Scientiae Mathematicae Japonicae

ON A MISUSE OF SUFFICIENT STATISTICS IN THE EXPONENTIAL FAMILY

Etsuo Hamada

Received Febuary 24, 2017

Abstract. With respect to the sufficient statistics in the transformed exponential family based on a continuous probability distribution, we examine a misuse that the sufficient statistics ought to be distributed with a k-dimensional exponential family where k is the dimension of the sufficient statistics. Under the irreducibility of the sufficient statistics, we define two types of the transformed exponential family, i.e., regular and pseudo, so that the misuse is made explicit.

1 Introduction The exponential and curved exponential families cover a wide range of distributions ([4], [5], [14], [15]), and are widely used for generalized mixed linear models [7], and for methods in information geometry ([1], [2], [3], [6], [8], [13]).

Under the framework of information geometry, it is well assumed that the sufficient statistics are *linearly independent* in order to make a one-to-one correspondence between the parameter and the density [12] and that the dimension of parameters are equal to the dimension of the sufficient statistics which are linearly independent under the duality in the statistical manifold. [10] showed that linear independence of the score function is not a sufficient condition for a distribution to belong to the curved exponential family, showing that there exists a gap between the parameter space and the observations.

When we regard an original probability distribution as one of the exponential family, we should carefully examine the assumptions in the exponential family. In this article, we show that an appropriate transformation from an original probability distribution of random variable X to a natural exponential family is restricted by a structure corresponding to the expectation $\mu = E(X)$ with respect to the original distribution at most. Thus an extended structure corresponding up to the sufficient statistics with respect to the original distribution

2010 *Mathematics Subject Classification.* 62B10.
Key words and phrases. Exponential family, Sufficient statistics, Misuse, Linearly independent, Irreducible, Transformed regular/pseudo- exponential family.

implies that the transformed exponential family contradicts the assumption of the natural exponential family in the viewpoint of the probability measure of the sufficient statistics.

We define a transformed regular/pseudo- exponential family with respect to the transformation from a continuous probability distribution and examine whether the sufficient statistics in the transformed exponential family is distributed with the k-dimensional exponential family where k is the dimension of the sufficient statistics. A vague application to the sufficient statistics implies a misuse in the exponential family.

2 A misuse in the sufficient statistics Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and a random vector $\boldsymbol{X}$ on the probability space $(\Omega, \mathcal{F}, \mathcal{P})$ induces a probability space $(\mathbf{R}^m, \mathcal{B}^m, \mu)$ by $\mu(A) = \mathcal{P}(\boldsymbol{X}^{-1}(\boldsymbol{A})) = \mathcal{P}\{\boldsymbol{X} \in \boldsymbol{A}\}$ for any $\boldsymbol{A} \in \mathcal{B}^m$. Symbolically we may write it $\mu = \mathcal{P} \circ \boldsymbol{X}^{-1}$ and this μ is called the "probability distribution measure" of $\boldsymbol{X}$ and we denote the probability space $(\mathbf{R}^m, \mathcal{B}^m, \mu)$ as $(\mathcal{X}, \mathcal{A}, \mu)$ and the probability (density) function of $\boldsymbol{X}$ as $f(\boldsymbol{x}|\boldsymbol{\xi})$ where $\boldsymbol{\xi}$ is an s-dimensional parameter with respect to the probability distribution of $\boldsymbol{X}$, i.e.,

$$f(\boldsymbol{x}|\boldsymbol{\xi}), \quad \boldsymbol{x} \in \mathbf{R}^m, \ \boldsymbol{\xi} \in \mathbf{R}^s, \tag{1}$$

where $\boldsymbol{x} = (x_1, x_2, \ldots, x_m)^T$, $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_s)^T$, and the notation ^{T} means the transpose.

