Minimal sufficient statistics in location-scale parameter models

Lutz Mattner
Universität Hamburg

May 7, 2000

Abstract

Let f be a probability density on the real line, let n be any positive integer, and assume the condition (R) that $\log f$ is locally integrable with respect to Lebesgue measure. Then either $\log f$ is almost everywhere equal to a polynomial of degree less than n, or the order statistic of n independent and identically distributed observations from the location-scale parameter model generated by f is minimal sufficient.

It follows, subject to (R) and $n \geq 3$, that a complete sufficient statistic exists in the normal case only. Also, for f with (R) infinitely divisible but not normal, the order statistic is always minimal sufficient for the corresponding location-scale parameter model.

The proof of the main result uses a theorem on the harmonic analysis of translation and dilation invariant function spaces, attributable to K.O. Leland (1968) and L. Schwartz (1947).

1 Introduction and main results

1.1 Aim. Perhaps the most natural first step in the analysis of a statistical model consists in determining a minimal sufficient σ-algebra. Unfortunately, this is not always easy, and systematic results are rare. The aim of the present paper is to treat the case of independent and identically distributed observations from a location-scale parameter model on the real line. Here, subject only to a regularity condition (R) discussed in 1.5 below, a complete analysis is possible. It turns out that the order statistic usually is minimal sufficient. Previously, this was known in special cases only.

1.2 Guide. The main result of the present paper is the implication (A) $\Rightarrow$ (C) in Theorem 1.6. The equivalence (B) $\Leftrightarrow$ (C) in Theorem 1.6, due to Dynkin (1951) and Ferguson (1962), is here merely stated to round off the picture. Corollary 1.8 then shows that several statistically desirable properties of a location-scale parameter model are equivalent to the normality of the generating density. In particular, subject to (R), Corollary 1.8 solves a problem of Ferguson (1962) and generalizes a theorem of Kelker

[0] *1991 Mathematics Subject Classification:* 62B05, 62E10, 60E07, 42A75.

[0] *Keywords and phrases:* Characterization, complete sufficient statistics, equivariance, exponential family, independence, infinitely divisible distribution, mean periodic functions, normal distribution, order statistics, transformation model.

& Matthes (1970). Corollary 1.9 then yields a new probabilistic characterization of the normal distribution via independence. More detailed remarks on related work are given in Subsection 1.12.

The proofs of Theorem 1.6 and Corollaries 1.8 and 1.9 are deferred to Section 3. Section 2 collects auxiliary results, the crucial one being the theorem in 2.1, attributable to Leland (1968) and Schwartz (1947). The lemma in 2.3 might be of independent interest, although it is a simple consequence of basic differential calculus.

1.3 Sufficiency, minimal sufficiency, and notation. We assume as known the definitions, notation, and basic facts concerning sufficiency and minimal sufficiency, here needed in the dominated case only, as described in Section 1.5 of Torgersen (1991). Let us just indicate the definition of minimal sufficiency of a statistic $T : (\mathcal{X}, \mathcal{A}) \to (\mathcal{Y}, \mathcal{B})$ for a statistical model $\mathcal{P} = (P_\vartheta : \vartheta \in \Theta)$ on the measurable space $(\mathcal{X}, \mathcal{A})$, where $(\mathcal{Y}, \mathcal{B})$ is any measurable space: It means that the σ-algebra $\sigma(T) := T^{-1}(\mathcal{B})$ generated by T is minimal sufficient, and this means that $\sigma(T)$ is sufficient and that for every σ-algebra $\mathcal{C} \subset \mathcal{A}$ which is sufficient for $\mathcal{P}$, we have $\sigma(T) \subset \mathcal{C}$ $[\mathcal{P}]$. The latter notation means that every element of the σ-algebra on the left hand side is up to $\mathcal{P}$-nullsets equal to an element of the right hand side.

We let $\mathcal{B}(\mathbb{R}^n)_{\mathrm{sym}}$ denote the σ-algebra generated by the order statistic on $\mathbb{R}^n$. Equivalently, $\mathcal{B}(\mathbb{R}^n)_{\mathrm{sym}}$ is the σ-algebra of all permutation invariant Borel sets on $\mathbb{R}^n$.

1.4 The model. Let f be a probability density with respect to Lebesgue measure $\boldsymbol{\lambda}$ on the Borel σ-algebra $\mathcal{B}(\mathbb{R})$ on the real line. We consider the location-scale parameter model for n independent and identically distributed observations, based on f. This is the family

$$\mathcal{P}^n \quad = \quad \left((\mathbb{R}^n \ni x \mapsto (\prod_{i=1}^n \frac{1}{b} f(\frac{x_i - a}{b})))\boldsymbol{\lambda}^n \ : \ a \in \mathbb{R},\, b \in]0, \infty[\right)$$

of probability measures on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, where $\boldsymbol{\lambda}^n$ denotes n-dimensional Lebesgue measure. For $n = 1$, we simply write $\mathcal{P}$ in place of $\mathcal{P}^1$. We let $X = (X_1, \ldots, X_n)$ denote the identity function on $\mathbb{R}^n$.

1.5 The regularity condition. Our results will be subject to the regularity assumption

(R) $\qquad \log f$ is locally integrable with respect to $\boldsymbol{\lambda}$.

This implies in particular the condition

(R_0) $\qquad f > 0$ $\boldsymbol{\lambda}$-almost everywhere,

and we conjecture that Theorem 1.6 and Corollaries 1.8 and 1.9 remain valid with the somewhat more natural condition (R_0) in place of (R). Without assuming (R) or (R_0), however, 1.6, 1.8 and 1.9 would be false, as the well-known counterexample in 1.11 shows. Concerning applications to specific densities of statistical or probabilistic interest, the difference between (R) and (R_0) appears to be slight.

