal{P}\rangle = J\Sigma^{-1}|\Delta\rangle$], ($\mu$-a.e.). Namely, we get the desired Equation (3.2). □

Related to the above, we have the following corollary:

COROLLARY 3.1. *Under the same conditions as those in the above theorem, we have an uncertainty relation in terms of a formal matrix quadratic form: for any $k \times k$ real symmetric matrix Y, it holds that*

$$Y^t(\Sigma - J^t I^{-1} J)Y \gg 0_{k\times k}.$$

PROOF. Since, from Theorem 3.1, $y^t\Sigma y \geq y^t J^t I^{-1} J y$ for any $y \in R^k$, pre- and post-multiplying a scaler value $y^t y$ to the both sides of the inequality and putting $y^t y = Y = Y^t$ we get $y^t Y^t \Sigma Y y \geq y^t Y^t J^t I^{-1} J Y y$, from which we have the matrix quadratic form expression of an uncertainty relation. □

Now, by integrating the parametric statistical fundamental Equation (3.3), we have the following *minimum uncertainty distribution*:

THEOREM 3.2. *Under the same conditions as those of the preceding theorem, we get*

$$\begin{aligned} p(A;\Theta,\Lambda &= c(A;\Lambda)\exp\left\{\int\langle d\Theta|K(\Theta,\Lambda)|\Delta(A;\Theta,\Lambda)\rangle\right\} \\ &= c(A;\Lambda)\exp\left\{\int\langle\Delta(A;\Theta,\Lambda)|K^t(\Theta,\Lambda)|d\Theta\rangle\right\}, \end{aligned}$$

where $\langle d\Theta|$ [resp. $|d\Theta\rangle$] stands for an l-dimensional differential row [resp. column] vector based on $\Theta = (\theta_{ij})_{u\times v}$ given by

$$\langle d\Theta| = (|d\Theta\rangle)^t = (d\theta^*_{11},\ldots,d\theta^*_{u_1 1}, d\theta^*_{12},\ldots,d\theta^*_{u_2 2},\ldots,d\theta^*_{1v},\ldots,d\theta^*_{u_n v}),$$

where $\sum_{i=1}^{v} u_i = l$. $c(\boldsymbol{A};\boldsymbol{\Lambda}) > 0$ *denotes a scaler-valued function of* $\boldsymbol{A}$ *and* $\boldsymbol{\Lambda}$, *if it exists, satisfying the condition*

$$\int_{R^k} p(|\boldsymbol{A}\rangle;\boldsymbol{\Theta},\boldsymbol{\Lambda})\mu\{(d|\boldsymbol{A}\rangle)\} = 1,$$

where $\mu\{\cdot\}$ *designates the* σ*-finite measure on the measurable space* $(R^k, \mathbb{B}^k)$, *and* $(d|\boldsymbol{A}\rangle)$ *stands for a volume element of the k-fold integral given by*

$$(d|\boldsymbol{A}\rangle) = da_{11}^* \cdots da_{m_1 1}^* da_{12}^*, da_{m_2 2}^* \cdots, da_{1n}^* \cdots da_{m_n n}^*$$

where $\boldsymbol{A} = (a_{ij})_{m\times n}$ *and* $\sum_{i=1}^{n} m_i = k$.

Remark 3.1. It should be noted that the integrand in the exponential part means a total differential of some scaler-valued function. In next section, we devise ways of integratings from some aspects of matrix calculation.

4. Specifications of multivariate distributions

In this section we construct basic multivariate distributions by making use of the previous theorem. First, we give some matrix-valued parametric statistical fundamental equations corresponding to the vector-valued Equation (3.3). To do this, let us consider the following matrix differential operators with respect to underlying parameter matrices:

(i) For a general real rectangular matrix $\boldsymbol{X} = (x_{ij})_{m\times n} \in \mathbf{G}$,

$$\frac{\partial}{\partial \boldsymbol{X}} = \left(\frac{\partial}{\partial x_{ij}}\right)_{m\times n} =: \begin{pmatrix} \frac{\partial}{\partial x_{11}} & \frac{\partial}{\partial x_{12}} & \cdots & \frac{\partial}{\partial x_{1n}} \\ \frac{\partial}{\partial x_{21}} & \frac{\partial}{\partial x_{22}} & \cdots & \frac{\partial}{\partial x_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_{m1}} & \frac{\partial}{\partial x_{m2}} & \cdots & \frac{\partial}{\partial x_{mn}} \end{pmatrix}. \tag{4.1}$$

(ii) For a real symmetric matrix $\boldsymbol{Y} = (y_{ij})_{m\times m} \in \mathbf{S}$, $(y_{ij} = y_{ji})$,

$$\frac{\partial^S}{\partial \boldsymbol{Y}} = \left(\frac{\partial^S}{\partial y_{ij}}\right)_{m\times m} =: \begin{pmatrix} \frac{\partial}{\partial y_{11}} & \frac{1}{2}\frac{\partial}{\partial y_{12}} & \cdots & \frac{1}{2}\frac{\partial}{\partial y_{1m}} \\ \frac{1}{2}\frac{\partial}{\partial y_{12}} & \frac{\partial}{\partial y_{22}} & \cdots & \frac{1}{2}\frac{\partial}{\partial y_{2m}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{2}\frac{\partial}{\partial y_{1m}} & \frac{1}{2}\frac{\partial}{\partial y_{2m}} & \cdots & \frac{\partial}{\partial y_{mm}} \end{pmatrix}. \tag{4.2}$$

Finally, (iii) for a real skew-symmetric matrix $\boldsymbol{Z} = (z_{ij})_{m\times m} \in \mathbf{K}$, $(z_{ij} = -z_{ji})$,

$$\frac{\partial^K}{\partial \boldsymbol{Z}} = \left(\frac{\partial^K}{\partial z_{ij}}\right)_{m\times m} =: \begin{pmatrix} 0 & \frac{1}{2}\frac{\partial}{\partial z_{12}} & \cdots & \frac{1}{2}\frac{\partial}{\partial z_{1m}} \\ -\frac{1}{2}\frac{\partial}{\partial z_{12}} & 0 & \cdots & \frac{1}{2}\frac{\partial}{\partial z_{2m}} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{2}\frac{\partial}{\partial z_{1m}} & -\frac{1}{2}\frac{\partial}{\partial z_{2m}} & \cdots & 0 \end{pmatrix}. \tag{4.3}$$

Next, we need to introduce the following matrix differentials corresponding to the above cases. For $\boldsymbol{X} = (x_{ij})_{m\times n} \in \mathbf{G}$,

$$dX := \begin{pmatrix} dx_{11} & dx_{12} & \cdots & dx_{1n} \\ dx_{21} & dx_{22} & \cdots & dx_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ dx_{m1} & dx_{m2} & \cdots & dx_{mn} \end{pmatrix} = (dx_{ij})_{m\times n} \tag{4.4}$$

and $(d\boldsymbol{X}) := \bigwedge_{i=1}^{m}\bigwedge_{j=1}^{n} dx_{ij}$ is the exterior product of the mn elements of $d\boldsymbol{X}$.

For $\boldsymbol{Y} = (y_{ij})_{m\times m} \in \mathbf{S}$, $(y_{ij} = y_{ji})$,

$$d\boldsymbol{Y} := \begin{pmatrix} dy_{11} & dy_{12} & \cdots & dy_{1m} \\ dy_{21} & dy_{22} & \cdots & dy_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ dy_{1m} & dy_{2m} & \cdots & dy_{mm} \end{pmatrix} \tag{4.5}$$

and $(d\boldsymbol{Y}) := \bigwedge_{1\le i\le j\le m} dy_{ij}$ is the exterior product of the $\frac{1}{2}m(m+1)$ elements of $d\boldsymbol{Y}$.

For $\boldsymbol{Z} = (z_{ij})_{m\times m} = (-z_{ji})_{m\times m} \in \mathbf{K}$,

$$d\boldsymbol{Z} := \begin{pmatrix} 0 & dz_{12} & \cdots dz_{1m} & \\ -dz_{12} & 0 & \cdots dz_{2m} & \\ \vdots & \vdots & \ddots & \vdots \\ -dz_{1m} & -dz_{2m} & \cdots & 0 \end{pmatrix} \tag{4.6}$$

and $(d\boldsymbol{Z}) := \bigwedge_{1\le i<j\le m} dz_{ij}$ is the exterior product of the $\frac{1}{2}m(m-1)$ elements of $d\boldsymbol{Z}$.

Under the above set-up, let us consider matrix representations of the parametric statistical fundamental equations. From Theorem 3.1, we expect to have a corresponding matrix-valued equation of the form

$$\frac{\partial \ln p(\boldsymbol{A};\boldsymbol{\Theta},\boldsymbol{\Lambda})}{\partial \boldsymbol{\Theta}} = \boldsymbol{\Xi}^{\dagger}(\boldsymbol{\Theta},\boldsymbol{\Lambda})\boldsymbol{\Delta}^{\dagger}(\boldsymbol{A};\boldsymbol{\Theta},\boldsymbol{\Lambda}),$$

where $\boldsymbol{\Delta}^{\dagger}(\boldsymbol{A};\boldsymbol{\Theta},\boldsymbol{\Lambda})$ and $\boldsymbol{\Xi}^{\dagger}(\boldsymbol{\Theta},\boldsymbol{\Lambda})$ are a modified measurement error function and a modified precision matrix, respectively, corresponding to $\boldsymbol{\Delta}(\boldsymbol{A};\boldsymbol{\Theta},\boldsymbol{\Lambda})$ and $\boldsymbol{K}(\boldsymbol{\Theta},\boldsymbol{\Lambda})$ appearing in (3.3). We have the following result.