Now we consider a transformed expression of $f(\boldsymbol{x}|\boldsymbol{\xi})$ as regarding an exponential family and we denote it as $f(\boldsymbol{y}|\boldsymbol{\eta})$, i.e.,

$$f(\boldsymbol{y}|\boldsymbol{\eta}) \ = \ \exp\left[C(\boldsymbol{x}) + \langle \boldsymbol{\eta}(\boldsymbol{\xi}), \boldsymbol{y}(\boldsymbol{x}) \rangle - \phi(\boldsymbol{\xi})\right], \tag{2}$$

where the notation $\langle , \rangle$ means the inner product, $\boldsymbol{\eta}(\boldsymbol{\xi}) = (\eta^1(\boldsymbol{\xi}), \ldots, \eta^k(\boldsymbol{\xi}))^T$ is a transformed parameter, $\boldsymbol{y} = \boldsymbol{y}(\boldsymbol{x}) = (y_1(\boldsymbol{x}), \ldots, y_k(\boldsymbol{x}))^T$ is a transformed variable, $\phi(\boldsymbol{\xi})$ is the normalizing term, and $C(\boldsymbol{x})$ is a constant term. We call (2) a transformed exponential family for the original density (1) and we define the following three kind of types in (2):

$$\begin{cases} \text{(Case A)} \quad f(\boldsymbol{y}|\boldsymbol{\eta}) \ = \ \displaystyle\sum_{\boldsymbol{x}:\boldsymbol{y}(\boldsymbol{x})=\boldsymbol{y}} f(\boldsymbol{x}|\boldsymbol{\xi}), \\ \text{(Case B)} \quad f(\boldsymbol{y}|\boldsymbol{\eta})\, d\boldsymbol{y} \ = \ \begin{cases} \text{(B1)} \quad f(\boldsymbol{x}|\boldsymbol{\xi}) \left|\dfrac{\partial \boldsymbol{x}}{\partial \boldsymbol{y}}\right| d\boldsymbol{y} & \text{under } \left|\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{y}}\right| \neq 0, \\ \text{(B2)} \quad f(\boldsymbol{y}(\boldsymbol{x})|\boldsymbol{\eta})\, d\boldsymbol{x} & \text{under } \left|\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{y}}\right| = 0. \end{cases} \end{cases} \tag{3}$$

Definition 2.1 *We define the above three types (3) of the transformed exponential family as follows:*

- *(Case A) a transformed discrete exponential family,*
- *(Case B1) a transformed regular exponential family,*
- *(Case B2) a transformed pseudo-exponential family.*

Remark that Case A is for a discrete random vector $\boldsymbol{X}$ and it does not need one-to-one correspondence between $\boldsymbol{X}$ and $\boldsymbol{Y}$ and Case B1 requires a one-to-one correspondence between $\boldsymbol{X}$ and $\boldsymbol{Y}$ in order to the non-zero Jacobian $\left|\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{y}}\right| \neq 0$, so that both dimensions are equal, i.e., $k = m$ and the sufficient statistics $\boldsymbol{Y}$ is a k-dimensional random vector distributed with a k-dimensional probability distribution.

Note that, because of $k \neq m$ in Case B2, the sufficient statistics $\boldsymbol{Y}(\boldsymbol{X})$ is a k-dimensional random vector distributed with $\left|\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{y}}\right| = 0$, so that we should recognize that the sufficient statistics $\boldsymbol{Y}$ in the transformed pseudo-exponential family does not correspond to $\boldsymbol{X}$ one-on-one.

We consider a k-dimensional canonical exponential family as follows:

$$g(\boldsymbol{z}|\boldsymbol{\theta}) \;=\; \exp\left[C(\boldsymbol{z}) + \langle \boldsymbol{\theta},\, \boldsymbol{z}\rangle - \phi(\boldsymbol{\theta})\right], \quad \boldsymbol{z} \in \mathbf{R}^k,\; \boldsymbol{\theta} \in \mathbf{R}^k. \tag{4}$$

It is obvious that the sufficient statistics is a k-dimensional $\boldsymbol{z}$. Under the framework of information geometry for the density (4), it is well assumed that $k+1$ functions $\{1, z_1, \ldots, z_k\}$ are *linearly independent* in order to make a one-to-one correspondence between $\boldsymbol{\theta}$ and $g(\boldsymbol{z}|\boldsymbol{\theta})$ for arbitrary $\boldsymbol{\theta} \in \mathbf{R}^k$, where 1 is the identity mapping. For example, [12] defined the linear independence for the density (2) as follows:

Definition 2.2 *The functions $\{1, y_1(\boldsymbol{x}), \ldots, y_k(\boldsymbol{x})\}$ in (2) are said to be linearly independent if the following holds: $a_0 + \sum_{j=1}^{k} a_j y_j(\boldsymbol{x}) = 0$ for any $\boldsymbol{x}$ in an open set if and only if $a_0 = \cdots = a_k = 0$.*