1.6 Theorem. *Let f be a probability density satisfying* (R), *and let $\mathcal{P}$ and $\mathcal{P}^n$ be defined as in 1.4. Then, for every $n \in \mathbb{N}$, the following three statements are equivalent.*

(A) *The order statistic is not minimal sufficient for $\mathcal{P}^n$.*

(B) *$\mathcal{P}$ is an exponential family of dimension less than n.*

(C) *$f = \exp p$ $\boldsymbol{\lambda}$-a.e. for some polynomial p of degree less than n.*

If (A), (B) *and* (C) *are true, then the dimension from* (B) *and the degree from* (C) *are the same, say k, k is even and ≥ 2, and the statistic $(\sum X_i, \ldots, \sum X_i^k)$ is minimal sufficient.*

1.7 Invariance and equivariance. To prepare for the statements of Corollaries 1.8 and 1.9, let us briefly recall some notions connected with transformation groups, specialized to the present context. For more on transformation groups and their use in statistics, we refer to Section 1.9 of Pfanzagl (1994) and to Section 2.1 of Wijsman (1990) as general introductions suitable for our present needs, and to Ramamoorthi (1990), Sapozhnikov (1998), Helland (1998), and to the references given there, for further developments.

The transformation group referred to in what follows is the group of affine transformations on coordinates, $G := \{g_{a,b} : a \in \mathbb{R}, b \in]0, \infty[\}$, where the transformations $g_{a,b}$ of $\mathbb{R}^n$ are defined by

$$g_{a,b}x \;:=\; (ax_1 + b, \ldots, ax_n + b) \qquad (a \in \mathbb{R}, b \in]0, \infty[, x \in \mathbb{R}^n).$$

We let $\mathcal{D}$ denote the σ-algebra of the G-invariant Borel sets. It is well known that $\mathcal{D}$ can be equivalently defined as

$$\mathcal{D} \;:=\; \sigma(\frac{X_1 - \overline{X}_n}{S_n}, \ldots, \frac{X_n - \overline{X}_n}{S_n}), \tag{1}$$

where $\overline{X}_n := \frac{1}{n}\sum X_i$, $S_n := (\frac{1}{n}\sum(X_i - \overline{X}_n)^2)^{1/2}$ and $0/0 := 0$, but it is important to keep in mind that here the choice of $\overline{X}_n$ and S_n is quite arbitrary: For example, $\overline{X}_n$ in (1) could be replaced by the sample median, leading to the same σ-algebra $\mathcal{D}$ generated by an essentially different statistic.

Let now T be a function on $\mathbb{R}^n$, with arbitrary range. T is called *equivariant* if we have $T(gx) = T(gy)$ whenever $g \in G$ and $x, y \in \mathbb{R}^n$ with $T(x) = T(y)$. For example, for any $k \in \mathbb{N}$, the statistic $(\sum X_i, \ldots, \sum X_i^k)$ of Theorem 1.6 is easily seen to be equivariant. [This is no accident, see Proposition 1.9.11 of Pfanzagl (1994).]

1.8 Corollary. *Let f be a probability density satisfying* (R), *and let $\mathcal{P}^n$ be defined as in 1.4. Then, for every $n \geq 3$, the following five statements are equivalent.*

(A) *There exists a σ-algebra $\mathcal{C} \subset \mathcal{B}(\mathbb{R}^n)$ which is complete and sufficient for $\mathcal{P}^n$.*

(B) *There exists a σ-algebra $\mathcal{C} \subset \mathcal{B}(\mathbb{R}^n)$ which is boundedly complete and sufficient for $\mathcal{P}^n$.*

(C) *There exists a σ-algebra $\mathcal{C} \subset \mathcal{B}(\mathbb{R}^n)$ which is sufficient for $\mathcal{P}^n$ and independent, under $\mathcal{P}^n$, of the σ-algebra $\mathcal{D}$ from* (1).

(D) *There exists, for some measurable space $(\mathcal{T}, \mathcal{A})$, a sufficient statistic $T : (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)) \to (\mathcal{T}, \mathcal{A})$ which is equivariant and satisfies, for some $\boldsymbol{\lambda}^n$-nullset $N \in \mathcal{D}$, the implication*

$$x \in \mathbb{R}^n \setminus N \;\Rightarrow\; \{T(bx_1 + a, \ldots, bx_n + a) : a \in \mathbb{R}, b \in]0, \infty[\} \;=\; T(\mathbb{R}^n \setminus N).$$

(E) *f is a normal density.*

1.9 Corollary. *Let P be a probability measure on $\mathbb{R}$ having a density f satisfying* (R). *Assume that for some $n \geq 3$ and for some countably generated and Hausdorff measurable space $(\mathcal{T}, \mathcal{A})$, there exists a statistic $T : (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)) \to (\mathcal{T}, \mathcal{A})$ with the properties:*

i) T is equivariant,

ii) for some $\boldsymbol{\lambda}^n$-nullset N, we have the implication

$$x \in \mathbb{R}^n \setminus N \quad \Longrightarrow \quad \text{the function } (a,b) \mapsto T(bx_1 + a, \ldots, bx_n + a) \text{ is injective,}$$

iii) under the product measure $P^{\otimes n}$, T and $\mathcal{D}$ are independent.

Then P is normal and $\sigma(T) = \sigma(\overline{X}_n, S_n)\,[\boldsymbol{\lambda}^n]$.

Here "$\mathcal{A}$ Hausdorff" means "If $s, t \in \mathcal{T}$ are different, then $s \in A$ and $t \notin A$ for some $A \in \mathcal{A}$". In the presence of the assumption "$\mathcal{A}$ countably generated", this is equivalent to "$\{t\} \in \mathcal{A}$ for every $t \in \mathcal{T}$".

1.10 Example: infinitely divisible and stable distributions. Let f be an infinitely divisible density on $\mathbb{R}$ satisfying (R), and let $\mathcal{P}^n$ be defined as in 1.4. Then either f is normal, or the order statistic is minimal sufficient for $\mathcal{P}^n$, for every $n \in \mathbb{N}$.

This follows easily from Theorem 1.6: It suffices to observe that a density of the form $f = \exp p$, with p a polynomial of degree ≥ 4, is too light tailed to be infinitely divisible. Compare, for example, Steutel (1974).

The result just proved applies in particular to most stable distributions on $\mathbb{R}$: Suppose that P is a stable distribution and neither normal nor with support bounded to one side. It is well known that then P has a Lebesgue density f which is continuous [by integrability of the corresponding characteristic function] and everywhere positive [see, for example, Zolotarev (1986), page 134, Theorem 2.7.6]. Hence f satisfies (R), and it follows that the order statistic is minimal sufficient for $\mathcal{P}^n$. Except when f is a Cauchy density, this seems to be a case where it would be difficult to compute a minimal sufficient σ-algebra directly via the usual approach, using formula (15) below.