THEOREM 4.1. *The following matrix-valued statistical fundamental equations hold*:

(i) *For a real primary parameter matrix* $\boldsymbol{\Theta}_1(m\times n)$ *and a secondary parameter matrix* $\boldsymbol{\Lambda}_1$, *we have a parametric statistical fundamental equation*

$$\frac{\partial \ln p(\boldsymbol{A}_1;\boldsymbol{\Theta}_1,\boldsymbol{\Lambda}_1)}{\partial \boldsymbol{\Theta}_1} = \boldsymbol{\Xi}_1(\boldsymbol{\Theta}_1,\boldsymbol{\Lambda}_1)(|\boldsymbol{\Delta}_1(\boldsymbol{A}_1;\boldsymbol{\Theta}_1,\boldsymbol{\Lambda}_1)\rangle \otimes \boldsymbol{I}_n) \qquad (\mu\text{-}a.e.), \tag{4.7}$$

where $\otimes \boldsymbol{I}_n$ *is the Kronecker product by the unit matrix of order* n; $|\boldsymbol{\Delta}_1\rangle(mn\times 1)$ *is a measurement error vector. The matrix* $\boldsymbol{\Xi}_1 = \boldsymbol{\Xi}_1(\boldsymbol{K}_1(\boldsymbol{\Theta}_1,\boldsymbol{\Lambda}_1)) = ({}_1\xi_{ij})(m\times mn^2)$

denotes the precision matrix depending on $\boldsymbol{K}_1(\boldsymbol{\Theta}_1, \boldsymbol{\Lambda}_1) = ({}_1\kappa_{\alpha\beta})(mn \times mn)$ *in* (3.3) *and satisfies the relations*

$$(4.8)\qquad {}_1\xi_{i,j+n(l-1)} = {}_1\kappa_{i+m(j-1),l} \quad (1 \le i \le m,\ 1 \le j \le n;\ l = 1, \dots, mn).$$

(ii) *For a symmetric real parameter matrix* $\boldsymbol{\Theta}_2(m \times m)$ *and a secondary parameter matrix* $\boldsymbol{\Lambda}_2$, *we have a parametric statistical fundamental equation*

$$(4.9)\qquad \frac{\partial^S \ln p(\boldsymbol{A}_2; \boldsymbol{\Theta}_2, \boldsymbol{\Lambda}_2)}{\partial \boldsymbol{\Theta}_2} = \boldsymbol{\Xi}(\boldsymbol{K}_2(\boldsymbol{\Theta}_2, \boldsymbol{\Lambda}_2))(|\operatorname{trs} \boldsymbol{\Delta}_2(\boldsymbol{A}_2; \boldsymbol{\Theta}_2, \boldsymbol{\Lambda}_2)\rangle \otimes \boldsymbol{I}_m) \qquad (\mu\text{-a.e.}),$$

where $\boldsymbol{\Delta}_2(m \times m)$ *is a symmetric measurement error matrix,* $\boldsymbol{\Xi}_2 = \boldsymbol{\Xi}_2(\boldsymbol{K}_2(\boldsymbol{\Theta})2, \boldsymbol{\Lambda}_2)) = ({}_2\xi_{ij})(m \times \frac{1}{2}m^2(m+1))$ *is the precision matrix depending on* $\boldsymbol{K}_2(\boldsymbol{\Theta}_2, \boldsymbol{\Lambda}_2) = ({}_2\kappa_{\alpha\beta})(\frac{1}{2}m(m+1) \times \frac{1}{2}m(m+1))$ *in* (3.3) *and satisfies the relations*

$$(4.10)\quad {}_2\xi_{i,j+m(l-1)} = {}_2\xi_{j,i+m(l-1)} \quad \text{and} \quad {}_2\xi_{i,j+m(l-1)} = {}_2\kappa_{i+m(j-1)-j(j-1)/2,l},$$

for $1 \le i \le m$, $1 \le j \le m$; $l = 1, \dots, \frac{1}{2}m(m+1)$.

(iii) *For a skew-symmetric real parameter matrix* $\boldsymbol{\Theta}_3(m \times m)$ *and a secondary parameter* $\boldsymbol{\Lambda}_3$, *we have a parametric statistical fundamental equations*

$$(4.11)\qquad \frac{\partial^K \ln p(\boldsymbol{A}_3; \boldsymbol{\Theta}_3, \boldsymbol{\Lambda}_3)}{\partial \boldsymbol{\Theta}_3} = \boldsymbol{\Xi}_3(\boldsymbol{\Theta}_3, \boldsymbol{\Lambda}_3)(|\operatorname{sds} \boldsymbol{\Delta}_3(\boldsymbol{A}_3; \boldsymbol{\Theta}_3, \boldsymbol{\Lambda}_3)\rangle \otimes \boldsymbol{I}_m) \qquad (\mu\text{-a.e.}),$$

where $\boldsymbol{\Delta}_3(m \times m)$ *is a skew-symmetric measurement error and* $\boldsymbol{\Xi}_3 = \boldsymbol{\Xi}_3(\boldsymbol{K}_3(\boldsymbol{\Theta}_3, \boldsymbol{\Lambda}_3)) = ({}_3\xi_{ij})(m \times \frac{1}{2}m^2(m-1))$ *is the precision matrix depending on* $\boldsymbol{K}_3(\boldsymbol{\Theta}_3, \boldsymbol{\Lambda}_3) = ({}_3\kappa_{\alpha\beta})(\frac{1}{2}m(m-1) \times \frac{1}{2}m(m-1))$ *in* (3.3) *and satisfies the relations for* $1 \le i \le m$, $1 \le j \le m$; $l = 1, \dots, \frac{1}{2}m(m-1)$

$${}_3\xi_{i,i+m(l-1)} = 0, \qquad {}_3\xi_{i,j+m(l-1)} = -{}_3\xi_{j,i+m(l-1)} \qquad (i \ne j),$$

and

$$(4.12)\qquad {}_3\xi_{i,j+m(l-1)} = {}_3\kappa_{i+m(j-1)-j(j+1)/2,l},$$

for $1 > j$; $2 \le i \le m$, $1 \le j \le m-1$; $l = 1, \dots, \frac{1}{2}m(m-1)$.

PROOF. (i) From (3.3), for a real rectangular parameter matrix $\boldsymbol{\Theta}_1(m \times n)$, we have a vector-represented parametric statistical fundamental equation

$$(4.13)\qquad \left|\operatorname{fls} \frac{\partial \ln p(\boldsymbol{A}_1; \boldsymbol{\Theta}_1, \boldsymbol{\Lambda}_1)}{\partial \boldsymbol{\Theta}_1}\right\rangle = \boldsymbol{K}_1(\boldsymbol{\Theta}_1, \boldsymbol{\Lambda}_1)|\operatorname{fls} \boldsymbol{\Delta}_1(\boldsymbol{A}_1; \boldsymbol{\Theta}_1, \boldsymbol{\Lambda})\rangle \qquad (\mu\text{-a.e.}),$$

where $\boldsymbol{K}_1(\boldsymbol{\Theta}_1, \boldsymbol{\Lambda}_1) = ({}_1\kappa_{\alpha\beta})$ is an $mn \times mn$ nonsingular matrix. Further, denote the $mn \times 1$ measurement error vector as

$$|\boldsymbol{\Delta}_1\rangle = (\Delta_{11}, \Delta_{21}, \dots, \Delta_{m1}, \Delta_{12}, \dots, \Delta_{m2}, \Delta_{13}, \dots, \Delta_{1n}, \dots, \Delta_{mn})^t$$

and renumber the suffices of the components as

$$|\boldsymbol{\Delta}_1\rangle =: (\Delta_1^*, \Delta_2^*, \dots, \Delta_m^*, \Delta_{m+1}^*, \dots, \Delta_{2m}^*, \Delta_{2m+1}^*, \dots, \Delta_{(n-1)m}^*, \dots, \Delta_{mn}^*)^t. \tag{4.14}$$

Then, we can represent the RHS vector of Equation (4.13) as

$$\begin{aligned} &\boldsymbol{K}_1(\boldsymbol{\Theta}_1, \boldsymbol{\Lambda}_1)|\operatorname{fls} \boldsymbol{\Delta}_1(\boldsymbol{A}_1; \boldsymbol{\Theta}_1, \boldsymbol{\Lambda})\rangle \\ &= \left(\sum_{l=1}^{mn} {}_1\kappa_{1l}\Delta_l^*, \sum_{l=1}^{mn} {}_1\kappa_{2l}\Delta_l^*, \dots, \sum_{l=1}^{mn} {}_1\kappa_{mn,l}\Delta_l^*\right)^t. \end{aligned} \tag{4.15}$$

Let us construct matrix expressions of the above vector-represented equation. To this end, we introduce an operator which may be called the *matrixization operator*:

$$\mathbb{M}(|\bullet\rangle) := (\boldsymbol{I}_m \otimes \langle \boldsymbol{I}_n|)(\mathbb{K}_{mn} \otimes \boldsymbol{I}_n)(|\bullet\rangle \otimes \boldsymbol{I}_n), \tag{4.16}$$

where $|\bullet\rangle$ denotes the full stacking vector of a matrix and where $\mathbb{K}_{mn}(mn \times mn)$ stands for a *commutation matrix* such that $\mathbb{K}_{mn}|\operatorname{fls} \boldsymbol{X}\rangle = |\operatorname{fls} \boldsymbol{X}^t\rangle$ for an $m \times n$ matrix $\boldsymbol{X}$. Applying this operator to the above equation, we have

$$\mathbb{M}\left(\left|\operatorname{fls} \frac{\partial \ln p(\boldsymbol{A}_1; \boldsymbol{\Theta}_1, \boldsymbol{\Lambda}_1)}{\partial \boldsymbol{\Theta}_1}\right\rangle\right) = \frac{\partial \ln p(\boldsymbol{A}_1; \boldsymbol{\Theta}_1, \boldsymbol{\Lambda}_1)}{\partial \boldsymbol{\Theta}_1} \tag{4.17}$$

and

$$\begin{aligned} &\mathbb{M}(\boldsymbol{K}_1(\boldsymbol{\Theta}_1, \boldsymbol{\Lambda}_1)|\operatorname{fls} \boldsymbol{\Delta}_1(\boldsymbol{A}_1; \boldsymbol{\Theta}_1, \boldsymbol{\Lambda}_1)\rangle) \\ &= \boldsymbol{\Xi}_1(\boldsymbol{K}_1(\boldsymbol{\Theta}_1, \boldsymbol{\Lambda}_1))(|\operatorname{fls} \boldsymbol{\Delta}_1(\boldsymbol{A}_1; (\boldsymbol{\Theta}_1, \boldsymbol{\Lambda}_1)\rangle \otimes \boldsymbol{I}_n) \qquad \text{(say)}, \end{aligned}$$

where $\boldsymbol{\Xi}_1(\boldsymbol{K}_1(\boldsymbol{\Theta}_1, \boldsymbol{\Lambda}_1)) = ({}_1\xi_{ij})(m \times mn^2)$ with $\boldsymbol{K}_1(\boldsymbol{\Theta}_1, \boldsymbol{\Lambda}_1) = ({}_1\kappa_{\alpha\beta})(mn \times mn)$. Thus, the above matrix has a representation as

$$\begin{aligned} &\boldsymbol{\Xi}_1(\boldsymbol{K}_1(\boldsymbol{\Theta}_1, \boldsymbol{\Lambda}_1))(|\operatorname{fls} \boldsymbol{\Delta}_1(\boldsymbol{A}_1; \boldsymbol{\Theta}_1, \boldsymbol{\Lambda}_1)\rangle \otimes \boldsymbol{I}_n) \\ &= (\xi_{ij}\Delta_{11} + \xi_{i,n+j}\Delta_{21} + \xi_{j,2n+j}\Delta_{31} + \dots + \xi_{i,(mn-1)n+j}\Delta_{mn})_{ij} \\ &=: \left(\sum_{l=1}^{mn} \xi_{i,(l-1)n+j}\Delta_l^*\right)_{ij}. \end{aligned} \tag{4.18}$$

