The Annals of Statistics
1983, Vol. 11, No. 3, 979–983

CHARACTERIZATION OF TYPE FROM MAXIMAL INVARIANT SPECTRA

BY CHRISTOPHER G. SMALL

Simon Fraser University

The affine type of distributions on the real line are represented as sequences of distributions of maximal invariants on spheres. It is shown that such a representation characterizes the affine type. A consistency condition is introduced, and it is shown that any sequence of maximal invariant distributions satisfying the condition is generated by some affine type on **R**.

1. Background and introduction. This paper is the second in a set of two that study characterization results for distributions of maximal invariant statistics. In (13), duality theory for locally compact abelian groups was used to derive characterization conditions for maximal invariants under actions of such groups. In this paper it will be shown that the class of affine types of distributions can be identified as the inverse limit of classes of maximal invariant distributions.

In a set of early papers, H. Hotelling [4] and E. J. G. Pitman [10, 11] introduced the use of invariant arguments into problems of inference. G. W. Brown (2) provided conditions under which maximal location and scale invariants of a sample possess power to distinguish hypotheses with location and scale nuisance parameters. His methods were analytic rather than statistical and effectively showed that affine type is characterized by the distribution of maximal invariants in most cases of statistical interest. A. A. Zinger and Yu V. Linnik [4] and L. Bondesson [1] have strengthened G. W. Brown's results. Yu. V. Prokhorov [12] developed similar results for distributions satisfying the Cramer condition.

The development of techniques by D. A. S. Fraser [3] has emphasized the importance of the affine group. Recent work by D. G. Kendall [5] makes use of maximal invariants for spatial data in what he calls the statistics of shape.

Let **R** be the real line endowed with the Borel σ-algebra. Let G be the group of transformations $\langle a, b\rangle$ of **R** of the form

$$\langle a, b\rangle x = ax + b$$

where $a > 0$ and b is any real number. Then G is called the affine group. Let $\mathscr{P}$ be the class of all (Borel) distributions on **R**. For each $P \in \mathscr{P}$ and $\langle a, b\rangle \in G$, define $(P\langle a, b\rangle)(A) = P(\langle a, b\rangle(A))$ where A is any Borel set. Then $P\langle a, b\rangle \in \mathscr{P}$. We define $[P]$, the affine type of P, by $[P] = \{P\langle a, b\rangle : a > 0, b \text{ any}\}$. Let $\mathscr{P}/G = \{[P] : P \in \mathscr{P}\}$. So $\mathscr{P}/G$, the orbit space under the action of G on $\mathscr{P}$, is the class of types in the sense of M. Loève [6].

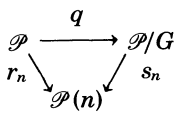
For each sample $(x_1, x_2, \cdots, x_n) \in \mathbf{R}^n$, define the maximal invariant

$$T_n(x_1, x_2, \cdots, x_n) = \begin{cases} \left(\dfrac{x_1 - \bar{x}}{v^{1/2}}, \dfrac{x_2 - \bar{x}}{v^{1/2}}, \cdots, \dfrac{x_n - \bar{x}}{v^{1/2}}\right) & \text{if } v \neq 0 \\ \{*_n\} & \text{if } v = 0. \end{cases}$$

In this definition, $\bar{x} = (\sum_{i=1}^n x_i)/n$, $v = \sum_{i=1}^n (x_i - \bar{x})^2$ and $\{*_n\}$ is some arbitrary point that is distinct from the points of $\mathbf{R}^n$. Then T_n contains all the information in the sample that is invariant under transformations of the kind $(x_1, x_2, \cdots, x_n) \to (\langle a, b\rangle x_1, \langle a, b\rangle x_2, \cdots, \langle a, b\rangle x_n)$. For $n \geq 1$, the range of T_n is $S^{n-2} \cup \{*_n\}$, where S^{n-2} is a unit $(n-2)$-sphere in $\mathbf{R}^n$. (The sphere of dimension -1 is defined as the empty set.) Let S^{n-2} be endowed with the restriction of the Borel σ-algebra from $\mathbf{R}^n$. The point $\{*_n\}$ shall be measurable.

Received September 1982; revised March 1983.
AMS 1980 *subject classifications.* Primary, 62A05; secondary, 62E10.
Key words and phrases. Maximal invariant statistic, affine group, type, weak convergence.