1.11 Counterexample: uniform distributions. Without assumption (R) or its putative substitute (R_0), a counterexample in case $n \geq 3$ for the implication (A) $\Rightarrow$ (C) in Theorem 1.6, and also for the implication (A) $\Rightarrow$ (E) in Corollary 1.8, would be given by the uniform density $f = 1_{[0,1]}$. To see this, recall that $\mathcal{C} := \sigma(\min X_i, \max X_i)$ is complete and sufficient in this case, and hence also minimal sufficient. Since we do not have $\mathcal{C} = \mathcal{B}(\mathbb{R}^n)\ [\mathcal{P}^n]$, by $n \geq 3$ and the corollary from 2.3 applied to, say, $G := \{x \in \mathbb{R}^n : 0 < x_1 < \ldots < x_n < 1\}$, it follows that the order statistic is not minimal sufficient.

Without (R) or (R_0), the uniform distribution would also be a counterexample to Corollary 1.9: Take $T := (\min X_i, \max X_i)$. To check the validity of assumption iii) in this case, one can use Basu's theorem as in the proof of Corollary 1.8 given in 3.2 below.

1.12 Remarks, in particular on related work.

a) As already indicated, the equivalence (B) $\Leftrightarrow$ (C) in 1.6 is due to Dynkin (1951) and Ferguson (1962), under stronger and weaker regularity conditions, respectively. Related works, yielding descriptions of the possible forms of exponential families generated by transformation groups, are Borges & Pfanzagl (1965), Maksimov (1967), Engert (1970), Sapozhnikhov (1970), Roy (1975), and Rukhin (1975; 1981).

b) Some readers might suspect that the implication (A) $\Rightarrow$ (C) of Theorem 1.6, stated to be the main result of this paper, could easily be deduced from results available in the statistical literature, perhaps under stronger regularity assumptions on the density f.

Indeed, Rukhin (1975) writes on page 153, after generalizing results of Dynkin (1951): "Thus distributions (I) are characterized by property of existence of nontrivial sufficient statistics within the class of all continuous and positive densities." Specialized to the present context, the "distributions (I)" are those of (C) in our Theorem 1.6, so that Rukhin's claim might appear to yield our implication (A) $\Rightarrow$ (C). This is, however, not the case: The definition of "nontriviality" of a sufficient statistic T adopted by Dynkin (1951) and Rukhin (1975) is a priori more exclusive than the condition "not $\sigma(T) = \mathcal{B}(\mathbb{R}^n)_{\mathrm{sym}}\ [\boldsymbol{\lambda}^n]$" corresponding to our condition (A). To see this, let us look at the corresponding situation for location parameter models. In that case, the "distributions (I)" are those with density $f = \exp p$ where now p is an exponential polynomial: $p(x) = \sum_{\lambda\in\Lambda}\sum_{k=0}^{K(\lambda)} x^k \exp(\lambda x)$ for some finite $\Lambda \subset \mathbb{C}$ and finite-valued K. On the other hand, as is observed on page 18 of Torgersen (1965), any density f of the form $f = f_1 f_2$ with f_1 a normal density and f_2 1-periodic admits as a sufficient statistic the pair consisting of $\sum X_i$ and of the order statistic of the fractional parts of the X_i. It is easy to see that, for such an f, the order statistic of the X_i is not minimal sufficient for sample size $n \geq 2$ and that, for a suitable choice of a continuous and positive f_2, the function $\log p$ is not an exponential polynomial. This shows that the perhaps expected location parameter analogue of our location-scale parameter implication (A) $\Rightarrow$ (C) is not valid. This does not contradict the claim of Rukhin (1975), due to his more exclusive definition of "nontriviality" of a sufficient statistic: According to the definitions adopted by Dynkin and Rukhin, the sufficient statistic given above for the Torgersen example is trivial, compare the definition of "trivial" on page 22 of the english translation of Dynkin (1951).

c) Some readers might wonder why, in Subsection 3.1 below, we prove the equivalence (A) $\Leftrightarrow$ (C) directly, that is, without referring to condition (B). Of course, this renders our contribution independent of the Dynkin-Ferguson theorem (B) $\Leftrightarrow$ (C), but the true reasons for our approach are the following two:

First, it is not possible to prove (A) $\Rightarrow$ (B) without somehow exploiting the location-scale parameter structure of $\mathcal{P}$. To see this, observe that the best theorems available yielding exponentiality of $\mathcal{P}$ as a conclusion from the existence of a sufficient reduction beyond the order statistic in the model $\mathcal{P}^n$ need dimensionality and regularity assumptions concerning a sufficient statistic, such as one-dimensionality and local Lipschitz continuity in the theorem of Hipp (1974) referring to models not necessarily involving a group structure, and one-dimensionality, continuity and equivariance in the theorems of Pfanzagl (1972) and Hipp (1975) referring to transformation models. Leaving aside the fact that at least two-dimensional analogues of these theorems would be needed in the present situation, it is still not a priori clear that similar assumptions are a consequence of (A). Indeed, the Torgersen example mentioned in Remark b) above shows that the implication (A) $\Rightarrow$ (B) is false in the analogous location parameter situation.

Second, it is not possible to prove (B) $\Rightarrow$ (A) without using again the location-scale structure of $\mathcal{P}$. To see this, observe that there exist one-parameter exponential families $\mathcal{P}$ with continuous Lebesgue densities such that, for every sample size n, the order statistic is minimal sufficient for the corresponding model $\mathcal{P}^n$ for n independent observations. Such families $\mathcal{P}$ have been constructed in Theorem 2.3 of Mattner (1999 b).

For these reasons, our approach of proving (A) $\Leftrightarrow$ (C) directly appears natural to us.

d) Subject to assumption (R), the implication (A) $\Rightarrow$ (E) of Corollary 1.8 solves the problem posed on page 997 of Ferguson (1962).

e) Subject to assumption (R), the implication (D) $\Rightarrow$ (E) of Corollary 1.8 generalizes Theorem 3 of Kelker & Matthes (1970), who assume from the beginning that $T = (\overline{X}_n, S_n^2)$, and also that $n \geq 4$. Similarly, again subject to assumption (R), Corollary 1.9 generalizes the Lemma on page 1088 of Kelker & Matthes (1970). Bondesson (1975) generalizes that Lemma in another direction.

f) Another reasonably large class of models, for which minimal sufficient σ-algebras have been computed for the corresponding models of independent observations, is the class of all convex models, see Mattner (2000).