It should be noted that the second suffices of ξ's compose an arithmetic progression with a common difference n and an initial term j which coincide with the column number of the above matrix. Vectorizing the matrix by fully stacking the columns, we find that the $\{(j-1)m + i\}$-th column of the resultant vector is

$$\sum_{l=1}^{mn} \xi_{i,(l-1)n+j}\Delta_l^* \qquad (1 \le i \le m,\ 1 \le j \le n).$$

Namely, considering (4.15), this must coincides with

$$\sum_{l=1}^{mn} {}_1\kappa_{i+m(j-1),l}\Delta_l^* \qquad (1 \le i \le m,\ 1 \le j \le n),$$

and, thus, comparing the coefficients of Δ_l^*, we have the following relation between the components of the matrix Ξ_1 and $\boldsymbol{K}_1$:

$${}_1\xi_{i,j+n(l-1)} = {}_1\kappa_{i+m(j-1),l} \qquad (1 \le i \le m,\ 1 \le j \le n;\ l = 1, \ldots, mn).$$

Therefore, there exists a precision matrix $\Xi_1(\boldsymbol{\Theta}_1, \boldsymbol{\Lambda}_1)$ satisfying the desired result (4.7), which completes the proof of (i).

(ii) For a symmetric real parameter matrix $\boldsymbol{\Theta}_2(m \times m)$, we have from (3.3)

$$\text{(4.19)} \quad \left|\operatorname{trs} \frac{\partial^S \ln p(\boldsymbol{A}_2; \boldsymbol{\Theta}_2, \boldsymbol{\Lambda}_2)}{\partial \boldsymbol{\Theta}_2}\right\rangle = \boldsymbol{K}_2(\boldsymbol{\Theta}_2, \boldsymbol{\Lambda}_2)|\operatorname{trs} \boldsymbol{\Delta}_2(\boldsymbol{A}_2; \boldsymbol{\Theta}_2, \boldsymbol{\Lambda}_2)\rangle \qquad (\mu\text{-a.e.}),$$

where $\boldsymbol{K}_2(\boldsymbol{\Theta}_2, \boldsymbol{\Lambda}_2) = ({}_2\kappa_{\alpha\beta})$ is some $\frac{1}{2}m(m+1) \times \frac{1}{2}m(m+1)$ nonsingular matrix. Let us consider the $\frac{1}{2}m(m+1) \times 1$ vector $|\boldsymbol{\Delta}_2\rangle$, corresponding to (4.14), given by

$$\begin{aligned} |\boldsymbol{\Delta}_2\rangle =: (&\Delta_1^{\#}, \ldots, \Delta_m^{\#}, \Delta_{m+1}^{\#}, \ldots, \Delta_{2m-1}^{\#}, \Delta_{2m}^{\#}, \ldots, \\ &\Delta_{3m-3}^{\#}, \Delta_{3m-2}^{\#}, \ldots, \Delta_{4m-6}^{\#}, \ldots, \Delta_{m(m+1)/2}^{\#})^t. \end{aligned}$$

Then, we can represent the RHS vector of equation (4.19) as

$$\text{(4.20)} \qquad \begin{aligned} \boldsymbol{K}_2|\operatorname{trs} \boldsymbol{\Delta}_2\rangle = \Bigg(&\sum_{l=1}^{m(m+1)/2} {}_2\kappa_{1l}\Delta_l^{\#}, \sum_{l=1}^{m(m+1)/2} {}_2\kappa_{2l}\Delta_l^{\#}, \ldots, \\ &\sum_{l=1}^{m(m+1)/2} {}_2\kappa_{m(m+1)/2,l}\Delta_l^{\#} \Bigg)^t. \end{aligned}$$

On the other hand, making use of matrixization operation (4.16) with $n = m$, we have

$$\text{(4.21)} \qquad \mathbb{M}\left(\mathbb{F}_m^S \left|\operatorname{trs} \frac{\partial^S \ln p(\boldsymbol{A}_3; \boldsymbol{\Theta}_3, \boldsymbol{\Lambda}_3)}{\partial \boldsymbol{\Theta}_2}\right\rangle\right) = \frac{\partial^S \ln p(\boldsymbol{A}_3; \boldsymbol{\Theta}_3, \boldsymbol{\Lambda}_3)}{\partial \boldsymbol{\Theta}_2},$$

and carrying out the same operation to the RHS of (4.19) we have

$$\text{(4.22)} \qquad \begin{aligned} &\mathbb{M}(\mathbb{F}_m^S(\boldsymbol{K}_2(\boldsymbol{\Theta}_2, \boldsymbol{\Lambda}_2)|\operatorname{trs} \boldsymbol{\Delta}_2(\boldsymbol{A}_2; \boldsymbol{\Theta}_2, \boldsymbol{\Lambda}_2)\rangle)) \\ &\quad = \Xi_2(\boldsymbol{\Theta}_2, \boldsymbol{\Lambda}_2)(|\operatorname{trs} \boldsymbol{\Delta}_2(\boldsymbol{A}_2; \boldsymbol{\Theta}_2, \boldsymbol{\Lambda}_2)\rangle \otimes \boldsymbol{I}_m) \qquad \text{(say)}, \end{aligned}$$

where $\mathbb{F}_m^S(m^2 \times \frac{1}{2}m(m+1))$ is the *full stacking operator* such that $\mathbb{F}_m^S|\operatorname{trs} \boldsymbol{Y}\rangle = |\operatorname{fls} \boldsymbol{Y}\rangle$ for an $m \times m$ symmetric matrix $\boldsymbol{Y}$, and $\Xi_2 = \Xi_2(\boldsymbol{K}_2(\boldsymbol{\Theta}_2, \boldsymbol{\Lambda}_2)) = ({}_2\xi_{ij})(m \times$

$\frac{1}{2}m^2(m+1)$). Vectorizing the matrix (4.22) by triangularly stacking the columns, we find that the $\{i+m(j-1)-\frac{1}{2}j(j-1)\}$-th row of the vector is

$$\sum_{l=1}^{m(m+1)/n} {}_2\xi_{i,j+(l-1)m}\Delta_l^{\#}, \qquad (1 \le i \le m, 1 \le j \le m).$$

Namely, considering (4.20), this must coincide with

$$\sum_{l=1}^{m(m+1)/n} {}_2\kappa_{i+m(j-1)-j(j-1)/2,l}\Delta_l^{\#}, \qquad (1 \le i \le m, 1 \le j \le m)$$

and, hence, comparing the resultant coefficients of $\Delta_l^{\#}$, we have the relation (4.10) between the components of the matrix Ξ_2 and $\boldsymbol{K}_2$. Thus, there exists a precision matrix $\Xi_2(\boldsymbol{\Theta}_2, \boldsymbol{\Lambda}_2)$ satisfying the desired result (4.9), which completes the proof of (ii).

(iii) For a skew-symmetric real parameter matrix $\boldsymbol{\Theta}_3(m \times m)$, we have from (3.3)

$$\left| \mathrm{sds}\, \frac{\partial^K \ln p(\boldsymbol{A}_3; \boldsymbol{\Theta}_3, \boldsymbol{\Lambda}_3)}{\partial \boldsymbol{\Theta}_3} \right\rangle = \boldsymbol{K}_3(\boldsymbol{\Theta}_3, \boldsymbol{\Lambda}_3)|\,\mathrm{sds}\,\boldsymbol{\Delta}_3(\boldsymbol{A}_3; \boldsymbol{\Theta}_3, \boldsymbol{\Lambda}_3)\rangle \qquad (\mu\text{-a.e.}), \tag{4.23}$$

where $\boldsymbol{K}_3(\boldsymbol{\Theta}_3, \boldsymbol{\Lambda}_3) = ({}_3\kappa_{\alpha\beta})$ is some $\frac{1}{2}m(m-1) \times \frac{1}{2}m(m-1)$ nonsingular matrix. Let us denote $|\boldsymbol{\Delta}_3\rangle$, corresponding to (4.14), as

$$|\boldsymbol{\Delta}_3\rangle =: (\Delta_1^{\dagger}, \Delta_2^{\dagger}, \ldots, \Delta_{m-1}^{\dagger}, \Delta_m^{\dagger}, \ldots, \Delta_{2m-2}^{\dagger}, \ldots, \Delta_{(m+1)(m-2)/2}^{\dagger}, \Delta_{m(m-1)/2}^{\dagger})^t \quad \left(\frac{1}{2}m(m-1) \times 1\right).$$

Then, we can represent the RHS vector of Equation (4.23) as

$$\boldsymbol{K}_3|\,\mathrm{sds}\,\boldsymbol{\Delta}_3\rangle = \left(\sum_{l=1}^{m(m-1)/2} {}_3\kappa_{1l}\Delta_l^{\dagger}, \sum_{l=1}^{m(m-1)/2} {}_3\kappa_{2l}\Delta_l^{\dagger}, \ldots, \sum_{l=1}^{m(m-1)/2} {}_3\kappa_{m(m-1)/2,l}\Delta_l^{\dagger} \right)^t. \tag{4.24}$$