Here we consider the following definition for Definition 2.2:

DEFINITION **2.3** *The functions* $\{y_1(\boldsymbol{x}), \ldots, y_k(\boldsymbol{x})\}$ *in (2) are said to be reducible if the following holds: for some* $y_i(\boldsymbol{x})$*, there exists constants* $d_i, c_j \ (j \neq i) \in \mathbf{R}$ *such that*

$$y_i(\boldsymbol{x}) \;=\; \sum_{j \neq i} c_j\, y_j(\boldsymbol{x}) + d_i\,,$$

where at least one of $\{c_j\}$ *is not zero. If the functions* $\{y_1(\boldsymbol{x}), \ldots, y_k(\boldsymbol{x})\}$ *are not reducible for arbitrary* $y_i(\boldsymbol{x})$*, then we call them irreducible.*

We have a relationship between Definition 2.2 and Definition 2.3 in the following lemma:

LEMMA **2.1** *If the functions* $\{y_1(\boldsymbol{x}), \ldots, y_k(\boldsymbol{x})\}$ *in (2) are irreducible for any* $\boldsymbol{x}$ *in an open set, then they satisfy the linear independence in Definition 2.2.*

Proof: If the functions $\{y_1(\boldsymbol{x}), \ldots, y_k(\boldsymbol{x})\}$ are irreducible, the condition $a_0+\sum_{j=1}^{k} a_j y_j(\boldsymbol{x}) = 0$ in Definition 2.2 can be regarded as follows:

$$\left\langle \mathbf{1},\ a_0\, \boldsymbol{e}_0 + \sum_{j=1}^{k} \left(a_j y_j(\boldsymbol{x})\right) \boldsymbol{e}_j \right\rangle \;=\; 0,$$

where the vector $\mathbf{1} = (1, 1, \ldots, 1)$ and $\boldsymbol{e}_j$ is the $(j+1)$-th unit vector $(j = 0, 1, \ldots, k)$. Suppose that there exists some $a_i y_i(\boldsymbol{x}) \neq 0$. Then $a_i y_i(\boldsymbol{x}) = -a_0 - \sum_{j \neq i} a_j y_j(\boldsymbol{x})$. If $a_i \neq 0$, then

$$y_i(\boldsymbol{x}) \;=\; -\frac{a_0}{a_i} - \sum_{j \neq i} \frac{a_j}{a_i}\, y_j(\boldsymbol{x}),$$

i.e., $y_i(\boldsymbol{x})$ is reducible and this is a contradiction, so that $a_i = 0$ and this also contradicts the assumption $a_i y_i(\boldsymbol{x}) \neq 0$. Thus we have $a_0 = a_1 y_1(\boldsymbol{x}) = \cdots = a_k y_k(\boldsymbol{x}) = 0$ and $\forall y_i(\boldsymbol{x}) \neq 0$ for any $\boldsymbol{x}$ in an open set, so that $a_0 = a_1 = \cdots = a_k = 0$. Therefore the functions $\{1, y_1(\boldsymbol{x}), \ldots, y_k(\boldsymbol{x})\}$ satisfy the linear independence in Definition 2.2. $\square$

Therefore, as a matter of principle, we suppose that $k+1$ functions $\{1, y_1(\boldsymbol{x}), \ldots, y_k(\boldsymbol{x})\}$ in (2) are irreducible. We show two typical examples as follows:

EXAMPLE **2.1** *In the multinomial distribution with* $k+1$ *cells, the probability function is*

$$f(\boldsymbol{x}|\boldsymbol{\xi}) \;=\; \binom{n}{x_1\, x_2\, \cdots\, x_{k+1}} \prod_{i=1}^{k+1} p_i^{x_i}$$

where $\boldsymbol{x} = (x_1, \ldots, x_k)$, $\boldsymbol{\xi} = (p_1, \ldots, p_k)$, $\sum_{i=1}^{k+1} p_i = 1 \quad (p_i \geq 0\ (i = 1, \ldots, k+1))$, *and where* $\forall x_i \geq 0$, $\sum_{i=1}^{k+1} x_i = n$, *and each integer* x_i $(i = 1, \ldots, k)$ *is the frequency in the* i*-th cell respectively. For* $f(\boldsymbol{x}|\boldsymbol{\xi})$, *the transformed exponential family is*