2 Auxiliary facts

2.1 The translation and dilation invariant closed subspaces of $\mathcal{C}(\mathbb{R})$. The following known theorem is the crucial tool for our proof of the implication (A) $\Rightarrow$ (C) of Theorem 1.6.

Theorem. *Every translation and dilation invariant closed subspace $\mathcal{H}$ of $\mathcal{C}(\mathbb{R})$ is either the entire space or, for some integer k, the set of all polynomial functions of degree at most k.*

Here $\mathcal{C}(\mathbb{R})$ denotes the set of all continuous functions on $\mathbb{R}$, "function" is to be read as "$\mathbb{K}$-valued function" with either $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$ throughout, "subspace" refers to the $\mathbb{K}$-vector space structure of $\mathcal{C}(\mathbb{R})$, "closed" refers to uniform convergence on compact sets, and to say that $\mathcal{H}$ is translation and dilation invariant means that $(x \mapsto h(ax+b)) \in \mathcal{H}$ whenever $h \in \mathcal{H}$, $a \in]0, \infty[$, $b \in \mathbb{R}$.

There exist at least two different proofs of the theorem. Let us first observe that the two cases $\mathbb{K} = \mathbb{R}$ and $\mathbb{K} = \mathbb{C}$ are easily reduced to each other.

For $\mathbb{K} = \mathbb{R}$, the theorem is contained in Theorem A of Leland (1968): Apply Leland's Theorem A to the set of all restrictions $\{h|_U : h \in \mathcal{H}, U \subset \mathbb{R} \text{ open}\}$.

Alternatively, the theorem for $\mathbb{K} = \mathbb{C}$ can easily be deduced from the following deep result of Schwartz (1947): Every translation invariant closed subspace $\mathcal{H}$ of $\mathcal{C}(\mathbb{R})$ is the closed span of the set of exponential monomials it includes, $\{x \mapsto x^k e^{\lambda x} : \lambda \in \Lambda, k \in \mathbb{N}_0, k \leq K(\lambda)\}$, where, unless $\mathcal{H} = \mathcal{C}(\mathbb{R})$, the set $\Lambda \subset \mathbb{C}$ is discrete and K is $\mathbb{N}_0$-valued. Here we have put $\mathbb{N}_0 := \{0\} \cup \mathbb{N}$.

2.2 A lemma on independence. *Let $(\mathcal{X}, \mathcal{A}, \mathbb{P})$ be a probability space and let $\mathcal{C}$, $\mathcal{D}$ be independent sub-σ-algebras of $\mathcal{A}$. If $\mathcal{E}$ is another sub-σ-algebra with*

(2) $$\mathcal{E} \subset \sigma(\mathcal{C}, \mathcal{D}) \ [\mathbb{P}],$$

(3) $$\mathcal{E} \text{ independent of } \mathcal{D},$$

(4) $$\mathcal{C} \subset \mathcal{E} \ [\mathbb{P}],$$

then $\mathcal{E} = \mathcal{C}$ $[\mathbb{P}]$.

Remark. Simple examples show that one can not in general conclude that $\mathcal{E} \subset \mathcal{C}\ [\mathbb{P}]$ if any one of the conditions (2), (3), (4) is omitted.

Proof. Under the assumptions stated, let $E \in \mathcal{E}$. By (2), the indicator 1_E is almost everywhere equal to any version of the conditional expectation $\mathbb{P}(E|\sigma(\mathcal{C},\mathcal{D}))$. The latter equivalence class contains $\mathbb{P}(E|\mathcal{C})$, as can be seen by using the independence of $\sigma(E,\mathcal{C}) \subset \mathcal{E}\ [\mathbb{P}]$ and $\mathcal{D}$, and by applying a standard property of conditional expectation [see Williams (1991), page 88, property (k)]. Hence there is a $C \in \mathcal{C}$ with $E = C\ [\mathbb{P}]$. ∎

2.3 Comparison of differentiably generated σ-algebras. For proving rigorously that certain σ-algebras are not equivalent modulo null sets, it appears worthwhile to state the following simple facts.

Lemma. *Let $n, m_1, m_2 \in \mathbb{N}$ and let $G \subset \mathbb{R}^n$ be open. For $i \in \{1,2\}$, let $f_i : G \to \mathbb{R}^{m_i}$ be continuously differentiable functions, and put*

$$p_i \ := \ \max\{\operatorname{rank} f_i'(x) \ : \ x \in G\}.$$

If $\sigma(f_1) \subset \sigma(f_2)\ [\boldsymbol{\lambda}^n]$, then $p_1 \leq p_2$.

Corollary. *Let $k, n \in \mathbb{N}$ with $k < n$, let $G \subset \mathbb{R}^n$ be open, and let $f : G \to \mathbb{R}^k$ be continuously differentiable. Then we do not have $\sigma(f) = \mathcal{B}(G)\ [\boldsymbol{\lambda}^n]$.*

Triviality. *Let $n, m \in \mathbb{N}$, let $G \subset \mathbb{R}^n$ be open, and let $f : G \to \mathbb{R}^m$ be continuous. If $\sigma(f) = \{\emptyset, G\}\ [\boldsymbol{\lambda}^n]$, then f is constant.*

We remark that the corollary becomes false if "continuously differentiable" is replaced by "continuous". This follows from a famous theorem of Denny (1964). An alternative proof of Denny's theorem is given in Mattner (1999 b).