On the other hand, making use of matrixization operation (4.14), with $n = m$, we have

$$\mathbb{M}\left(\mathbb{F}_m^K \left| \mathrm{sds}\, \frac{\partial^K \ln p(\boldsymbol{A}_3; \boldsymbol{\Theta}_3, \boldsymbol{\Lambda}_3)}{\partial \boldsymbol{\Theta}_3} \right\rangle \right) = \frac{\partial^S \ln p(\boldsymbol{A}_3; \boldsymbol{\Theta}_3, \boldsymbol{\Lambda}_3)}{\partial \boldsymbol{\Theta}_3},$$

and

$$\mathbb{M}(\mathbb{F}_m^K(\boldsymbol{K}_3(\boldsymbol{\Theta}_3, \boldsymbol{\Lambda}_3)|\,\mathrm{sds}\,\boldsymbol{\Delta}_3(\boldsymbol{A}_3; \boldsymbol{\Theta}_3, \boldsymbol{\Lambda}_3)\rangle)) = \Xi_3(\boldsymbol{\Theta}_3, \boldsymbol{\Lambda}_3)(|\,\mathrm{sds}\,\boldsymbol{\Delta}_3(\boldsymbol{A}_3; \boldsymbol{\Theta}_3, \boldsymbol{\Lambda}_3)\rangle \otimes \boldsymbol{I}_m) \qquad (\text{say}), \tag{4.25}$$

where $\mathbb{F}_m^K(m^2 \times \frac{1}{2}m(m-1))$ is the *full stacking operator* such that $\mathbb{F}_m^K|\operatorname{sds} \boldsymbol{Z}\rangle = |\operatorname{fls} \boldsymbol{Z}\rangle$ for an $m \times m$ skew-symmetric matrix $\boldsymbol{Z}$, and $\Xi_3 = ({}_3\xi_{ij})(m \times \frac{1}{2}m^2(m-1))$. Vectorizing (4.25) by sub-diagonal stacking the columns, we find that the $\{i + m(j-1) - \frac{1}{2}j(j-1)\}$-th row of the vector is

$$\sum_{l=1}^{m(m-1)/2} {}_3\xi_{i,j+(l-1)m}\Delta_l^\dagger \qquad (1 \le i \le m,\ 1 \le j \le m).$$

Namely, by (4.24), this must coincides with

$$\sum_{l=1}^{m(m-1)/2} {}_3\kappa_{i+m(j-1)-j(j+1)/2,l}\Delta_l^\dagger \qquad (1 \le i \le m,\ 1 \le j \le m),$$

and comparing the resultant coefficients of $\Delta_l^\dagger$ with those of (4.24), we have the relation (4.12) between the components of the matrix Ξ_3 and $\boldsymbol{K}_3$. Thus, there exists a precision matrix $\Xi_3(\boldsymbol{\Theta}_3, \boldsymbol{\Lambda}_3)$ satisfying the desired result (4.11), which completes the proof. □

Remark 4.1. The commutation operator $\mathbb{K}_{mn}(mn \times mn)$ for general matrices, the full stacking operators $\mathbb{F}_m^S(m^2 \times \frac{1}{2}m(m+1))$ for symmetric matrices and $\mathbb{F}_m^K(m^2 \times \frac{1}{2}m(m-1))$ for skew-symmetric matrices can be, respectively, given by

$$\mathbb{K}_{mn} = \sum_{i=1}^{m}\sum_{j=1}^{n}(|i\rangle_m\langle j|_n) \otimes (|j\rangle_n\langle i|_m),$$

$$\text{with} \quad \langle i|_m = (|i\rangle_m) = \Big(\underbrace{0,\ldots,0}_{i-1},1,\underbrace{0,\ldots,0}_{m-i}\Big),$$

$$\mathbb{F}_m^S = \sum_{i>j} \big||i\rangle\langle j| + |j\rangle\langle i|\big\rangle \left\langle (j-1)m + i - \frac{1}{2}j(j-1)\right| + \sum_{i=1}^{m} \big||i\rangle\langle i|\big\rangle \left\langle (i-1)m + i - \frac{1}{2}i(i-1)\right|,$$

and

$$\mathbb{F}_m^K = \sum_{i>j} \big|\operatorname{fls}(|i\rangle\langle j| - |j\rangle\langle i|)\big\rangle \left\langle (j-1)m + i - \frac{1}{2}j(j+1)\right|,$$

where $|i\rangle = |i\rangle_m$ and $\langle i| = \langle i|_m$.

Now we are in a position to represent parametric minimum uncertainty distributions based on the corresponding parametric statistical fundamental equations given in Theorem 4.1. Their resultant distributions belong to some modified exponential type families whose densities have different expressions from the usual ones (cf. Brown (1986)).

THEOREM 4.2. *Under the conditions of Theorem* 3.1 *and the same notations as those of Theorem* 4.1, *the probability density functions with respect to the* σ-*finite measure* μ *of the minimum uncertainty distributions of* $m \times n$ *real random matrices* $\boldsymbol{A}_i$ $(i = 1, 2, 3)$, *are given by*

$$p(\boldsymbol{A}_i; \boldsymbol{\Theta}_i, \boldsymbol{\Lambda}_i) = c_i(\boldsymbol{\Theta}_i, \boldsymbol{\Lambda}) \exp \int \mathrm{tr}\{\tilde{\boldsymbol{\Xi}}_i(\boldsymbol{\Theta}_i, \boldsymbol{\Lambda})(|\tilde{\boldsymbol{\Delta}}_i(\boldsymbol{A}_i; \boldsymbol{\Theta}_i, \boldsymbol{\Lambda}_i)\rangle \otimes \boldsymbol{I}_r) d\boldsymbol{\Theta}_i^t\}$$

$(i = 1, 2, 3)$, *where* $\tilde{\boldsymbol{\Delta}}_i$ *and* $\tilde{\boldsymbol{\Xi}}_i$ *are the corresponding quantities* $\boldsymbol{\Delta}_i$ *and* $\boldsymbol{\Xi}_i$ *considered in the previous theorem, and where* $\boldsymbol{I}_r$ *is the unit matrix such that* $r = n$ *if* $i = 1$ *and that* $r = m$ *if* $i = 2$ *and* 3. *Further,* $c_i(\boldsymbol{A}_i; \boldsymbol{\Lambda}_i)$ $(i = 1, 2, 3)$ *are the scaler-valued normalizing functions of* $\boldsymbol{A}_i$, $\boldsymbol{\Lambda}_i$ *which are independent of* $\boldsymbol{\Theta}_i$ *and satisfy the conditions*

$$\int p(\boldsymbol{A}_i; \boldsymbol{\Theta}_i, \boldsymbol{\Lambda}_i)(d\boldsymbol{A}_i) = 1 \qquad (i = 1, 2, 3)$$

where integration is done over the domain of $\boldsymbol{A}_i$ *and the notation* $(d\boldsymbol{A}_i)$ $(i = 1, 2, 3)$ *is the measure of the differential matrices* $d\boldsymbol{A}_i$ $(i = 1, 2, 3)$ *expressed by the exterior products defined as before.*

PROOF. It suffices to prove the case where the general real parameter $\boldsymbol{\Theta}_1(m \times n)$ is the underlying parameter matrix, because we can prove the cases of $i = 2$ and 3 in the same manner. From the matrix-valued parametric statistical fundamental equation in Theorem 4.1, we have

$$\frac{\partial \ln p(\boldsymbol{A}_1; \boldsymbol{\Theta}_1, \boldsymbol{\Lambda}_1)}{\partial \boldsymbol{\Theta}_1} = \boldsymbol{\Xi}_1(\boldsymbol{\Theta}_1, \boldsymbol{\Lambda}_1)(|\boldsymbol{\Delta}_1(\boldsymbol{A}_1; \boldsymbol{\Theta}_1, \boldsymbol{\Lambda}_1)\rangle \otimes \boldsymbol{I}_n) \qquad (\mu\text{-a.e.}).$$

Hence,

$$\mathrm{tr}\left\{\frac{\partial \ln p(\boldsymbol{A}_1; \boldsymbol{\Theta}_1, \boldsymbol{\Lambda}_1)}{\partial \boldsymbol{\Theta}_1} d\boldsymbol{\Theta}_1^t\right\} = \mathrm{tr}\{\boldsymbol{\Xi}_1(\boldsymbol{\Theta}_1, \boldsymbol{\Lambda}_1)(|\boldsymbol{\Delta}_1(\boldsymbol{A}_1; \boldsymbol{\Theta}_1, \boldsymbol{\Lambda}_1)\rangle \otimes \boldsymbol{I}_n) d\boldsymbol{\Theta}_1^t\}.$$

Integrating with respect to $\boldsymbol{\Theta}_1(m \times n)$, we have

$$\ln p(\boldsymbol{A}_1; \boldsymbol{\Theta}_1, \boldsymbol{\Lambda}_1) \propto \int \mathrm{tr}\{\boldsymbol{\Xi}_1(\boldsymbol{\Theta}_1, \boldsymbol{\Lambda}_1)(|\boldsymbol{\Delta}_1(\boldsymbol{A}_1; \boldsymbol{\Theta}_1, \boldsymbol{\Lambda}_1)\rangle \otimes \boldsymbol{I}_n) d\boldsymbol{\Theta}_1^t\}.$$

Namely, we get the desired expression:

$$p(\boldsymbol{A}_1, \boldsymbol{\Theta}_1, \boldsymbol{\Lambda}_1) = c_1(\boldsymbol{A}_1; \boldsymbol{\Lambda}_1) \int \mathrm{tr}\{\boldsymbol{\Xi}_1(\boldsymbol{\Theta}_1, \boldsymbol{\Lambda}_1)(|\boldsymbol{\Delta}_1(\boldsymbol{A}_1; \boldsymbol{\Theta}_1, \boldsymbol{\Lambda}_1)\rangle \otimes \boldsymbol{I}_n) d\boldsymbol{\Theta}_1^t\}. \ \square$$

The following result is immediately obtained:

COROLLARY 4.1. *Under the same conditions as in Theorem* 4.2, *if we take*

$$\tilde{\boldsymbol{\Xi}}_i(\boldsymbol{\Theta}_i, \boldsymbol{\Lambda}_i) =: \boldsymbol{\Xi}_i(\boldsymbol{\Theta}_i, \boldsymbol{\Lambda}_i)(\boldsymbol{I}_m \otimes \langle \boldsymbol{I}_r|)(\mathbb{K}_{mr} \otimes \boldsymbol{I}_r),$$

where $\Xi_i(\boldsymbol{\Theta}_i, \boldsymbol{\Lambda}_i)$ *is an* $m \times m$ *matrix, then we have*

$$\frac{\partial \ln p(\boldsymbol{A}_i; \boldsymbol{\Theta}_i, \boldsymbol{\Lambda}_i)}{\partial \boldsymbol{\Theta}_i} = \boldsymbol{\Xi}_i(\boldsymbol{\Theta}_i, \boldsymbol{\Lambda}_i)\tilde{\boldsymbol{\Delta}}_i(\boldsymbol{A}_i; \boldsymbol{\Theta}_i, \boldsymbol{\Lambda}_i)$$

and then

$$p(\boldsymbol{A}_i; \boldsymbol{\Theta}_i, \boldsymbol{\Lambda}_i) = c_i(\boldsymbol{A}_i; \boldsymbol{\Lambda}_i) \exp \int \mathrm{tr}\{\boldsymbol{\Xi}_i(\boldsymbol{\Theta}_i, \boldsymbol{\Lambda}_i)\tilde{\boldsymbol{\Delta}}_i(\boldsymbol{A}_i; \boldsymbol{\Theta}_i, \boldsymbol{\Lambda}_i) d\boldsymbol{\Theta}_i^t\},$$

for $i = 1, 2, 3$.