For a sample $(x_1, x_2, \cdots, x_n)$ drawn independently from $\mathbf{R}$ with distribution $P \in \mathscr{P}$, the maximal invariant T_n will induce the distribution $P^n T_n^{-1}$ on $S^{n-2} \cup \{*_n\}$. Define $\mathscr{P}(n) = \{P^n T_n^{-1} : P \in \mathscr{P}\}$. These spaces can be represented in the following diagram



where $q(P) = [P]$, $r_n(P) = P^n T_n^{-1}$ and $s_n([P]) = P^n T_n^{-1}$. Note that s_n is well defined and that $r_n = s_n q$.

The characterization problem is to determine whether the equality $P^n T_n^{-1} = Q^n T_n^{-1}$ implies $[P] = [Q]$. The existence of counterexamples for all $n \geq 1$ shall be discussed in Section 3. Such examples for the translation group were first noticed by I. N. Kovalenko [7]. By restriction to subclasses of $\mathscr{P}$, characterization results were proved in [2], [12] and [13], among others.

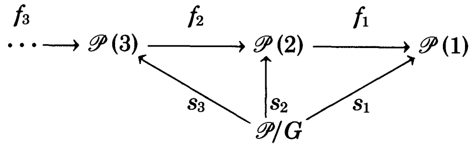
In [1], L. Bondesson noted without proof that if the distributions of the maximal translation invariants are known for all sample sizes then the additive type of the generating distribution is determined. For affine type, given $[P] \neq [Q]$ does there exist a number N such that for all $n \geq N$, $P^n T_n^{-1} \neq Q^n T_n^{-1}$? It can be shown that for other groups, even this sequential characterization can sometimes fail. For example, for the group of strictly increasing continuous transformations of $\mathbf{R}$ onto itself, the maximal invariants of a sample are the rank statistics. However, the rank statistics do not characterize the equivalence class of generating distributions under the group action. Information as to whether the distribution has compact support is not contained in the rank statistics of a finite sample.

However, for the affine group, such sequential characterization will be proved in Section 2 as part of a stronger result. For all $n \geq 1$, we can write T_n as a function of T_{n+1}. So the distribution of T_n is determined by the distribution of T_{n+1}. Suppose $R_1, R_2, \cdots$ is a sequence of distributions of $T_1, T_2, T_3, \cdots$, respectively, such that $R_n = P_n^n T_n^{-1}$ for some P_n. The main result of this paper establishes that if R_n is determined from R_{n+1} as above, then there is a P on $\mathbf{R}$ such that $R_n = P^n T_n^{-1}$ for all $n = 1, 2, \cdots$. Hence the class of affine types will be identified as the inverse limit of the classes of maximal invariant distributions.

2. The identification result. We shall use the symbol "$\Rightarrow$" to denote weak convergence in $\mathscr{P}$ and convergence in type in $\mathscr{P}/G$. A sequence $[P_n]$ in $\mathscr{P}/G$ is said to converge in type to P, $[P_n] \Rightarrow [P]$, if there exists a sequence $\langle a_n, b_n\rangle$ in G such that $P_n\langle a_n, b_n\rangle \Rightarrow P$. See M. Loève [6]. For each $n \geq 1$, let $\pi_n : \mathbf{R}^{n+1} \to \mathbf{R}^n$ be the projection defined as $\pi_n(\mathrm{x}_1, \mathrm{x}_2, \cdots, \mathrm{x}_n, \mathrm{x}_{n+1}) = (x_1, x_2, \cdots, x_n)$. We construct a sequence of mappings $\beta_n : S^{n-1} \cup \{*_{n+1}\} \to S^{n-2} \cup \{*_n\}$ by defining implicitly $T_n \pi_n = \beta_n T_{n+1}$. Then β_n can be shown to be well defined. Also $\beta_n(\{*_{n+1}\}) = \{*_n\}$.

The mappings β_n induce functions between the classes $\mathscr{P}(n)$. For all $n \geq 1$, $f_n : \mathscr{P}(n+1) \to \mathscr{P}(n)$ is defined by $f_n(R_{n+1}) = R_{n+1}\beta_n^{-1}$, where $R_{n+1} \in \mathscr{P}(n+1)$. Now R_{n+1} can be written as $P^{n+1}T_{n+1}^{-1}$ for some $P \in \mathscr{P}$. So $f_n(P^{n+1}T_{n+1}^{-1}) = P^{n+1}T_{n+1}^{-1}\ \beta_n^{-1} = P^{n+1}\pi_n^{-1}T_n^{-1} = P^n T_n^{-1}$.

Therefore $(\mathscr{P}(n), f_n)n = 1, 2, \cdots$ is a directed sequence of classes of distributions.