$$f(\boldsymbol{y}|\boldsymbol{\eta}) \;=\; \exp\left\{C(\boldsymbol{x}) + \sum_{i=1}^{k} \eta^i(\boldsymbol{\xi}) y_i(\boldsymbol{x}) - \phi(\boldsymbol{\xi})\right\}$$

with respect to $\boldsymbol{Y} = \boldsymbol{Y}(\boldsymbol{X})$, *where* $y_i(\boldsymbol{x}) = x_i$, $\eta^i(\boldsymbol{\xi}) = \log(\xi_i/(1 - \sum_{j=1}^{k} \xi_j)$, $(i = 1, \ldots, k)$,

$$\phi(\boldsymbol{\xi}) = -n \log\left(1 - \sum_{j=1}^{k} \xi_j\right), \quad \textit{and} \quad C(\boldsymbol{x}) = \log\binom{n}{x_1\, x_2\, \cdots\, x_{k+1}}.$$

Here the linear independence of $\{1, y_1(\boldsymbol{x}), \ldots, y_k(\boldsymbol{x})\}$ *holds.*

Example 2.2 *In the normal distribution* $N(\mu, \sigma^2)$ *whose density is*

$$f(x|\boldsymbol{\xi}) \;=\; \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} \quad (x \in \mathbf{R})$$

with $\boldsymbol{\xi} = (\mu, \sigma)$ *for* $-\infty < \mu < \infty$ *and* $0 < \sigma < \infty$, *the transformed exponential family is*

$$f(\boldsymbol{y}|\boldsymbol{\eta}) \;=\; \exp\left\{C(x) + \sum_{i=1}^{2} \eta^i(\xi)\, y_i(x) - \phi(\boldsymbol{\xi})\right\}$$

with respect to $\boldsymbol{Y} = \boldsymbol{Y}(X)$, *where* $\eta^1(\boldsymbol{\xi}) = \mu/\sigma^2$, $\eta^2(\boldsymbol{\xi}) = -1/(2\sigma^2)$, $y_1(x) = x$, $y_2(x) = x^2$,

$$\phi(\boldsymbol{\xi}) = \frac{\mu^2}{2\sigma^2} + \log(\sigma), \quad \textit{and} \quad C(x) = \log\frac{1}{\sqrt{2\pi}}.$$

Here the region $\{(y_1(x), y_2(x)) : x \in \mathbf{R}\}$ *is equivalent to a parabolic curve in the 2-dimensional space and the functions* $\{1, y_1(x), y_2(x)\}$ *are irreducible by Definition 2.3. This example is of (Case B2) in the relationship (3) unless the parameter* σ^2 *is supposed to be known.*

Since it is well known that the exponential family includes a lot of probability distributions, we are apt to confuse the sufficient statistics in the exponential family with the sufficient statistics distributed with the exponential family. The following theorem shows a solution to the above confusion.

THEOREM **2.1** *Let $\boldsymbol{X}$ be an m-dimensional random vector distributed with a probability density function (1) and, for a transformed exponential family (2) of $\boldsymbol{X}$, let the k-dimensional random vector $\boldsymbol{Y}(\boldsymbol{X}) = (\boldsymbol{Y}^{(1)}(\boldsymbol{X}), \boldsymbol{Y}^{(2)}(\boldsymbol{X}))$ which is the sufficient statistics in (2). Suppose that these $k+1$ functions $\{1, \boldsymbol{y}^{(1)}, \boldsymbol{y}^{(2)}\}$ are irreducible. There exists three cases as follows: (Case 1) $k < m$, (Case 2) $k = m$, (Case 3) $k > m$. In Case 1, since we have a loss of information based on $\boldsymbol{X}$, it contradicts that $\boldsymbol{Y}$ is the sufficient statistics. In Case 2, since $\boldsymbol{X}$ and $\boldsymbol{Y}$ have a one-to-one correspondence, this is a transformed regular exponential family. In Case 3, since $\boldsymbol{Y}^{(1)}(\boldsymbol{X})$ is regarded as corresponding to $\boldsymbol{X}$ under $k > m$ and the dimension of $\boldsymbol{Y}^{(2)}(\boldsymbol{X})$ is $k-m$, there exists a measurable and irreducible function u such that $\boldsymbol{y}^{(2)} = u\left(\boldsymbol{y}^{(1)}\right)$ and the conditional density of $\boldsymbol{y}^{(2)}$ given $\boldsymbol{y}^{(1)}$ is the indicator function, i.e., this is a transformed pseudo-exponential family.*

Proof: Both Case 1 and Case 2 are obvious, so we prove Case 3 only.