PROOF. Applying the relation $(\boldsymbol{I}_m \otimes \langle \boldsymbol{I}_r|)(\mathbb{K}_{mr} \otimes \boldsymbol{I}_r)(|\boldsymbol{\Delta}_i\rangle \otimes \boldsymbol{I}_r) = \boldsymbol{\Delta}_i$ (cf. Matsunawa and Zhao (1994)) to Theorem 4.1, we immediately obtain the desired result. □

5. Examples of multivariate parametric statistical models

In this section we construct multivariate distributions as basic parametric statistical models based on the fundamental statistical equations and the matrixization operation $\mathbb{M}(|\bullet\rangle)$, considered in the previous section. In the following examples, when μ is the Lebesgue measure, we use $f(\boldsymbol{A}; \boldsymbol{\Theta}, \cdot)$ as the functional notation of probability density functions instead of $p(\boldsymbol{A}; \boldsymbol{\Theta}, \cdot)$.

5.1 *Multivariate normal distribution*

$$f(\boldsymbol{A}; \boldsymbol{\Theta}, \boldsymbol{\Sigma}) = (2\pi)^{-mn/2}|\boldsymbol{\Sigma}|^{-n/2} \,\mathrm{etr}\left\{-\frac{1}{2}(\boldsymbol{A} - \boldsymbol{\Theta})^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{A} - \boldsymbol{\Theta})\right\},$$

where $\boldsymbol{A}$ is an $m \times n$ real random matrix, $\boldsymbol{\Theta}$ is an $m \times n$ real primarily interested parameter matrix, $\boldsymbol{\Sigma}(n \times n)$ is a positive definite secondary parameter matrices.

Derivation. Set the functions in Theorem 4.1 (i) without suffices as follows:

$$\boldsymbol{\Delta}(\boldsymbol{A}; \boldsymbol{\Theta}) =: -(\boldsymbol{A} - \boldsymbol{\Theta}) \qquad (m \times n)$$

and

$$\boldsymbol{\Xi}(\boldsymbol{\Sigma}) := \boldsymbol{\Sigma}^{-1}(\boldsymbol{I}_m \otimes \langle \boldsymbol{I}_n|)(\mathbb{K}_{mn} \otimes \boldsymbol{I}_n) \qquad (m \times mn^2).$$

Then, from Theorem 4.2, we get

$$\begin{aligned} f(\boldsymbol{A}; \boldsymbol{\Theta}, \boldsymbol{\Sigma}) &\propto \exp \int \mathrm{tr}\{\boldsymbol{\Xi}(\boldsymbol{\Sigma})(|\,\mathrm{fls}\,\boldsymbol{\Delta}(\boldsymbol{A}; \boldsymbol{\Theta})\rangle \otimes \boldsymbol{I}_n) d\boldsymbol{\Theta}^t\} \\ &= \exp \int \mathrm{tr}\{\boldsymbol{\Sigma}^{-1}\mathbb{M}(|\boldsymbol{\Delta}(\boldsymbol{A}; \boldsymbol{\Theta})\rangle) d\boldsymbol{\Theta}^t\} = \exp \int \mathrm{tr}\{\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}(\boldsymbol{A}; \boldsymbol{\Theta}) d\boldsymbol{\Theta}^t\} \\ &= \exp \int \mathrm{tr}\left[-\frac{1}{2}\left(\frac{\partial}{\partial \boldsymbol{\Theta}} \mathrm{tr}\{(\boldsymbol{A} - \boldsymbol{\Theta})^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{A} - \boldsymbol{\Theta})\}\right) d\boldsymbol{\Theta}^t\right] \\ &= \mathrm{etr}\left\{-\frac{1}{2}(\boldsymbol{A} - \boldsymbol{\Theta})^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{A} - \boldsymbol{\Theta})\right\}. \end{aligned}$$

Therefore, we can represent

$$f(\boldsymbol{A};\boldsymbol{\Theta},\boldsymbol{\Sigma}) = c_{MN}(\boldsymbol{A};\boldsymbol{\Sigma})\,\mathrm{etr}\left\{-\frac{1}{2}(\boldsymbol{A}-\boldsymbol{\Theta})^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{A}-\boldsymbol{\Theta})\right\}.$$

To determine $c_{MN}(\boldsymbol{A};\boldsymbol{\Sigma},\boldsymbol{\Gamma})$ we can use the well-known integral formula (cf. Muirhead (1982)):

$$\int \mathrm{etr}\left\{-\frac{1}{2}(\boldsymbol{A}-\boldsymbol{\Theta})^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{A}-\boldsymbol{\Theta})\right\}(d\boldsymbol{A}) = (2\pi)^{mn/2}|\boldsymbol{\Sigma}|^{n/2}$$

and, consequently,

$$c_{MN}(\boldsymbol{A};\boldsymbol{\Sigma}) = (2\pi)^{-mn/2}|\boldsymbol{\Sigma}|^{-n/2}.$$

5.2 *Multivariate symmetric normal distribution*

$$f(\boldsymbol{A};\boldsymbol{\Theta},\boldsymbol{\Sigma}) = \frac{1}{2^{m/2}(2\pi)^{m(m+1)/4}|\boldsymbol{\Sigma}|^{(m+1)/2}}\,\mathrm{etr}\left\{-\frac{1}{4}\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{A}-\boldsymbol{\Theta})\}^2\right\},$$

where $\boldsymbol{A}$ is an $m\times m$ symmetric real random matrix, $\boldsymbol{\Theta}$ is an $m\times m$ real symmetric primary parameter matrix, and $\boldsymbol{\Sigma}(m\times m)$ is a positive definite secondary parameter matrix.

Derivation. Set the functions in Theorem 4.1 (ii) as follows:

$$\boldsymbol{\Delta}(\boldsymbol{A};\boldsymbol{\Theta}) = -(\boldsymbol{A}-\boldsymbol{\Theta}) \quad \text{and} \quad \boldsymbol{\Xi}(\boldsymbol{\Sigma}) = \frac{1}{2}(\boldsymbol{\Sigma}^{-1})^2(\boldsymbol{I}_m\otimes\langle\boldsymbol{I}_m|)(\mathbb{K}_{mm}\otimes\boldsymbol{I}_m),$$

then from Theorem 4.2

$$\begin{aligned}
f(\boldsymbol{A};\boldsymbol{\Theta},\boldsymbol{\Sigma}) &\propto \exp\int \mathrm{tr}\Big\{-\frac{1}{2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{I}_m\otimes\langle\boldsymbol{I}_m|)(\mathbb{K}_{mm}\otimes\boldsymbol{I}_m)\\
&\qquad\cdot[\{\mathbb{F}_m^S|\,\mathrm{trs}(\boldsymbol{A}-\boldsymbol{\Theta})\rangle\}\otimes\boldsymbol{I}_m]\boldsymbol{\Sigma}^{-1}d\boldsymbol{\Theta}^t\Big\}\\
&= \exp\int -\frac{1}{2}\,\mathrm{tr}\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{A}-\boldsymbol{\Theta})\boldsymbol{\Sigma}^{-1}d\boldsymbol{\Theta}^t\}\\
&= \exp\int -\frac{1}{4}\,\mathrm{tr}\left\{\frac{\partial^S}{\partial\boldsymbol{\Theta}}[\mathrm{tr}\{(\boldsymbol{A}-\boldsymbol{\Theta})^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{A}-\boldsymbol{\Theta})\boldsymbol{\Sigma}^{-1}\}]d\boldsymbol{\Theta}^t\right\}\\
&= \mathrm{etr}\left\{-\frac{1}{4}(\boldsymbol{A}-\boldsymbol{\Theta})^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{A}-\boldsymbol{\Theta})\boldsymbol{\Sigma}^{-1}\right\}\\
&= \mathrm{etr}\left\{-\frac{1}{4}\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{A}-\boldsymbol{\Theta})\}^2\right\}.
\end{aligned}$$

Therefore, we can represent

$$f(\boldsymbol{A};\boldsymbol{\Theta},\boldsymbol{\Sigma}) = c_{SN}(\boldsymbol{A};\boldsymbol{\Sigma})\,\mathrm{etr}\left\{-\frac{1}{4}\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{A}-\boldsymbol{\Theta})\}^2\right\}.$$