Proof of the triviality. If f is not constant, then we can choose $a, b \in f(G)$ and $U, V \subset \mathbb{R}^m$ open and disjoint with $a \in U$, $b \in V$. Then $f^{-1}(U)$, $f^{-1}(V) \in \sigma(f)$ are nonempty and open in $\mathbb{R}^n$, hence of positive Lebesgue measure, and disjoint. It follows that we do not have $f^{-1}(U) \in \{\emptyset, G\}\ [\boldsymbol{\lambda}^n]$. ∎

Proof of the corollary. Apply the lemma to $f_1 =$ the identity on G and $f_2 = f$. ∎

Proof of the lemma. Let us assume, with the aim of arriving at a contradiction, that

$$\sigma(f_1) \ \subset \ \sigma(f_2) \quad [\boldsymbol{\lambda}^n] \tag{5}$$

and that

$$p_1 \ > \ p_2. \tag{6}$$

Let us put $I :=]-1,1[$. By the rank theorem [see Dieudonné (1960), Section 10.3], there exist open sets $U \subset G$, $V \subset \mathbb{R}^{m_1}$ and $\mathcal{C}^1$-diffeomorphisms $u : U \to I^n$ and $v : I^{m_1} \to V$, such that

$$f_1|_U \ = \ v \circ \pi \circ u \tag{7}$$

where $\pi : I^n \to I^{m_1}$ is given by

$$\pi(x_1, \ldots, x_n) \ = \ (x_1, \ldots, x_{p_1}, \underbrace{0, \ldots, 0}_{m_1 - p_1}) \qquad ((x_1, \ldots, x_n) \in I^n). \tag{8}$$

If we replace G by U and the f_i by their restrictions $f_i|_U$, then p_2 might decrease, but (5) and (6) clearly remain valid. Hence we may assume that $U = G$ in what follows. Since v is in particular a Borel isomorphism of I^{m_1} onto the Borel set V, we have $\sigma(v \circ \pi \circ u) = \sigma(\pi \circ u)$, and hence may assume that v is the identity, so that

$$f_1 \quad = \quad \pi \circ u. \tag{9}$$

Further, $\sigma(\pi \circ u)$ does not change if replace m_1 by p_1, so that (8) is replaced by

$$\pi(x_1, \ldots, x_n) \quad = \quad (x_1, \ldots, x_{p_1}) \qquad ((x_1, \ldots, x_n) \in I^n). \tag{10}$$

Now assume for a moment that $p_1 < n$, and put

$$g \quad := \quad ((x_1, \ldots, x_n) \mapsto (x_{p_1+1}, \ldots, x_n)) \circ u. \tag{11}$$

By (5), we have $\sigma((f_1, g)) \subset \sigma((f_2, g))$ $[\boldsymbol{\lambda}^n]$. Further, from (9), (10), (11), we have $\operatorname{rank}(f_1, g)' = n$ in G, while, using (6), we get $\operatorname{rank}(f_2, g)' \leq p_2 + (n - p_1) < n$. Hence, by passing from f_1 and f_2 to (f_1, g) and (f_2, g), we arrive at (5) and (6) with $p_1 = n$.

Thus we may assume that $p_1 = n$ for the rest of this proof. Then $f_1 = u$ is a $\mathcal{C}^1$-diffeomorphism, so that

$$\sigma(f_1) = \mathcal{B}(G). \tag{12}$$

By applying now the rank theorem to f_2, and by repeating the arguments that led to (9), we may assume that

$$f_2 \quad = \quad \pi \circ u$$

where the new function $u : G \to I^n$ is again a $\mathcal{C}^1$-diffeomorphism, and now

$$\pi(x_1, \ldots, x_n) \quad = \quad (x_1, \ldots, x_{p_2}) \qquad ((x_1, \ldots, x_n) \in I^n). \tag{13}$$

In view of (12), assumption (5) now reads $\sigma(f_2) = \mathcal{B}(G)$ $[\boldsymbol{\lambda}^n]$. Since the $\mathcal{C}^1$-diffeomorphism u transforms $\boldsymbol{\lambda}^n$-nullsets into $\boldsymbol{\lambda}^n$-nullsets [see, for example, Lemma 7.25 of Rudin (1987)], it follows that $\sigma(f_2 \circ u^{-1}) = \mathcal{B}(I^n)$ $[\boldsymbol{\lambda}^n]$. Since $f_2 \circ u^{-1} = \pi$ from (13), we deduce [using well known facts given as Theorem II.5.2 (i) in Heyer (1982) and Theorem 1.5.1 (ii) in Torgersen (1991)] the existence of a Borel function $h : I^{p_2} \to I$ with

$$h(x_1, \ldots, x_{p_2}) \quad = \quad x_n \qquad \text{for } \boldsymbol{\lambda}^n\text{-a.e. } x \in I^n.$$

Using $p_2 < n$ and $\int_I x_n \, dx_n = 0$, we derive

$$0 < \int_{I^n} x_n^2 \, dx_1 \ldots dx_n = \int_{I^{n-1}} \left[h(x_1, \ldots, x_{p_2}) \int_I x_n \, dx_n \right] dx_1 \ldots dx_{n-1} = 0,$$

a contradiction proving our claim. ■

3 Proofs of the main results.

3.1 Proof of Theorem 1.6. Let us fix $n \in \mathbb{N}$.

PROOF OF (A) $\Rightarrow$ (C): By assumption (R), we have in particular (R_0), and hence may assume without loss of generality that

$$f \;>\; 0 \text{ everywhere.} \tag{14}$$

By Bahadur's version of a theorem of Lehmann & Scheffé [see Torgersen (1991), p. 69], the σ-algebra

$$\mathcal{C}_0 \;:=\; \sigma\Big(\Big\{x \mapsto \frac{\prod_{i=1}^n \frac{1}{b} f(\frac{x_i - a}{b})}{\prod_{i=1}^n f(x_i)} \;:\; a \in \mathbb{R},\, b \in]0,\infty[\Big\}\Big) \tag{15}$$

is minimal sufficient for $\mathcal{P}^n$. Let us put

$$g \;:=\; \log f. \tag{16}$$

It is then easy to check that

$$\mathcal{C}_0 \;=\; \sigma\Big(\Big\{x \mapsto \sum_{i=1}^n \Big(g(\frac{x_i - a}{b} - y) - g(\frac{x_i - c}{d} - y)\Big) \;:\; a, c, y \in \mathbb{R},\, b, d \in]0,\infty[\Big\}\Big). \tag{17}$$