Since the joint density $h(\boldsymbol{y}^{(1)}, \boldsymbol{y}^{(2)})$ of $\boldsymbol{Y}$ is equivalent to the transformed exponential family (2), i.e.,

$$h(\boldsymbol{y}^{(1)}, \boldsymbol{y}^{(2)}) \;=\; f(\boldsymbol{y}|\boldsymbol{\eta}) \;=\; \exp\{C(\boldsymbol{x}) + \langle \boldsymbol{\eta}^{(1)}(\boldsymbol{\xi}),\, \boldsymbol{y}^{(1)}\rangle + \langle \boldsymbol{\eta}^{(2)}(\boldsymbol{\xi}),\, \boldsymbol{y}^{(2)}\rangle - \phi(\boldsymbol{\xi})\},$$

where $\boldsymbol{\eta}(\boldsymbol{\xi}) = (\boldsymbol{\eta}^{(1)}(\boldsymbol{\xi}),\, \boldsymbol{\eta}^{(2)}(\boldsymbol{\xi}))$ and the dimension of $\boldsymbol{\eta}^{(1)}(\boldsymbol{\xi})$ is m, we have the following representation:

$$f(\boldsymbol{y}|\boldsymbol{\eta}) \;=\; f(\boldsymbol{y}^{(2)} \,|\, \boldsymbol{y}^{(1)})\, f(\boldsymbol{y}^{(1)}),$$

where the marginal of $\boldsymbol{Y}^{(1)}$ and the conditional density of $\boldsymbol{Y}^{(2)}$ given $\boldsymbol{Y}^{(1)}$ are

$$f(\boldsymbol{y}^{(1)}) \;=\; \exp\{\langle \boldsymbol{\eta}^{(1)}(\boldsymbol{\xi}),\, \boldsymbol{y}^{(1)}\rangle - \phi(\boldsymbol{\xi})\} \int \exp\{C(\boldsymbol{x}) + \langle \boldsymbol{\eta}^{(2)}(\boldsymbol{\xi}),\, \boldsymbol{y}^{(2)}\rangle\} d\boldsymbol{y}^{(2)},$$

$$f(\boldsymbol{y}^{(2)} \,|\, \boldsymbol{y}^{(1)}) \;=\; \frac{\exp\{C(\boldsymbol{x}) + \langle \boldsymbol{\eta}^{(2)}(\boldsymbol{\xi}),\, \boldsymbol{y}^{(2)}\rangle\}}{\int \exp\{C(\boldsymbol{x}) + \langle \boldsymbol{\eta}^{(2)}(\boldsymbol{\xi}),\, \boldsymbol{y}^{(2)}\rangle\}\, d\boldsymbol{y}^{(2)}} \tag{5}$$

under the assumption that $0 < \int \exp\{C(\boldsymbol{x}) + \langle \boldsymbol{\eta}^{(2)}(\boldsymbol{\xi}),\, \boldsymbol{y}^{(2)}\rangle\} d\boldsymbol{y}^{(2)} < \infty$. Since these $k+1$ functions $\{1, \boldsymbol{y}^{(1)}, \boldsymbol{y}^{(2)}\}$ are irreducible, $\boldsymbol{Y}^{(1)} = \boldsymbol{Y}^{(1)}(\boldsymbol{X})$ corresponds $\boldsymbol{X}$ one-on-one, and $\boldsymbol{Y}^{(2)} = \boldsymbol{Y}^{(2)}(\boldsymbol{X})$ is a measurable function of $\boldsymbol{X}$, any element of $\boldsymbol{Y}^{(2)}$ is not of a linear combination of $\boldsymbol{Y}^{(1)}$, so that there exists a measurable function u such that $\boldsymbol{Y}^{(2)} = u(\boldsymbol{Y}^{(1)})$ and $\{1, \boldsymbol{y}^{(1)}, u(\boldsymbol{y}^{(1)})\}$ are irreducible.