It remains to determine the normalized constant put $c_{SN}(\boldsymbol{A};\boldsymbol{\Sigma})$. Since $\boldsymbol{\Sigma}^{-1}$ is positive definite and $\boldsymbol{A}-\boldsymbol{\Theta}$ belongs to the family $\mathbf{S}^{m\times m}$ of all $m\times m$ real symmetric matrices, the following transformation $\boldsymbol{Y} := 2^{-1/2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{A}-\boldsymbol{\Theta})\boldsymbol{\Sigma}^{-1} \in \mathbf{S}^{m\times m}$. Hence, the Jacobian is given by $J(\boldsymbol{A}\to\boldsymbol{Y}) = 2^{m(m+1)/4}|\boldsymbol{\Sigma}|^{(m+1)/2}$ and

$$\begin{aligned}&\int_{\mathbf{S}^{m\times m}} \mathrm{etr}\left\{-\frac{1}{4}\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{A}-\boldsymbol{\Theta})\}^2\right\}(d\boldsymbol{A})\\ &\quad= \int_{\mathbf{S}^{m\times m}} \mathrm{etr}\left\{-\frac{1}{2}\boldsymbol{Y}^2\right\}(d\boldsymbol{Y})2^{m(m+1)/4}|\boldsymbol{\Sigma}|^{(m+1)/2}\\ &\quad= 2^{m/2}\pi^{m(m+1)/4}2^{m(m+1)/4}|\boldsymbol{\Sigma}|^{(m+1)/2} = 2^{m/2}(2\pi)^{m(m+1)/4}|\boldsymbol{\Sigma}|^{(m+1)/2}.\end{aligned}$$

Consequently,

$$c_{SN}(\boldsymbol{\Sigma}) = 1/\{2^{m/2}(2\pi)^{m(m+1)/4}|\boldsymbol{\Sigma}|^{(m+1)/2}\}.$$

Remark 5.1. If $m=1$ and taking the correspondence $\boldsymbol{A}\leftrightarrow x$, $\boldsymbol{\Theta}\leftrightarrow\theta$, $\boldsymbol{\Sigma}\leftrightarrow\sigma/\sqrt{2}$, we get a univariate normal distribution $N(\theta,\sigma^2)$.

5.3 *Multivariate symmetric log-normal distribution*

$$\begin{aligned}f(\boldsymbol{A};\boldsymbol{\Theta},\boldsymbol{\Sigma}) &= \frac{1}{2^{m/2}\pi^{m(m+1)/4}|\boldsymbol{\Sigma}|^{(m+1)/2}}\left(\prod_{i<j}\frac{\ln\alpha_i-\ln\alpha_j}{\alpha_i-\alpha_j}\right)\\ &\quad\times\frac{1}{|\boldsymbol{A}|}\cdot\mathrm{etr}\left\{-\frac{1}{2}\{\boldsymbol{\Sigma}^{-1}(\ln\boldsymbol{A}-\boldsymbol{\Theta})\}^2\right\},\end{aligned}$$

where

$$\begin{aligned}&\boldsymbol{A},\boldsymbol{\Theta},\boldsymbol{\Sigma}\in\mathbf{S}^{m\times m},\quad \boldsymbol{A},\boldsymbol{\Theta},\boldsymbol{\Sigma}>0;\quad \boldsymbol{A}=\boldsymbol{T}\boldsymbol{A}_\Delta\boldsymbol{T}^t,\quad \boldsymbol{T}\boldsymbol{T}^t=\boldsymbol{I}_m\\ &\boldsymbol{A}_\Delta=\mathrm{diag}(\alpha_1,\alpha_2,\ldots,\alpha_m),\quad(\alpha_1>\alpha_2>\cdots>\alpha_m>0),\\ &\ln\boldsymbol{A}:=\boldsymbol{T}(\ln\boldsymbol{A}_\Delta)\boldsymbol{T}^t.\end{aligned}$$

Derivation. Set the functions in Theorem 4.1 (ii) as follows:

$$\boldsymbol{\Delta}(\boldsymbol{A};\boldsymbol{\Theta}) = -(\ln\boldsymbol{A}-\boldsymbol{\Theta})\quad\text{and}\quad\boldsymbol{\Xi}(\boldsymbol{\Sigma}) = (\boldsymbol{\Sigma}^{-1})^2(\boldsymbol{I}_m\otimes\langle\boldsymbol{I}_m|)(\mathbb{K}_{mm}\otimes\boldsymbol{I}_m),$$

then, from Theorem 4.2,

$$\begin{aligned}f(\boldsymbol{A};\boldsymbol{\Theta},\boldsymbol{\Sigma}) &\propto \exp\int -\mathrm{tr}[\boldsymbol{\Sigma}^{-1}(\boldsymbol{I}_m\otimes\langle\boldsymbol{I}_m|)(\mathbb{K}_{mm}\otimes\boldsymbol{I}_m)\\ &\qquad\cdot[\{\mathbb{F}_m^S|\,\mathrm{trs}(\ln\boldsymbol{A}-\boldsymbol{\Theta})\rangle\}\otimes\boldsymbol{I}_m]\boldsymbol{\Sigma}^{-1}d\boldsymbol{\Theta}^t]\\ &= \exp\int -\mathrm{tr}\{\boldsymbol{\Sigma}^{-1}(\ln\boldsymbol{A}-\boldsymbol{\Theta})\boldsymbol{\Sigma}^{-1}d\boldsymbol{\Theta}^t\}\\ &= \exp\int -\mathrm{tr}\left[\frac{1}{2}\frac{\partial^S}{\partial\boldsymbol{\Theta}}\mathrm{tr}\{(\ln\boldsymbol{A}-\boldsymbol{\Theta})^t\boldsymbol{\Sigma}^{-1}(\ln\boldsymbol{A}-\boldsymbol{\Theta})\boldsymbol{\Sigma}^{-1}\}\cdot d\boldsymbol{\Theta}^t\right]\\ &= \mathrm{etr}\left\{-\frac{1}{2}((\ln\boldsymbol{A}-\boldsymbol{\Theta})^t\boldsymbol{\Sigma}^{-1}(\ln\boldsymbol{A}-\boldsymbol{\Theta})\boldsymbol{\Sigma}^{-1})\right\}.\end{aligned}$$

Therefore,

$$f(\boldsymbol{A};\boldsymbol{\Theta},\boldsymbol{\Sigma}) = c_{SLN}(\boldsymbol{A};\boldsymbol{\Sigma})\,\mathrm{etr}\left\{-\frac{1}{2}\{\boldsymbol{\Sigma}^{-1}(\ln \boldsymbol{A}-\boldsymbol{\Theta})\}^2\right\}.$$

To determine the normalizing constant $c_{SLN}(\boldsymbol{A};\boldsymbol{\Sigma})$, let $\boldsymbol{Y} = \ln \boldsymbol{A} (\in \mathbf{S}^{m\times m})$, then the Jacobian of the transformation is

$$J(\boldsymbol{Y}\to\boldsymbol{A}) = \left|\frac{\partial\,\mathrm{vec}(\boldsymbol{Y})^t}{\partial\,\mathrm{vec}(\boldsymbol{A})}\right|_+ = \left(\prod_{i<j}\frac{\ln\alpha_i-\ln\alpha_j}{\alpha_i-\alpha_j}\right)\frac{1}{|\boldsymbol{A}|}.$$

Suppose that $\boldsymbol{Y}$ is distributed according to a multivariate symmetric log-normal distribution, then

$$\int_{\mathbf{S}^{n\times n}}\mathrm{etr}\left\{-\frac{1}{2}\{\boldsymbol{\Sigma}^{-1}(\ln\boldsymbol{A}-\boldsymbol{\Theta})\}^2\right\}\cdot\left(\prod_{i<j}\frac{\ln\alpha_i-\ln\alpha_j}{\alpha_i-\alpha_j}\right)\frac{1}{|\boldsymbol{A}|}(d\boldsymbol{A})$$
$$= \int_{\mathbf{S}^{n\times n}}\mathrm{etr}\left\{-\frac{1}{2}\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}-\boldsymbol{\Theta})\}^2\right\}(d\boldsymbol{Y}) = 2^{m/2}\pi^{m(m+1)/4}|\boldsymbol{\Sigma}|^{(m+1)/2}.$$

Hence, we have

$$c_{SLN}(\boldsymbol{A};\boldsymbol{\Sigma}) = \frac{1}{2^{m/2}\pi^{m(m+1)/4}|\boldsymbol{\Sigma}|^{(m+1)/2}}\left(\prod_{i<j}\frac{\ln\alpha_i-\ln\alpha_j}{\alpha_i-\alpha_j}\right)\frac{1}{|\boldsymbol{A}|}.$$

Remark 5.2. If $m = 1$ and taking the correspondence $\boldsymbol{A}\leftrightarrow x$, $\boldsymbol{\Theta}\leftrightarrow\theta$, $\boldsymbol{\Sigma}\leftrightarrow\sigma/\sqrt{2}$, we get a univariate log-normal distribution:

$$f(x;\theta,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma x}\exp\left\{-\frac{1}{2}\left(\frac{\ln x-\theta}{\sigma}\right)^2\right\} \qquad (x>0,\theta>0,\sigma>0).$$

5.4 *Multivariate symmetric logarithmic gamma distribution*

$$f(\boldsymbol{A};\boldsymbol{\Theta},\boldsymbol{M},b) = \frac{|\boldsymbol{\Theta}|^b}{\Gamma_n(b)}|\ln\boldsymbol{A}-\ln\boldsymbol{M}|^{b-(n+1)/2}$$
$$\times\left(\prod_{i<j}\frac{\ln\alpha_i-\ln\alpha_j}{\alpha_i-\alpha_j}\right)\frac{1}{|\boldsymbol{A}|}\cdot\mathrm{etr}\{-\boldsymbol{\Theta}(\ln\boldsymbol{A}-\ln\boldsymbol{M})\},$$

where

$$\boldsymbol{A},\boldsymbol{\Theta},\boldsymbol{M}\in\mathbf{S}^{n\times n},\qquad \boldsymbol{A},\boldsymbol{\Theta},\boldsymbol{M}>0,\qquad \ln\boldsymbol{A}-\ln\boldsymbol{M}>0,\qquad b>(n-1)/2,$$
$$\boldsymbol{A}=\boldsymbol{T}\boldsymbol{A}_\Delta\boldsymbol{T}^t,\qquad \boldsymbol{A}_\Delta=\mathrm{diag}(\alpha_1,\alpha_2,\ldots,\alpha_n),\qquad (\alpha_1>\alpha_2>\cdots>\alpha_n>0),$$

and

$$\Gamma_n(b) = \int_{\boldsymbol{X}>0} \operatorname{etr}(-\boldsymbol{X})|\boldsymbol{X}|^{b-(n+1)/2}(d\boldsymbol{X}) \qquad (\boldsymbol{X} > 0).$$