Let now $\varphi : \mathbb{R} \to \mathbb{R}$ be continuous with compact support, let $h : \mathbb{R} \to \mathbb{R}$ be the convolution of g with φ,

$$h(x) \;=\; \int_{\mathbb{R}} g(x - y)\varphi(y)\, dy \qquad (x \in \mathbb{R}), \tag{18}$$

which is a well defined continuous function thanks to assumption (R), and consider the σ-algebra

$$\mathcal{C}_1 \;:=\; \sigma\Big(\Big\{x \mapsto \sum_{i=1}^n \Big(h(\frac{x_i - a}{b}) - h(\frac{x_i - c}{d})\Big) \;:\; a, c \in \mathbb{R},\, b, d \in]0,\infty[\Big\}\Big). \tag{19}$$

One might think that necessarily $\mathcal{C}_1 \subset \mathcal{C}_0$, since the functions H generating $\mathcal{C}_1$ as in (19) are of the form

$$H(x) \;=\; \int_{\mathbb{R}} G(x, y)\varphi(y)\, dy \qquad (x \in \mathbb{R}^n), \tag{20}$$

with functions $G(\cdot, y)$ generating $\mathcal{C}_0$ as in (17). From this, however, the wanted inclusion $\mathcal{C}_1 \subset \mathcal{C}_0$ does not follow, since H from (20) need not be measurable with respect to $\sigma(\{G(\cdot, y) \,:\, y \in \mathbb{R}\})$. [See Section 2.1 of Mattner (1999 a) for a counterexample.] On the other hand, it is true that the function H from (20) is $\boldsymbol{\lambda}^n$-almost everywhere equal to a $\sigma(\{G(\cdot, y) \,:\, y \in \mathbb{R}\})$-measurable function. [This follows from Theorem 3.1 of Mattner (1999 a), using the product measurability of G with respect to $\mathcal{B}(\mathbb{R}^n) \otimes \mathcal{B}(\mathbb{R})$.] Hence we can conclude that

$$\mathcal{C}_1 \;\subset\; \mathcal{C}_0 \quad [\boldsymbol{\lambda}^n]. \tag{21}$$

Now assume (A). This means that we do not have $\mathcal{C}_0 = \mathcal{B}(\mathbb{R}^n)_{\mathrm{sym}}\ [\boldsymbol{\lambda}^n]$. In view of (21) and of $\mathcal{C}_0$, $\mathcal{C}_1 \subset \mathcal{B}(\mathbb{R}^n)_{\mathrm{sym}}$, we get the strict inclusion

$$\mathcal{C}_1 \;\subsetneq\; \mathcal{B}(\mathbb{R}^n)_{\mathrm{sym}}. \tag{22}$$

By the continuity of h from (18), the σ-algebra $\mathcal{C}_1$ is countably generated. Using a theorem of Blackwell [see Dellacherie & Meyer (1975), p. 80], we deduce the existence of $x, y \in \mathbb{R}^n$ satisfying

$$x \text{ is not a permutation of } y, \tag{23}$$

and

$$\sum_{i=1}^{n} \left(h(\tfrac{x_i-a}{b}) - h(\tfrac{x_i-c}{d})\right) \;=\; \sum_{i=1}^{n} \left(h(\tfrac{y_i-a}{b}) - h(\tfrac{y_i-c}{d})\right) \qquad (a, c \in \mathbb{R},\, b, d \in]0, \infty[). \tag{24}$$

Let us consider

$$\mathcal{H} \;:=\; \{H \in \mathcal{C}(\mathbb{R}) : (24) \text{ holds with } H \text{ in place of } h\}.$$

Obviously, $\mathcal{H}$ is a translation and dilation invariant closed subspace of $\mathcal{C}(\mathbb{R})$. Further, $\mathcal{H}$ does not contain every monomial $x, \ldots, x^n$, for otherwise we could insert these monomials into (24) and take $a = c = 0$ and $b \neq d$ to deduce $\sum_{i=1}^n x_i^k = \sum_{i=1}^n y_i^k$ for $k = 1, \ldots, n$, which is well known to contradict (23). Hence, by the theorem from 2.1, there is an integer $m \leq n-1$ such that $\mathcal{H}$ consists of the polynomials of degree at most m. Thus, in particular, h is such a polynomial. Since this is true for every choice of φ in the definition (18) of h, it follows that g is $\boldsymbol{\lambda}$-almost everywhere equal to a polynomial of degree at most $n-1$, so that (C) is true.

Proof that in case of (C) the degree k of p satisfies the claim "k is even and ≥ 2 and $(\sum X_i, \ldots, \sum X_i^k)$ is minimal sufficient.": That k must be even and ≥ 2 follows from integrability of $f = \exp p$.

To prove minimal sufficiency of $T := (\sum X_i, \ldots, \sum X_i^k)$, we may use formula (15), which here yields the minimal sufficient σ-algebra

$$\mathcal{C} \;=\; \sigma(\{P(\cdot, a, b) \,:\, a \in \mathbb{R},\, b \in]0, \infty[\}),$$

where

$$P(x, a, b) \;:=\; \sum_{i=1}^{n} (p(\frac{x_i - a}{b}) - p(x_i)) \qquad (x \in \mathbb{R}^n,\, a \in \mathbb{R},\, b \in]0, \infty[).$$

Obviously, $\mathcal{C} \subset \sigma(T)$. To prove $\sigma(T) \subset \mathcal{C}$, consider for $c \in \mathbb{R}$ the difference operator Δ_c defined on functions $F : \mathbb{R}^n \to \mathbb{R}$ by $(\Delta_c F)(x) = F(x_1{+}c, \ldots, x_n{+}c) - F(x_1, \ldots, x_n)$. For $m \geq 1$, we clearly have $\Delta_c^m \sum p(X_i) \in \operatorname{span}\{P(\cdot, a, 1) \,:\, a \in \mathbb{R}\}$. Since $\Delta_c^{k-1} \sum p(X_i) = \alpha_1 \sum X_i + \alpha_2$ with $\alpha_1 \neq 0$, it follows that $\sum X_i$ is $\mathcal{C}$-measurable. In case $k \geq 3$, we then use $\Delta_a^{k-2} \sum p(X_i) = \beta_1 \sum X_i^2 + \beta_2 \sum X_i + \beta_3$ with $\beta_1 \neq 0$ to deduce $\mathcal{C}$-measurability of $\sum X_i^2$. Continuing in this way, we arrive at the $\mathcal{C}$-measurability of $(\sum X_i, \ldots, \sum X_i^{k-1})$. To finally prove $\mathcal{C}$-measurability of $\sum X_i^k$, we observe that $P(\cdot, 0, 1) - P(\cdot, 0, 2) = \gamma_1 \sum X_i^k + \gamma_2 \sum X_i^{k-1} + \ldots \gamma_{k+1}$ with $\gamma_1 \neq 0$.

