Metrika, Volume 33, 1986, page 299–305

# Non-Exponential Families of Distributions

By K. C. Klauer[1]

*Summary*: The problem discussed is whether a given set $\mathcal{P}$ of mutually equivalent probability distributions constitutes an exponential family or not. As an alternative to the well known characterization by linear independence an analytic method is pointed out which allows one to demonstrate in a relatively easy way that many familiar classes of distributions (e.g. Cauchy-, $t_n$-, logistic-) are not exponential families.

*Key words and Phrases*: Exponential families, exponential representations, minimal statistics.

## 1 Introduction and Terminology

For the well known examples of exponential families it is fairly easy to find an exponential representation. The complementary problem, i.e. to demonstrate that a given class $\mathcal{P}$ of distributions does not constitute an exponential family, requires quite different methods.

Apart from a characterization by linear independence, characterizations of exponential families are either confined to one-parameter exponential families (e.g. Pfanzagl 1968) or yield conditions that cannot be falsified more easily than the conditions implied by the original definition of an exponential family (e.g. Barndorff-Nielsen and Pedersen 1968).

In Sect. 2, the well known characterization by linear independence (cf. Barndorff-Nielsen 1978, p. 112) is briefly reviewed and applied in two cases to show that a family of distributions in non-exponential. In Sect. 3, we shall provide a new method which on the basis of a separation property allows one to identify many familiar classes of distributions as non-exponential. Contrasting Sect. 2 and the examples given in Sect. 3, it is seen that the new criterion is often easier to apply.

In the following we will essentially use the terminology of Barndorff-Nielsen (1978). Let $(\mathcal{X}, \mathcal{B}, \mathcal{P})$ be a statistical field, where $\mathcal{P}$ is a family of probability measures on the $\sigma$-algebra $\mathcal{B}$ on $\mathcal{X}$. $\mathcal{P}$ is said to be *parametrizable*, if there exist a subset $\Omega$ of a Euclidean space $\mathbb{R}^m$ and a bijective mapping from $\Omega$ onto $\mathcal{P}$. In this case,

[1] K. C. Klauer, Institut für Statistik und Wirtschaftsmathematik, RWTH Aachen, Wüllnerstraße 3, 5100 Aachen, West-Germany.

$(P_\omega)_{\omega \in \Omega}$ will be called a parametrization of $\mathcal{P}$. $\mathcal{P}$ is said to be ***relevant***, if it is parametrizable and if its elements are mutually equivalent.

*Definition 1:* Let $(\mathcal{X}, \mathcal{B}, \mathcal{P})$ be a statistical field. $\mathcal{P}$ is an exponential family if

(i) $\mathcal{P}$ is parametrizable and – $(P_\omega)_{\omega \in \Omega}$ being a parametrization of $\mathcal{P}$ –

(ii) there exist $k \in \mathbb{N}$, mappings $a : \Omega \to \mathbb{R}$, $\alpha : \Omega \to \mathbb{R}^k$ $T : (\mathcal{X}, \mathcal{B}) \to (\mathbb{R}^k, \mathcal{B}^k)$ and a $Q \in \mathcal{P}$ such that

$$\frac{dP_\omega}{dQ} = a(\omega) \exp(\alpha(\omega) \cdot T) \qquad \mathcal{P} - \text{a.e.} \tag{1}$$

*Remark 1:* The expression (1) is called an *exponential representation* of (the densities of) $\mathcal{P}$. The smallest integer $k$, for which there exists an exponential representation (1) is the *order* of $\mathcal{P}$ denoted by ord $\mathcal{P}$. The statistic $T$ is called a canonical statistic and, in case $k$ = ord $\mathcal{P}$, the representation (1) as well as $T$ are said to be *minimal*.

## 2 A Characterization by Linear Independence

Let $(\mathcal{X}, \mathcal{B}, \mathcal{P})$ be a relevant statistical field, $(P_\omega)_{\omega \in \Omega}$ be a parametrization of $\mathcal{P}$ and $Q \in \mathcal{P}$. For each $\omega \in \Omega$ we choose a version $p(\cdot; \omega)$ of the $Q$-densities of $P_\omega$ that is positive for all $x \in \mathcal{X}$. For a real-valued statistic $f$ we denote the $\mathcal{P}$-equivalence class containing $f$ by $[f]$.

*Proposition 1* (cf. Barndorff-Nielsen 1978, p 112): $\mathcal{P}$ is an exponential family of order $k$ if and only if span $A$, the linear hull of $A := \{[\ln p(\cdot; \omega)] : \omega \in \Omega\} \cup \{[1_\mathcal{X}]\}$ has dimension $k + 1$.