Derivation. Set the functions in Theorem 4.1 (ii) as follows:

$$\boldsymbol{\Delta}(\boldsymbol{A};\boldsymbol{\Theta},\boldsymbol{M},b) = -(\ln\boldsymbol{A} - \ln\boldsymbol{M}) + b\boldsymbol{\Theta}^{-1} \quad \text{and} \quad \boldsymbol{\Xi} = (\boldsymbol{I}_n \otimes \langle \boldsymbol{I}_n|)(\mathbb{K}_{nn} \otimes \boldsymbol{I}_n)$$

then, from Theorem 4.2,

$$\begin{aligned} f(\boldsymbol{A};\boldsymbol{\Theta},\boldsymbol{M},b) &\propto \exp\int \operatorname{tr}\{-(\ln\boldsymbol{A} - \ln\boldsymbol{M}) + b\boldsymbol{\Theta}^{-1}\}d\boldsymbol{\Theta} \\ &= \exp\int \operatorname{tr}\left\{\frac{\partial^S \operatorname{tr}(-\boldsymbol{\Theta}(\ln\boldsymbol{A} - \ln\boldsymbol{M}))}{\partial\boldsymbol{\Theta}} + \frac{\partial^S \ln|\boldsymbol{\Theta}|^b}{\partial\boldsymbol{\Theta}}\right\}d\boldsymbol{\Theta}. \end{aligned}$$

Since for $\boldsymbol{\Theta}$, $\boldsymbol{Y} \in \mathbf{S}^{n\times n}$

$$\frac{\partial^S \operatorname{tr}(\boldsymbol{\Theta}\boldsymbol{Y})}{\partial\boldsymbol{\Theta}} = \boldsymbol{Y} \quad \text{and} \quad \frac{\partial^S \ln|\boldsymbol{\Theta}|^b}{\partial\boldsymbol{\Theta}} = b\cdot\boldsymbol{\Theta}^{-1},$$

then we have

$$\begin{aligned} f(\boldsymbol{A};\boldsymbol{\Theta},\boldsymbol{M},b) &\propto \exp\int \operatorname{tr}\{-(\ln\boldsymbol{A} - \ln\boldsymbol{M}) + b\boldsymbol{\Theta}^{-1}\}d\boldsymbol{\Theta} \\ &= c_{SLG}(\boldsymbol{A};\boldsymbol{M},b)\cdot\exp(\operatorname{tr}(-\boldsymbol{\Theta}(\ln\boldsymbol{A} - \ln\boldsymbol{M})) + \ln|\boldsymbol{\Theta}|^b) \\ &= c_{SLG}(\boldsymbol{A};\boldsymbol{M},b)\cdot|\boldsymbol{\Theta}|^b \operatorname{etr}(-\boldsymbol{\Theta}(\ln\boldsymbol{A} - \ln\boldsymbol{M})). \end{aligned}$$

To determine the normalizing constant $c_{SLG}(\boldsymbol{A};\boldsymbol{M},b)$, we modify the multivariate gamma distribution as

$$\begin{aligned} \Gamma_n(b) = \int_D &|\boldsymbol{\Theta}|^b \operatorname{etr}(-\boldsymbol{\Theta}(\ln\boldsymbol{A} - \ln\boldsymbol{M}))\cdot|\ln\boldsymbol{A} - \ln\boldsymbol{M}|^{b-(n+1)/2} \\ &\cdot\left|\frac{\partial^S}{\partial\boldsymbol{A}}(\ln\boldsymbol{A} - \ln\boldsymbol{M})\right|_+ (d\boldsymbol{A}), \end{aligned}$$

where $D = \{\boldsymbol{A} > 0, \boldsymbol{M} > 0, \boldsymbol{\Theta} > 0, \ln\boldsymbol{A} - \ln\boldsymbol{M} > 0\}$. Then, we get

$$\begin{aligned} c_{SLG}(\boldsymbol{A};\boldsymbol{M},b) &= \frac{1}{\Gamma_n(b)}|\ln\boldsymbol{A} - \ln\boldsymbol{M}|^{b-(n+1)/2}\cdot\left|\frac{\partial^S \ln\boldsymbol{A}}{\partial\boldsymbol{A}}\right|_+ \\ &= \frac{1}{\Gamma_n(b)}|\ln\boldsymbol{A} - \ln\boldsymbol{M}|^{b-(n+1)/2}\cdot\left(\prod_{i<j}\frac{\ln\alpha_i - \ln\alpha_j}{\alpha_i - \alpha_j}\right)\frac{1}{|\boldsymbol{A}|}. \end{aligned}$$

5.5 *Multivariate skew-symmetric normal distribution*

$$f(\boldsymbol{A};\boldsymbol{\Theta},\boldsymbol{\Sigma}) = \frac{1}{\pi^{m(m-1)/4}2^{m(m-1)/4}|\boldsymbol{\Sigma}|^{(m-1)/2}}\operatorname{etr}\left\{\frac{1}{4}\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{A}-\boldsymbol{\Theta})\}^2\right\},$$

where $\boldsymbol{A}$ is an $m \times m$ skew-symmetric real random matrix, $\boldsymbol{\Theta}$ is an $m \times m$ real skew-symmetric primary parameter matrix, and $\boldsymbol{\Sigma}(m \times m)$ is a positive definite secondary parameter matrix.

Derivation. Set the functions in Theorem 4.1 (ii) as follows:

$$\boldsymbol{\Delta}(\boldsymbol{A};\boldsymbol{\Theta},\boldsymbol{\Sigma}) = -(\boldsymbol{A}-\boldsymbol{\Theta}) \quad \text{and} \quad \Xi(\boldsymbol{\Sigma}) = \frac{1}{2}(\boldsymbol{\Sigma}^{-1})^2(\boldsymbol{I}_m \otimes \langle \boldsymbol{I}_m|)(\mathbb{K}_{mm}\otimes \boldsymbol{I}_m),$$

then from Theorem 4.2 for skew-symmetric $\boldsymbol{\Theta}$

$$\begin{aligned} f(\boldsymbol{A};\boldsymbol{\Theta},\boldsymbol{\Sigma}) &\propto \exp\int -\frac{1}{2}\operatorname{tr}[\boldsymbol{\Sigma}^{-1}(\boldsymbol{I}_m \otimes \langle \boldsymbol{I}_m|)(\mathbb{K}_{mm}\otimes \boldsymbol{I}_m) \\ &\qquad \cdot [\{\mathbb{F}_m^k|\operatorname{sds}(\boldsymbol{A}-\boldsymbol{\Theta})\rangle\}\otimes \boldsymbol{I}_m]\boldsymbol{\Sigma}^{-1}d\boldsymbol{\Theta}^t] \\ &= \exp\int -\frac{1}{2}\operatorname{tr}\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{A}-\boldsymbol{\Theta})\boldsymbol{\Sigma}^{-1}d\boldsymbol{\Theta}^t\} \\ &= \exp\int -\frac{1}{4}\operatorname{tr}\left\{\frac{\partial^k}{\partial\boldsymbol{\Theta}}[\operatorname{tr}\{(\boldsymbol{A}-\boldsymbol{\Theta})^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{A}-\boldsymbol{\Theta})\boldsymbol{\Sigma}^{-1}\}]d\boldsymbol{\Theta}^t\right\} \\ &= \operatorname{etr}\left\{-\frac{1}{4}(\boldsymbol{A}-\boldsymbol{\Theta})^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{A}-\boldsymbol{\Theta})\boldsymbol{\Sigma}^{-1}\right\} = \operatorname{etr}\left\{\frac{1}{4}\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{A}-\boldsymbol{\Theta})\}^2\right\}. \end{aligned}$$

Therefore, we can represent as

$$f(\boldsymbol{A};\boldsymbol{\Theta},\boldsymbol{\Sigma}) = c_{SSN}(\boldsymbol{A};\boldsymbol{\Sigma})\operatorname{etr}\left\{\frac{1}{4}\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{A}-\boldsymbol{\Theta})\}^2\right\}.$$

It remains to determine the normalized constant put $c_{SSN}(\boldsymbol{A};\boldsymbol{\Sigma})$. Since $\boldsymbol{\Sigma}^{-1}$ is positive definite and $\boldsymbol{A}-\boldsymbol{\Theta}$ belongs to the family $\mathbf{K}^{m\times m}$ of all $m\times m$ real symmetric matrices, then the following transformation $\boldsymbol{Z} := 2^{-1/2}\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{A}-\Theta)\boldsymbol{\Sigma}^{-1/2} \in \mathbf{K}^{m\times m}$. Hence, the Jacobian is given by $J(\boldsymbol{A}\to\boldsymbol{Z}) = 2^{m(m-1)/4}|\boldsymbol{\Sigma}|^{(m-1)/2}$ and

$$\begin{aligned} &\int_{\mathbf{K}^{m\times m}} \operatorname{etr}\left\{\frac{1}{4}\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{A}-\boldsymbol{\Theta})\}^2\right\}(d\boldsymbol{A}) \\ &= \int_{\mathbf{K}^{m\times m}} \operatorname{etr}\left\{\frac{1}{2}\boldsymbol{Z}^2\right\}(d\boldsymbol{Z})2^{m(m-1)/4}|\boldsymbol{\Sigma}|^{(m-1)/2} \\ &= \int_{\mathbf{K}^{m\times m}} \exp\left\{-\left\langle d\boldsymbol{Z}\left|\frac{1}{2}\boldsymbol{I}_{k\times k}\right|\boldsymbol{Z}^2\right\rangle\right\}2^{m(m-1)/4}|\boldsymbol{\Sigma}|^{(m-1)/2} \\ &= \prod_{i<j}^{m}\left[\int_{\mathbf{K}^{m\times m}} \exp\left\{-\frac{1}{2}(\sqrt{2}z_{ij})^2\right\}dz_{ij}\right]\cdot 2^{m(m-1)/4}|\boldsymbol{\Sigma}|^{(m-1)/2} \\ &= \prod_{i<j}^{m}\left[\int_{\mathbf{K}^{m\times m}} \exp\left(-\frac{t_{ij}^2}{2}\right)\frac{dt_{ij}}{\sqrt{2}}\right]\cdot 2^{m(m-1)/4}|\boldsymbol{\Sigma}|^{(m-1)/2} \\ &= \pi^{m(m-1)/4}2^{m(m-1)/4}|\boldsymbol{\Sigma}|^{(m-1)/2}. \end{aligned}$$