Proof that (C) $\Rightarrow$ (A): Assume (C). Then, by what has already been proved, $T := (\sum X_i, \ldots, \sum X_i^k)$ is minimal sufficient for $\mathcal{P}_n$. Put $G := \{x \in \mathbb{R} \,:\, x_1 < x_2 < \ldots < x_n\}$. If the order statistic were minimal sufficient too, then we would have in particular $\sigma(T|_G) = \mathcal{B}(G)$ $[\boldsymbol{\lambda}^n]$. Since $k < n$ and since $T|_G$ is continuously differentiable, this would contradict the corollary in 2.3.

Proof that (B) $\Leftrightarrow$ (C) and that the dimension equals the degree: This is due to Dynkin (1951) and Ferguson (1962), see Theorem 4 of the latter. ∎

3.2 Proof of Corollary 1.8.

(E) $\Rightarrow$ (A): It is well known that $(\overline{X}_n, S_n^2)$ is complete sufficient for $\mathcal{P}^n$ if f is normal.

(A) $\Rightarrow$ (B): Trivial.

(B) $\Rightarrow$ (C): This follows from Basu's theorem [Theorem 1 in Basu (1982)], since $\mathcal{D}$ is ancillary.

(C) $\Rightarrow$ (E): Here we use the assumptions (R) and $n \geq 3$. To avoid treating two cases separately, let us first observe that Theorem 1.6 always yields a $k \in \{2, \ldots, n\}$ such that $T := (\sum X_i, \ldots, \sum X_i^k)$ is minimal sufficient: If (A) of 1.6 is false, we take $k = n$ and use $\sigma(T) = \mathcal{B}(\mathbb{R}^n)_{\mathrm{sym}}$.

If $k = 2$, then f is normal.

If $k > 2$, then the statistic

$$U \ := \ \sum_{i=1}^{n} \left(\frac{X_i - \overline{X}_n}{S_n} \right)^3 \ = \ h(\sum X_i, \sum X_i^2, \sum X_i^3),$$

for some measurable function $h : \mathbb{R}^3 \to \mathbb{R}$, is $\mathcal{D}$-measurable and T-measurable. By assumption (C) of 1.8 and by minimal sufficiency of T, this implies that U is $\mathcal{P}^n$-independent of itself. Hence $U = c \ [\boldsymbol{\lambda}^n]$ for some $c \in \mathbb{R}$. Since $n \geq 3$, this is impossible. To prove rigorously the claimed impossibility, we may compute that $U(1, 0, \ldots, 0) > 0$ and hence $U(-1, 0, \ldots, 0) < 0$, and apply the triviality from 2.3 to $G := \{S_n > 0\}$ and $f = U|_G$.

(E) $\Rightarrow$ (D): Take $T := (\overline{X}_n, S_n)$ and $N := \{S_n = 0\}$.

(D) $\Rightarrow$ (C): Apply the Corollary 1.9.15 to Proposition 1.9.11 of Pfanzagl (1994). ∎

3.3 Proof of Corollary 1.9. Let $\mathcal{P}^n$ be defined as in 1.4. Let us put $\tilde{N} := N \cup \{x \in \mathbb{R}^n : S_n(x) = 0\}$, with N as in assumption ii). Then, since $n \geq 2$, $\tilde{N}$ is a $\boldsymbol{\lambda}^n$-nullset. Using assumption ii), the statistic

$$U \ := \ (T, \frac{X_1 - \overline{X}_n}{S_n}, \ldots \frac{X_n - \overline{X}_n}{S_n})$$

is easily seen to be injective on $\mathbb{R}^n \setminus \tilde{N}$. Hence, using the assumptions on $(\mathcal{T}, \mathcal{A})$ and again the theorem of Blackwell [Dellacherie & Meyer (1975), page 80], we conclude that $\sigma(U|_{(\mathbb{R}^n \setminus \tilde{N})}) = \mathcal{B}(\mathbb{R}^n \setminus \tilde{N})$. Since $\tilde{N}$ is a nullset, U is in particular sufficient for $\mathcal{P}^n$.

By assumption iii), T is independent of $\mathcal{D}$ under $P^{\otimes n}$. To deduce independence also under $\mathcal{P}^n$, we argue as follows: Fix $g \in G$ for the moment. By the equivariance assumption i), the functions T and $x \mapsto T(gx)$ generate the same partition on $\mathbb{R}^n$. By the assumptions on $(\mathcal{T}, \mathcal{A})$ and once more using Blackwell's theorem, it follows that the two functions generate the same σ-algebra. It follows that $\sigma(T)$ is G-invariant [in the sense of "$B \in \sigma(T)$, $g \in G \Rightarrow gB \in \sigma(T)$"]. Since $\mathcal{D}$ is trivially G-invariant, and since $\mathcal{P}^n$ is generated from $P^{\otimes n}$ via G, we get indeed the independence of T and $\mathcal{D}$ under $\mathcal{P}^n$.

This independence taken together with the sufficiency of U and the ancillarity of $\mathcal{D}$ yields, by Theorem 3 of Basu (1982), that already T is sufficient for $\mathcal{P}^n$. Hence (C) of Corollary 1.8 holds with $\mathcal{C} = \sigma(T)$, so that P must be normal.

Thus we know that $(\overline{X}_n, S_n)$ is minimal sufficient for $\mathcal{P}_n$, for example from Theorem 1.6. With $\mathcal{C} := \sigma(\overline{X}_n, S_n)$ and $\mathcal{E} := \sigma(T)$, sufficiency of $\mathcal{E}$ yields $\mathcal{C} \subset \mathcal{E} \ [\mathcal{P}^n]$. Now an application of Lemma 2.2, to $\mathbb{P} := P^{\otimes n}$, $\mathcal{C}$ and $\mathcal{E}$ as just defined, and $\mathcal{D}$ from (1), yields $\sigma(T) = \sigma(\overline{X}_n, S_n) \, [P^{\otimes n}]$, which is equivalent to the final claim. ∎

Acknowledgement. The present work was supported by a Heisenberg grant of the Deutsche Forschungsgemeinschaft. I thank P. Mörters, J. Pfanzagl, and two referees for helpful remarks on earlier versions.