Proposition 1 immediately yields

*Corollary 1:* If $\Omega$ or $\mathcal{X}$ is a finite set, then $\mathcal{P}$ is an exponential family fulfilling

$$\text{ord } \mathcal{P} < |\Omega| + 1 \quad \text{or} \quad \text{ord } \mathcal{P} < |\mathcal{X}|, \quad \text{respectively.}$$

*Remark 2:* For the proof of Proposition 1 it is useful to note that – using a basis of span $A$ – a minimal canonical statistic $T$ can be chosen as

$$T(x) = (\ln p(x; \omega_1), \ldots, \ln p(x; \omega_k))^t.$$

Thus, when $\mathcal{X}$ is a topological space and $p(\cdot; \omega)$ is continuous for each $\omega \in \Omega$, then a *continuous* minimal canonical statistic exists.

In order to prove that $\mathcal{P}$ is not an exponential family, one has to demonstrate the existence of linearly idependent systems of arbitrary size in the Set $A$ as defined in Proposition 1. In doing so, the handling of $\mathcal{P}$-equivalence classes tends to be rather cumbersome. In many situations this can be avoided as follows: Let $\mathcal{X}$ be a topological space, $\mathcal{P}$ be the Borel $\sigma$-algebra and $\mathcal{P}$ be relevant; then one can assume that $\mathcal{X}$ is the common support of all $P \in \mathcal{P}$. If the functions $p(\cdot\,;\omega)$ are continuous w.r.t. the topology on $\mathcal{X}$, then they are uniquely determined, i.e. we can replace $[\ln p(\cdot\,;\omega)]$ by $\ln p(\cdot\,;\omega)$. Furthermore, using Remark 2 and interchanging the roles of $x$ and $\omega$, we get similarly to Proposition 1

*Corollary 2:* Let $(\mathcal{X}, \mathcal{B}, \mathcal{P})$ be a statistical field, where $(\mathcal{X}, \mathcal{B})$ is a Borel space and $\mathcal{X}$ is the common support of all $P \in \mathcal{P}$, $p(\cdot\,;\omega)$ being continuous w.r.t. the topology on $\mathcal{X}$. Then $\mathcal{P}$ is an exponential family if and only if span $A'$ where $A' = \{\ln p(x;\cdot) : x \in \mathcal{X}\}$, has finite dimension.

We now give two examples of how the preceding results can be used to show that certain classes of distributions are non-exponential. For this purpose the following remark is useful.

*Remark 3:* If $p(\cdot\,;\omega)$ is of the form $p(\cdot\,;\omega) = a(\cdot\,;\omega)b(\cdot\,;\omega)$ where $a(\cdot\,;\omega)$ and $b(\cdot\,;\omega)$ are continuous functions and $a(\cdot\,;\omega)$ already is of exponential form (i.e. the functions $\ln a(\cdot\,;\omega)$ span a vector space of finite dimension) then it is sufficient to find linearly independent systems of arbitrary size in the vector space spanned by the functions $\ln b(\cdot\,;\omega)$. Analogous results hold when in applying Corollary 2 the roles of $x$ and $\omega$ are interchanged.

*Example 1:* For parameters $(c_0, c_1) \in (0, \infty) \times (0, 1/2)$ the following functions are densities of probability measures with respect to Lebesgue measure

$$f(x;(c_0,c_1)) = K(c_0,c_1)(c_0 + c_1 x^2)^{-\frac{1}{2c_1}},$$

where $K(c_0,c_1)$ is a suitably chosen constant. Subfamilies of this family, i.e. those with $(c_0, c_1) \in \Omega \subset (0, \infty) \times (0, 1/2)$, are called Pearson – type VII – distributions (Johnson and Kotz 1970). Such subfamilies do not constitute exponential families if the set $A := \{c_0/c_1 : (c_0, c_1) \in \Omega\}$ is not finite.

For $n \in \mathbb{N}$, the functions $x \to \ln(a_i + x^2)$, $i = 1, \ldots, n$, constitute a linearly independent system where $(a_n)_{n \in \mathbb{N}}$ is a sequence of pairwise different elements of $A$. This can be seen by, for example, proving the linear independence of the derivatives of these functions. According to Proposition 1 and Remark 3, this proves our claim. An important member of this class is the family of Student's $t_\nu$-distributions with parameter $\nu \in \mathbb{N}$ as can be seen by representing the Lebesgue-densities in the form

$$f(x;\nu) = \left(\frac{\nu+1}{\nu}\right)^{-\frac{\nu+1}{2}} \frac{1}{\sqrt{\nu}\, B\left(\frac{1}{2}, \frac{\nu}{2}\right)} \left(\frac{\nu}{\nu+1} + \frac{x^2}{\nu+1}\right)^{-\frac{\nu+1}{2}}.$$

Letting $c_0 = \nu/(\nu+1)$, $c_1 = 1/(\nu+1)$ and observing that $A = \mathbb{N}$, it is seen that the family of Student's $t_\nu$-distributions is not an exponential family.