For the joint density function $h(\boldsymbol{y}^{(1)}, \boldsymbol{y}^{(2)})$ with the relationship $\boldsymbol{y}^{(2)} = u(\boldsymbol{y}^{(1)})$, it holds that $h(\boldsymbol{y}^{(1)}, \boldsymbol{y}^{(2)}) = \delta_{u(y^{(1)})}(\boldsymbol{y}^{(2)})\, h(\boldsymbol{y}^{(1)})$, where $h(\boldsymbol{y}^{(1)})$ is the marginal function of $\boldsymbol{Y}^{(1)}$, the function $\delta_{u(y^{(1)})}(\boldsymbol{y}^{(2)})$ is the Dirac's delta function at the point $u(\boldsymbol{y}^{(1)})$, that is, the conditional density function of $\boldsymbol{Y}^{(2)}$ given $\boldsymbol{y}^{(1)}$ is the delta function $\delta_{u(y^{(1)})}(\boldsymbol{Y}^{(2)})$:

$$
(6) \qquad h(\boldsymbol{y}^{(2)} \,|\, \boldsymbol{y}^{(1)}) \;=\; \frac{h(\boldsymbol{y}^{(1)}, \boldsymbol{y}^{(2)})}{h(\boldsymbol{y}^{(1)})} \;=\; \frac{\delta_{u(y^{(1)})}(\boldsymbol{y}^{(2)})\, h(\boldsymbol{y}^{(1)})}{h(\boldsymbol{y}^{(1)})} \;=\; \delta_{u(y^{(1)})}(\boldsymbol{y}^{(2)}),
$$

so that, since the conditional probability function (5) is equivalent to the conditional (6), we have the following relationship:

$$
(7) \qquad \begin{aligned} \delta_{u(y^{(1)})}(\boldsymbol{y}^{(2)}) \;&=\; \frac{\exp\{C(\boldsymbol{x}) + \langle \boldsymbol{\eta}^{(2)}(\boldsymbol{\xi}),\, \boldsymbol{y}^{(2)} \rangle\}}{\int \exp\{C(\boldsymbol{x}) + \langle \boldsymbol{\eta}^{(2)}(\boldsymbol{\xi}),\, \boldsymbol{y}^{(2)} \rangle\} d\boldsymbol{y}^{(2)}} \\ &=\; \begin{cases} 1, & \text{if } \boldsymbol{Y}^{(2)} = u(\boldsymbol{y}^{(1)}), \\ 0, & \text{if } \boldsymbol{Y}^{(2)} \neq u(\boldsymbol{y}^{(1)}). \end{cases} \end{aligned}
$$

If $\boldsymbol{Y}^{(2)} = u(\boldsymbol{y}^{(1)})$, then the numerator is equivalent to the denominator in (7), i.e.,

$$
\exp\{C(\boldsymbol{x}) + \langle \boldsymbol{\eta}^{(2)}(\boldsymbol{\xi}),\, \boldsymbol{y}^{(2)} \rangle\} \;=\; \int \exp\{C(\boldsymbol{x}) + \langle \boldsymbol{\eta}^{(2)}(\boldsymbol{\xi}),\, \boldsymbol{y}^{(2)} \rangle\} d\boldsymbol{y}^{(2)},
$$

which implies that the probability distribution of $\boldsymbol{Y}^{(2)}$ should be one point distribution because the based random variable $\boldsymbol{X}$ is a continuous distribution. If $\boldsymbol{Y}^{(2)} \neq u(\boldsymbol{y}^{(1)})$, then the numerator in (7) should be zero, i.e.,

$$
\exp\{C(\boldsymbol{x}) + \langle \boldsymbol{\eta}^{(2)}(\boldsymbol{\xi}),\, \boldsymbol{y}^{(2)} \rangle\} \;=\; 0,
$$

and this is impossible, but we need not directly consider the conditional density of $\boldsymbol{Y}^{(2)}$ given $\boldsymbol{Y}^{(1)}$ in this situation and the transformed density is zero, i.e., $f(\boldsymbol{y}|\boldsymbol{\eta}) = 0$.