Consequently,

$$c_{SSN}(\mathbf{\Sigma}) = 1/\{\pi^{m(m-1)/4}2^{m(m-1)/4}|\mathbf{\Sigma}|^{(m-1)/2}\}.$$

5.6 *Multivariate generalized power series distribution*

$$q(\boldsymbol{A};\boldsymbol{\Theta}) = C(\boldsymbol{A}) \cdot \prod_{i=1}^{k} \theta_i^{a_i}/\omega(\boldsymbol{A};\boldsymbol{\Theta}), \quad \text{with} \quad \omega(\boldsymbol{A},\boldsymbol{\Theta}) = \sum_{\boldsymbol{A}\in T} C(\boldsymbol{A}) \prod_{i=1}^{k} \theta_i^{a_i} > 0,$$

where $C(\boldsymbol{A}) > 0$; $\boldsymbol{\Theta} = (\theta_1, \dots, \theta_k)^t$, $\theta_i > 0$; $\boldsymbol{A} = (a_1, \dots, a_k)^t \in T$, a countable subset without any limit point of a k-fold Cartesian product of the set R of real numbers. Namely,

$$T = \{(a_1, \dots, a_k)\} \subset R_1 \times R_2 \times \cdots \times R_k, \quad \text{with} \quad R_i = R, \ i = 1, 2, \dots, k.$$

Derivation. Set the functions in Corollary 4.1 as

$$\tilde{\boldsymbol{\Delta}}(\boldsymbol{A};\boldsymbol{\Theta}) = \left(\frac{a_1}{\theta_1} - \frac{\partial}{\partial\theta_1}\ln(\omega(\boldsymbol{A};\boldsymbol{\Theta}))^{1/k}, \dots, \frac{a_k}{\theta_k} - \frac{\partial}{\partial\theta_k}\ln(\omega(\boldsymbol{A};\boldsymbol{\Theta}))^{1/k}\right) \quad \text{and}$$

$\boldsymbol{\Xi}^\dagger(\theta) = 1$.

Then, we have

$$\begin{aligned}
q(\boldsymbol{A};\boldsymbol{\Theta}) &\propto \exp\int \operatorname{tr}[\boldsymbol{\Xi}^\dagger(\theta)\tilde{\boldsymbol{\Delta}}(\boldsymbol{A};\boldsymbol{\Theta})d\boldsymbol{\Theta}^t] \\
&= \exp\int \sum_{i=1}^{k}\left\{\frac{a_i}{\theta_i} - \frac{\partial}{\partial\theta_i}\ln(\omega(\boldsymbol{A};\boldsymbol{\Theta}))^{1/k}\right\}d\theta_i \\
&= \exp\left[\sum_{i=1}^{k}\left\{a_i\ln\theta_i - \frac{1}{k}\ln\omega(\boldsymbol{A};\boldsymbol{\Theta})\right\}\right] = \left(\prod_{i=1}^{k}\theta_i^{a_i}\right)\bigg/\ \omega(\boldsymbol{A};\boldsymbol{\Theta}).
\end{aligned}$$

Thus, we can take the functional form of the probability density function as

$$\begin{aligned}
q(\boldsymbol{A};\boldsymbol{\Theta}) &= c_{\mathrm{KPS}}(\boldsymbol{A})\left(\prod_{i=1}^{k}\theta_i^{a_i}\right)\bigg/\ \omega(\boldsymbol{A};\boldsymbol{\Theta}) \\
&= c_{\mathrm{KPS}}(\boldsymbol{A})\left(\prod_{i=1}^{k}\theta_i^{a_i}\right)\bigg/\ \sum_{\boldsymbol{A}\in T}C(\boldsymbol{A})\prod_{i=1}^{k}\theta_i^{a_i},
\end{aligned}$$

hence we get $c_{\mathrm{KPS}}(\boldsymbol{A}) = C(\boldsymbol{A})$.

Remark 5.1. This distribution contains important multivariate discrete distributions. If we replace R with I, we get the set of nonnegative integers so that $\sum_{\boldsymbol{A}\in T} C(\boldsymbol{A})\prod_{i=1}^{k}\theta_i^{a_i}$ is the power series expansion of $\omega(\cdot;\theta_1, \dots, \theta_k)$ in $\theta_1, \dots, \theta_k$.

Under this situation, (i) if we set $\omega(\boldsymbol{A}, \boldsymbol{\Theta}) = (1 + \theta_1 + \cdots + \theta_k)^n$, we have the *multinomial distribution*:

$$q(a_1, \ldots, a_k; p_1, \ldots, p_k) = \frac{n!}{a_1! \cdots a_k!(n - \sum_{j=1}^k a_j)!} p_1^{a_1} \cdots p_k^{a_k} \left(1 - \sum_{j=1}^k p_j\right)^{n - \sum_{j=1}^k a_j},$$

where $p_i = \theta_i/(1 + \sum_{j=1}^k \theta_j)$, $i = 1, \ldots, k$, with $0 < \sum_{j=1}^k p_j < 1$ and $0 \leq \sum_{j=1}^k a_j \leq n$,

(ii) if we set $\omega(\boldsymbol{A}, \boldsymbol{\Theta}) = (1 - \theta_1 - \cdots - \theta_k)^{-n}$, we have the *negative multinomial distribution*:

$$q(a_1, \ldots, a_k; p_1, \ldots, p_k) = \frac{(n + \sum_{j=1}^k a_j - 1)!}{a_1! \cdots a_k!(n-1)!} p_1^{a_1} \cdots p_k^{a_k} \left(1 - \sum_{j=1}^k p_j\right)^n,$$

where $p_i = \theta_i$, $i = 1, \ldots, k$, with $0 < \sum_{j=1}^k p_j < 1$ and $(a_1, \ldots, a_k) \in I \times I \times \cdots \times I$.

6. Conclusion

In this paper, the theory of parametric statistical model building without assuming the existence of true distributions has been discussed. The understanding of the models and the theory developed here are new and are expected to be strengthened in future fundamental statistical theory. As was mentioned in Section 1, the specification of a parametric statistical basic model and the statistical parameter estimation are fundamentally different concepts. The problems of model specification scrutinized in this paper are not those of model selection, because the latter are generally considered under the assumption of the existence of true distribution and its candidate distributions are given in advance.

As far as the author knows, we have not so far obtained systematic statistical theories for providing the candidate distributions. This fact annoys the intellectual users of statistical theory or its methods when they are faced with new phenomena and are urged to build suitable statistical basic models based on available observations. In overcoming such difficulties, the parametric statistical fundamental equations in this paper may be useful. As a future research subject, however, we have to proceed with the practical statistical model building. In order to do that, we need to develop the investigation of the observation error $\boldsymbol{\Delta}$ and the measurement scale $\boldsymbol{\Xi}$, which is required to connect directly to the practical or imaginable data. At that stage, some suitable statistical estimation theory will be helpful to adjust the parameters of the distribution obtained as a basic parametric model. In respect to this, as is shown in Theorem 4.2, we have a modified multivariate exponential family as the form of a parametric statistical basic model. So it is expected that the existing estimation theories, for example, sufficient estimation and the properties of exponential families (cf. Barndorff-Nielsen (1978), Brown (1986)), may strengthen our model specification theory. Currently, it may be helpful to make use of new aspects in the developments of statistical models (cf. Matsunawa (1995)).

In addition, the aim of this paper is to give a new introduction to multivariate statistical distribution. Compared with most textbooks on multivariate analysis, the model specification based on the fundamental statistical equations given here seems to be very helpful in introducing multivariate distributions systematically. An extension from univariate distributions to multivariate ones is systematically realized by resorting to some new implementations with vectorization and matrixization. Related nonparametric investigations to this paper can be also discussed (cf. Matsunawa (1994)).

Acknowledgements

The author would like to thank the referee and Editor of this journal for their constructive comments on an earlier version of this article.

References

Aitken, A. C. and Silverstone, H. (1942). On the estimation of statistical parameters, *Proc. Roy. Soc. Edinburgh, Sect. A*, **61**, 186–194.

Barndorff-Nielsen, O. (1978). *Information and Exponential Families in Statistical Theory*, Wiley, New York.

Bohr, N. (1928). The quantum postulate and the recent development of atomic theory, *Nature*, **121**, 580–590.

Brown, L. D. (1986). *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory*, IMS, Lectue Notes-Monograph Series, Vol. 9, Hayward, California.

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics, *Philos. Trans. Roy. Soc. London Ser. A*, **222**, 309–368.

Fisher, R. A. (1936). Uncertain inference, Proceedings of American Academy of Arts and Sciences, **71**, 245–258.

Heisenberg, W. (1927). Über den anschaulichen Inhalt der quantentheoretishen Kinematik und Mechanik, *Zeitschrift fur Physik*, **43**, 172–198.

Kapur, J. N. (1989). *Maximum-Entropy Models in Science and Engineering*, Wiley Easter, New Delhi.

Matsunawa, T. (1994). Origin of distributions —Nonparametric statistical uncertainty relation and a statistical fundamental equation—, *Proc. Inst. Statist. Math.*, **42**, 197–214 (in Japanese).

Matsunawa, T. (1995). Development of distributions —The Legendre transformation and canonical information criteria—, *Proc. Inst. Statist. Math.*, **43**, 293–311 (in Japanese).

Matsunawa, T. (1997). Basic parametric statistical model building without assuming existence of true distributions, *Proc. Inst. Statist. Math.*, **45**, 107–123 (in Japanese).

Matsunawa, T. and Zhao, Y. (1994). Specification of models for observations in multivariate case, Research Memo., No. 500, The Institute of Statistical Mathematics, Tokyo.

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*, Wiley, New York.

Pearson, K. (1895). Contribution to the mathematical theory of evolution II, Skew variation in homogeneous material, *Philos. Trans. Roy. Soc. London Ser. A*, **186**, 343–414.

Pearson, K. (1916). Contribution to the mathematical theory of evolution XIX, Second supplement to a memoir on Skew variation, *Philos. Trans. Roy. Soc. London Ser. A*, **216**, 429–457.