References

Basu, D. (1982). Basu theorems. In S. Kotz and N.L. Johnson (eds.), *Encyclopedia of Statistical Sciences, Vol. 1*, pp. 193-196. New York: Wiley. Also in J.K. Ghosh (1988) (ed.), *Statistical Information and Likelihood. A Collection of Critical Essays of Dr. D. Basu.*, Lecture Notes in Statistics, Vol. 45, pp. 342-349. New York: Springer.

Bondesson, L. (1975). Normal distribution, gamma distribution, and quadratic polynomial statistics. *Scand. J. Statist.*, **2**, 138-144.

Borges, R. & Pfanzagl, J. (1965). One-parameter exponential families generated by transformation groups. *Ann. Math. Statist.*, **36**, 261-271.

Dellacherie, C. & Meyer, P.-A. (1975). *Probabilités et potentiel. Chapitres I à IV.* Paris: Hermann.

Denny, J.L. (1964). A continuous real-valued function on E^n almost everywhere 1-1. *Fund. Math.*, **60**, 95-99.

Dieudonné, J. (1960). *Foundations of Modern Analysis.* New York: Academic Press.

Dynkin, E.B. (1951). Necessary and sufficient statistics for a family of probability distributions. *Uspehi Mat. Nauk (N.S.)*, **6**, 68-90. (In russian. Here cited after the english translation in: *Selected Translations in Mathematical Statistics and Probability*, Vol. 1, American Mathematical Society, Providence, 1961, pp. 17-40.)

Engert, M. (1970). Finite dimensional translation invariant subspaces. *Pacific J. Math.*, **32**, 333-343.

Ferguson, T.S. (1962). Location and scale parameters in exponential families of distributions. *Ann. Math. Statist.*, **33**, 986-1001. Correction: **34** (1963), 1603.

Helland, I. (1998). Statistical inference under a fixed symmetry group. Preprint, available at `http://www.math.uio.no/~ingeh/publ.html`.

Heyer, H. (1982). *Theory of Statistical Experiments.* New York: Springer.

Hipp, C. (1974). Sufficient statistics and exponential families. *Ann. Statist.*, **2**, 1283-1292.

Hipp, C. (1975). Note on the paper "Transformation groups and sufficient statistics" by J. Pfanzagl. *Ann. Statist.*, **3**, 478-482.

Kelker, D. & Matthes, T.K. (1970). A sufficient statistics characterization of the normal distribution. *Ann. Math. Statist.*, **41**, 1086-1090.

Leland, K.O. (1968). A characterization of analyticity. III. *J. Math. Mech.*, **18**, 109-123.

Maksimov, V.M. (1967). Necessary and sufficient statistics for a family of shifts of probability distributions on continuous bicompact groups. *Theor. Probab. Appl.*, **12**, 267-280.

Mattner, L. (1999 a). Product measurability, parameter integrals, and a Fubini counterexample. *L'Enseignement Mathématique*, **45**, 271-279.

Mattner, L. (1999 b). Sufficiency, exponential families, and algebraically independent numbers. *Math. Meth. Statist.*, **8**, 397-406.

Mattner, L. (2000). Minimal sufficient order statistics in convex models. *Proc. Amer. Math. Soc.* (to appear).

Pfanzagl, J. (1972). Transformation groups and sufficient statistics. *Ann. Math. Statist.*, **43**, 553-568.

Pfanzagl, J. (1994). *Parametric Statistical Theory.* Berlin: de Gruyter.

Ramamoorthi, R.V. (1990). Sufficiency, ancillarity and independence in invariant models. *J. Statist. Planning Inference*, **26**, 59-63.

Roy, K.K. (1975). Exponential families of densities on an analytic group and sufficient statistics. *Sankhyā, Ser. A*, **37**, 82-92.

Rudin, W. (1987). *Real and Complex Analysis.* 3rd. ed. New York: McGraw-Hill.

Rukhin, A.L. (1975). Characterizations of distributions by statistical properties on groups. In G.P. Patil et al. (eds.), *A Modern Course on Statistical Distributions in Scientific Work. Volume 3 - Characterizations and Applications (Proceedings Calgary 1974)*, pp. 149-161. Dordrecht: Reidel.

Rukhin, A.L. (1981). Distributions with sufficient statistics for multivariate location parameter and transformation parameter. In C. Taillie et al. (eds.), *Statistical Distributions in Scientific Work, Vol. 4 (Proceedings Trieste 1980)*, pp. 243-254. Dordrecht: Reidel.

Sapozhnikhov, P.N. (1970). Invariant finite-dimensional exponential families on homogeneous spaces. *Math. Notes*, **7**, 427-431.

Sapozhnikhov, P.N. (1998). Shift families that admit nontrivial sufficient statistics. *J. Math. Sci. (New York)*, **88**, 848-855.

Schwartz, L. (1947). Théorie générale des fonctions moyenne-périodiques. *Ann. Math.*, **48**, 857-929.

Steutel, F.W. (1974). On the tails of infinitely divisible distributions. *Zeitschrift Wahrscheinlichkeitstheorie verw. Geb.*, **28**, 273-276.

Torgersen, E. (1965). Minimal sufficiency of the order statistic in the case of translation- and scale parameters. *Skand. Aktuarietidsskrift*, **48**, 16-21.

Torgersen, E. (1991). *Comparison of Statistical Experiments.* Cambridge: Cambridge University Press.

Wijsman, R.A. (1990). *Invariant Measures on Groups and Their Use in Statistics.* IMS Lecture Notes-Monograph Series, Vol. 14. Hayward, California: Institute of Mathematical Statistics.

Williams, D. (1991). *Probability with Martingales.* Cambridge: Cambridge University Press.

Zolotarev, V.M. (1986). *One-dimensional stable distributions.* Translations of Mathematical Monographs, Volume 65. Providence, Rhode Island: American Mathematical Society.

Universität Hamburg, Institut für Mathematische Stochastik, Bundesstr. 55, 20146 Hamburg, Germany

E-mail address: `mattner@math.uni-hamburg.de`