Therefore, for the sufficient statistics $\boldsymbol{Y} = (\boldsymbol{Y}^{(1)}, \boldsymbol{Y}^{(2)})$, the transformed density (2) is represented by

$$
(8) \qquad f(\boldsymbol{y}|\boldsymbol{\eta}) \;=\; \begin{cases} f(\boldsymbol{y}^{(1)}), & \text{if } \boldsymbol{Y}^{(2)} = u(\boldsymbol{y}^{(1)}), \\ 0, & \text{if } \boldsymbol{Y}^{(2)} \neq u(\boldsymbol{y}^{(1)}), \end{cases}
$$

where $f(\boldsymbol{y}^{(1)}) \;=\; \exp\{C(\boldsymbol{x}) + \langle \boldsymbol{\eta}^{(1)}(\boldsymbol{\xi}),\, \boldsymbol{y}^{(1)} \rangle + \langle \boldsymbol{\eta}^{(2)}(\boldsymbol{\xi}),\, u(\boldsymbol{y}^{(1)}) \rangle - \phi(\boldsymbol{\xi})\}$, so that we can regard the second element $\boldsymbol{Y}^{(2)}$ as either a random variable with one point distribution or a non-random (deterministic) variable given $u(\boldsymbol{y}^{(1)})$. In the equation (8), the left-hand side is the density of k-dimensional random variable and the right-hand side is that of m-dimensional random variable ($k > m$), which implies that the k-dimensional sufficient statistics $\boldsymbol{Y}$ is distributed with a transformed pseudo-exponential family in (3). $\square$

Although the conditional expectation and variance of $\boldsymbol{Y}^{(2)}$ given $\boldsymbol{Y}^{(1)}$ in Case 3 are

$$E[\boldsymbol{Y}^{(2)} \,|\, \boldsymbol{Y}^{(1)}] \;=\; u(\boldsymbol{Y}^{(1)}) \quad \text{and} \quad V[\boldsymbol{Y}^{(2)} \,|\, \boldsymbol{Y}^{(1)}] \;=\; \boldsymbol{0},$$

the expectation and variance of $\boldsymbol{Y}^{(2)}$ are

$$\begin{aligned} E[\boldsymbol{Y}^{(2)}] &= E\left[E[\boldsymbol{Y}^{(2)} \,|\, \boldsymbol{Y}^{(1)}]\right] \;=\; E[u(\boldsymbol{Y}^{(1)})], \\ V[\boldsymbol{Y}^{(2)}] &= V\left[E[\boldsymbol{Y}^{(2)} \,|\, \boldsymbol{Y}^{(1)}]\right] + E\left[V[\boldsymbol{Y}^{(2)} \,|\, \boldsymbol{Y}^{(1)}]\right] \;=\; V[u(\boldsymbol{Y}^{(1)})], \end{aligned}$$

and the covariance between $\boldsymbol{Y}^{(1)}$ and $\boldsymbol{Y}^{(2)}$ is $Cov[\boldsymbol{Y}^{(1)}, \boldsymbol{Y}^{(2)}] = Cov[\boldsymbol{Y}^{(1)}, u(\boldsymbol{Y}^{(1)})]$. On the other hand, based on the structure of exponential family, we have the following relationships with respect to the sufficient statistics $\boldsymbol{Y}$:

$$\begin{aligned} E[\boldsymbol{Y}] &= \frac{\partial \phi(\boldsymbol{\xi})}{\partial \boldsymbol{\eta}(\boldsymbol{\xi})} = \begin{pmatrix} E[\boldsymbol{Y}^{(1)}] \\ E[\boldsymbol{Y}^{(2)}] \end{pmatrix}, \\ V[\boldsymbol{Y}] &= \frac{\partial^2 \phi(\boldsymbol{\xi})}{\partial \boldsymbol{\eta}(\boldsymbol{\xi})\, \partial \boldsymbol{\eta}(\boldsymbol{\xi})^T} = \begin{pmatrix} V[\boldsymbol{Y}^{(1)}] & Cov[\boldsymbol{Y}^{(1)}, \boldsymbol{Y}^{(2)}] \\ Cov[\boldsymbol{Y}^{(1)}, \boldsymbol{Y}^{(2)}]^T & V[\boldsymbol{Y}^{(2)}] \end{pmatrix}, \end{aligned}$$

so that the transformed exponential family (2) with respect to k-dimensional $\boldsymbol{Y} = (\boldsymbol{Y}^{(1)}, \boldsymbol{Y}^{(2)})$ has the same properties with respect to the usual exponential family (4) with respect to $\boldsymbol{Z} = (Z_1, \ldots, Z_k)$ on the surface, but the transformed pseudo-exponential family in (3) is not like the k-dimensional exponential density (4) because the pseudo-density is reduced to the density (8) with respect to only $\boldsymbol{Y}^{(1)}$. Thus, even if the sufficient statistics $\boldsymbol{Y}(\boldsymbol{X})$ in the transformed exponential family (2) is irreducible, it might belong to the k-dimensional transformed pseudo-exponential family.