*Example 2:* The class of hyper-Poisson distributions given by

$$f(x;(\theta,\lambda)) = a(\theta,\lambda)\,\frac{\theta^x}{\lambda(\lambda+1)\ldots(\lambda+x-1)},\quad x\in\mathbb{N},(\theta,\lambda)\in(0,\infty)^2$$

(cf. Johnson and Kotz 1963) is not exponential: According to Corollary 2 and Remark 3 it is sufficient to show that the functions

$$g_x(\lambda) = \sum_{i=0}^{x-1} \ln(\lambda+i),\quad x\in\mathbb{N},$$

are linearly independent. This can be seen by observing that the functions $g_{x+1}-g_x = \ln(\cdot+x), x\in\mathbb{N}$, constitute a linearly independent system as can be shown by looking at their derivatives.

## 3 A Separation Criterion

In this section $\mathcal{X}$ is a Borel subset of an $\mathbb{R}^m$ and $\mathcal{B}$ is the Borel $\sigma$-algebra on $\mathcal{X}$. Furthermore, $\mathcal{P}$ is relevant and $\mathcal{X}$ is assumed to be the common support of all $P\in\mathcal{P}$. Let Let $\mu$ be a $\sigma$-finite measure equivalent to $\mathcal{P}$ and suppose there exist positive and continuous versions $f(\cdot;\omega)$ of the $\mu$-densities of $P_\omega$, $\omega\in\Omega$.

*Definition 2* (cf. Barndorff-Nielsen 1978, p. 12): Let $D$ be a subset of $\mathcal{X}$. $\mathcal{P}$ is said to *distinguish on D* if for all elements $x$ and $y$ of $D$ the condition

"there are positive constants $c$ and $\hat{c}$ such that (3)
for all $\omega\in\Omega$ $cf(x;\omega) = \hat{c}f(y;\omega)$"

implies that $x = y$.

Note that $f(\cdot;\omega)$ is uniquely determined by the assumptions listed at the beginning of this section.

The idea of the criterion to be derived resides in the fact that, under certain regularity conditions, the partition of $\mathcal{X}$ induced by (3) (note that (3) defines an equivalence relation on $\mathcal{X}$) is the same as that induced by a minimal sufficient statistic $T$, i.e. that $T(x) = T(y)$ if and only if (3) holds (cf. Barndorff-Nielsen 1978, Corollary 4.3). Thus, if $T:\mathcal{X}\to\mathbb{R}^k$ is a continuous minimal sufficient statistic and if $\mathcal{P}$ distinguishes on an open set $D\subset\mathbb{R}^m$, then $k$ must not be smaller than $m$ (see Lemma 1). Now assume that $\mathcal{P}$ is an exponential family and that $T:\mathcal{X}\to\mathbb{R}^k$ is a continuous minimal canonical statistic. If $\mathcal{P}$ distinguishes on an open set $D\subset\mathbb{R}^m$, then our argument suggests that $k\geqslant m$, since, under certain regularity conditions, $T$ is a minimal sufficient statistic for $\mathcal{P}$. Recalling that for minimal canonical statistics $k = \operatorname{ord}\mathcal{P}$, we have $\operatorname{ord}\mathcal{P}\geqslant m$. If the same assumptions hold for all families $\mathcal{P}^n := \{\bigotimes_{i=1}^n P : P\in\mathcal{P}\}$, we have ord $\mathcal{P}^n \geqslant n \cdot m$ for all $n \in \mathbb{N}$. Since ord $\mathcal{P}$ = ord $\mathcal{P}^n$ (cf. Barndorff-Nielsen 1978, p. 127), the assumption that $\mathcal{P}$ is an exponential family yields a contradiction. In the sequel, we will show that these heuristics give rise to a valid criterion under the assumptions listed at the beginning of this section. We will make use of a result from analytic topology.

*Lemma 1:* Let $m$, $n \in \mathbb{N}$, $n < m$, and $A \subset \mathbb{R}^m$, $A$ having non-empty interior. If $f: A \to \mathbb{R}^n$ is continuous, then $f$ is not injective.

*Proof:* The assertion is an immediate consequence of the so-called "antipodal theorem" which states (cf. e.g. Eisenack and Fenske 1978, Corollary 5.2.6): If $f$ is a continuous function defined on the boundary $\partial S$ of a symmetric bounded zero-neighbourhood $S$ in $\mathbb{R}^m$ and maps into a proper linear subspace of $\mathbb{R}^m$, then there exists a $p \in \partial S$ such that $f(p) = f(-p)$.

*Theorem 1:* Let the conditions stated at the beginning of this section be satisfied (in particular $\mathcal{X} \subset \mathbb{R}^m$) and let $\mathcal{P}$ be an exponential family. If there exists a subset $D$ of $\mathcal{X}$ the interior of which is not empty and such that $\mathcal{P}$ distinguishes on $D$, then ord $\mathcal{P} \geqslant m$.