For the m-dimensional normal distribution $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the density is represented as follows:

$$f(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \;=\; \exp\left\{ \langle \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu},\, \boldsymbol{x} \rangle - \frac{1}{2} \langle \boldsymbol{\Sigma}^{-1}\boldsymbol{x},\, \boldsymbol{x} \rangle - \frac{\langle \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu},\, \boldsymbol{\mu} \rangle + \log(|\boldsymbol{\Sigma}|)}{2} \right\} \frac{1}{(2\pi)^{m/2}},$$

so that the sufficient statistics $\boldsymbol{Y}$ in the transformed exponential family under unknown parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ is

$$\boldsymbol{Y} \;=\; (X_1, \ldots, X_m,\, X_1^2, \ldots, X_m^2,\, X_1X_2, \ldots, X_{m-1}X_m)^T$$

whose dimension is $2m + (m^2 - m)/2$. Note that [9] studied the circular mechanism as a limitation to the transformed exponential family.

3 Conclusion In this article, we considered a misuse of the sufficient statistics in the transformed exponential family from a continuous probability distribution based on the linear independence of the sufficient statistics which are assumed in the information geometry. We defined new two terms, the transformed regular exponential family and the transformed pseudo-exponential family and we determined properties of the sufficient statistics under the irreducibility of the transformed exponential family. We recognized an importance of the Jacobian matrix with respect to the transformation of random variables.

We hope that it is decreasing to misuse that the sufficient statistics in a transformed pseudo-exponential family ought to be distributed with the k-dimensional regular exponential family where k is the dimension of the sufficient statistics.

References

[1] S.Amari (1985), *Differential-Geometrical Methods in Statistics*, Springer.

[2] S.Amari and H.Nagaoka (2000), *Methods of Information Geometry*, American Mathematical Society.

[3] O. Barndorff-Nielsen (1978), *Information and Exponential Families in Statistical Theory*, John Wiley & Sons

[4] D.D.Boos and L.A.Stefanski (2013), *Essential Statistical Inference Theory and Methods*, Springer.

[5] L.D.Brown (1986), *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory*, Institute of Mathematical Statistics.

[6] B.Efron (1975), Defining the Curvature of a Statistical Problem (with Applications to Second Order Efficiency), *Ann. Stat.*, **3**, 1189–1242.

[7] J.Jiang (2007), *Linear and Generalized Linear Mixed Models and Their Applications*, Springer.

[8] R.E.Kass and P.W.Voss (1997), *Geometrical Foundations of Asymptotic Inference*, John Wiley & Sons, Inc.

[9] E.Kumagai and N.Inagaki (1999), The circular mechanism in the curved exponential family, *Mathematicae Japonicae*, **49**(1), 27–38.

[10] E.Kumagai (2015), A comment on a regularity condition in a curved exponential family, *Scientiae Mathematicae Japonicae*, **78**(1), 97–101.

[11] M.K.Murray and J.W.Rice (1993), *Differential Geometry and Statistics*, Chapman & Hall.

[12] T.Ohnishi and T.Yanagimoto (2010), Duality induced from conjugacy in the curved exponential family, *Journal of Japan Statistical Society*, **40**(1), 23–43.

[13] G.Pistone and M.P.Rogantin (1999), The exponential statistical manifold: mean parameters, orthogonality and space transformations, *Bernoulli*, **5**(4), 721–760.

[14] M.J.Schervish (1995), *Theory of Statistics*, Springer.

[15] J.Shao (2003), *Mathematical Statistics 2nd ed.*, Springer.

Communicated by *Hisao Nagao*

Division of Mathematical Science,
Graduate School of Engineering Science,
Osaka University,
1-3 Machikaneyama, Toyonaka, Osaka 560-8531, Japan.