*Proof:* Let

$$f(x;\omega) = h(x)a(\omega) \exp(\alpha(\omega) \cdot T) \quad \mu - \text{a.e.} \tag{4}$$

be a minimal representation of $\mathcal{P}$. According to Remark 2 we can assume $T$ to be continuous, hence $h$ can be assumed to be continuous and positive. This shows that (4) holds everywhere. If $T(x) = T(y)$, we have, setting $c = 1/h(x)$ and $\hat{c} = 1/h(y)$,

$$cf(x;\omega) = \hat{c}f(y;\omega) \quad \text{for all } \omega \in \Omega.$$

Since $\mathcal{P}$ distinguishes on $D$, the restriction of $T$ on $D$ must be injective. By Lemma 1 the assertion follows.

The criterion implied by Theorem 1 is often easier to use than that outlined in Sect. 2. We give two examples.

*Example 3:* The family of logistic distributions (Johnson and Kotz 1970) with parameter $(\alpha, \beta) \in \mathbb{R} \times (0, \infty)$ is given by the Lebesgue-densities

$$f(x;(\alpha,\beta)) = \frac{1}{\beta} \exp(-(x-\alpha)/\beta)/(\exp(-(x-\alpha)/\beta) + 1)^2.$$

The condition

$$\text{“}c \prod_{j=1}^{n} f(x_j;(\alpha,\beta)) = \hat{c} \prod_{j=1}^{n} f(y_j;(\alpha,\beta)) \quad \text{for all } (\alpha,\beta) \in \mathbb{R} \times (0,\infty)\text{”}$$

implies

$$\text{“}c \exp\left(-\sum_{j=1}^{n}(x_j-\alpha)\right)\prod_{j=1}^{n}(1+e^{-(y_j-\alpha)})^2$$

$$=\hat{c}\exp\left(-\sum_{j=1}^{n}(y_j-\alpha)\right)\prod_{j=1}^{n}(1+e^{-(x_j-\alpha)})^2 \quad \text{for all } \alpha\in\mathbb{R}\text{”}. \tag{5}$$

This equality has to hold even for complex $\alpha$. Thus, the zeros of the left-hand side and right-hand side of (5) have to coincide so that $\mathcal{P}^n$ distinguishes on the open set

$$D^{(n)} := \{(x_1,\ldots,x_n)^t : x_1 < x_2 < \ldots < x_n\}.$$

Thus, the family of logistic distributions is not an exponential family.

*Example 4:* The family of Cauchy distributions is not an exponential family. The Lebesgue densities of this family are given by

$$f(x;(\theta,\lambda)) = \frac{1}{\pi\lambda}\,\frac{1}{1+((x-\theta)/\lambda)^2}, \quad (\theta,\lambda)\in\mathbb{R}\times(0,\infty).$$

The condition

“there exist positive constants $c$ and $\hat{c}$ such that

$$c\prod_{j=1}^{n}f(x_j;(\theta,\lambda)) = \hat{c}\prod_{j=1}^{n}f(y_j;(\theta,\lambda)) \quad \text{for all } (\theta,\lambda)\in\mathbb{R}\times(0,\infty)\text{”}$$

implies

$$c\prod_{j=1}^{n}(1+(y_j-\theta)^2) = \hat{c}\prod_{j=1}^{n}(1+(x_j-\theta)^2) \quad \text{for all } \theta\in\mathbb{R}. \tag{6}$$

This equality has to hold even for complex $\theta$. Thus, the zeros of the left-hand side and right-hand side of (6) have to coincide so that $\mathcal{P}^n$ distinguishes on

$$D^{(n)} := \{(x_1,\ldots,x_n) : x_1 < x_2 < \ldots < x_n\}.$$

*Remark 4:* In a similar way, it can be seen that the class of Laplace-distributions (cf. Johnson and Kotz 1970) is not exponential.

*Acknowledgement*: I wish to thank Prof. O. Krafft for drawing my attention to the problem discussed in this paper.

## References

Barndorff-Nielsen O, Pedersen K (1968) Sufficient data reductions and exponential families. Math Scand 22:197–202

Barndorff-Nielsen O (1978) Information and exponential families. Wiley, New York

Eisenack G, Fenske Ch (1978) Fixpunkttheorie. Bibl Institut, Mannheim

Johnson N, Kotz S (1963) Discrete distributions. Wiley, New York

Johnson N, Kotz S (1970) Continuous univariate distributions. Wiley, New York

Pfanzagl J (1968) A characterization of the one parameter exponential family by existence of uniformly most powerful tests. Sankhyā A 30:147–156

Received September 10, 1984
(Revised version October 16, 1984)