The *inverse limit* of the directed sequence, $\varprojlim \mathscr{P}(n)$ is

$$\varprojlim \mathscr{P}(n) = \{(R_1, R_2, \cdots) \in \mathscr{P}(1) \times \mathscr{P}(2) \times \cdots : f_n(R_{n+1}) = R_n \text{ for all } n\}.$$

We can define a function J: $\mathscr{P}/G \to \varprojlim \mathscr{P}(n)$ by setting $J([P]) = (P^1 T_1^{-1}, P^2 T_2^{-1}, \cdots)$.

Then J is the natural function into the inverse limit induced by the diagram above. Our main result shows that J is a bijection from $\mathscr{P}/G$ to $\varprojlim \mathscr{P}(n)$.

THEOREM. *Let J be defined as above. Then $J: \mathscr{P}/G \to \varprojlim \mathscr{P}(n)$ is a bijection.*

PROOF. The proof is divided into two parts.

Part 1. The proof that J is $1-1$ is given first and establishes the characterization result. For any $\mathbf{x}_n \in \mathbf{R}^n$, let $E(\mathbf{x}_n)$ be the empirical distribution on the sample, defined by assigning mass $1/n$ to each coordinate of $\mathbf{x}_n$. Then $E(\mathbf{x}_n) \in \mathscr{P}$. It can be seen that $[E(\mathbf{x}_n)] = [E(\mathbf{y}_n)]$ whenever $T_n(\mathbf{x}_n) = T_n(\mathbf{y}_n)$. So we write $[E(\mathbf{z}_n)]$ where $\mathbf{z}_n = T_n(\mathbf{x}_n)$ to denote all such equivalent representations of the type.

Suppose $P^n T_n^{-1} = Q^n T_n^{-1}$ for all $n \geq 1$. The degenerate case where P assigns unit mass to some point is dispensed with first. In that case $P^n T_n^{-1}(\{*_n\}) = Q^n T_n^{-1}(\{*_n\}) = 1$ for all $n \geq 1$. So Q is also degenerate.

Henceforth assume that P is not degenerate. Let $(\mathbf{z}_n)_{n=1,2,\cdots}$ be an independent sequence of points where each $\mathbf{z}_n$ is drawn with distribution $P^n T_n^{-1}$ from $S^{n-2} \cup \{*_n\}$. Without loss of generality, set $\mathbf{z}_n = T_n(\mathbf{x}_n)$, where $\mathbf{x}_n$ is an independent sequence and each $\mathbf{x}_n$ is drawn with distribution P^n in $\mathbf{R}^n$. Almost sure empirical convergence $E(\mathbf{x}_n) \Rightarrow P$ holds. So $\text{Prob}([E(\mathbf{z}_n)] \Rightarrow [P]) = 1$. Similarly it can be seen that $\text{Prob}([E(z_n)] \Rightarrow [Q]) = 1$. By Khintchine's Convergence of Types theorem, $[P] = [Q]$, as neither is degenerate.

Part 2. To prove J is onto, let $(R_n)_{n=1,2,\cdots}$ be a point in $\varprojlim \mathscr{P}(n)$. Then $f_n(R_{n+1}) = R_n$ for all $n \geq 1$. Once again, the case where $R_2(\{*_n\}) = 1$ can be dispensed with first. If P is degenerate, $P^n T_n^{-1} = R_n$ for all $n \geq 1$. Henceforth assume that this is not the case. For each n, choose $Q_n \in \mathscr{P}$ such that $Q_n^n T_n^{-1} = R_n$. Without loss of generality, Q_n can be chosen so as to satisfy

(i) $Q_n(x \leq 0) \geq ½$, $Q_n(x \geq 0) \geq ½$

(ii) $Q_n^2(|x_2 - x_1| \leq 1) \geq [1 - Q_n^2(x_1 = x_2)]/2$
$Q_n^2(|x_2 - x_1| \geq 1) \geq [1 - Q_n^2(x_1 = x_2)]/2.$

Each distribution, appropriately scaled and translated will satisfy (i) and (ii). For $n \geq 4$ the statistic $(\sum_{i=2}^{n-2} |x_i - x_1|)/|x_n - x_{n-1}|$ is a function of the maximal invariant $T_n(x_1, x_2, \cdots, x_n)$. Therefore for all $\varepsilon > 0$ there exists some $L_n > 0$ such that

$$Q_m^n(0 < (\textstyle\sum_{i=2}^{n-2} |x_i - x_1|)/|x_n - x_{n-1}| < 1/L_n) < \frac{\varepsilon}{2}[1 - Q_m^2(x_1 = x_2)]$$

for all $m \geq 1$. Note that $Q_2^2(x_1 = x_2) = Q_3^2(x_1 = x_2) = \cdots$ and so the r.h.s. does not depend upon the choice of $m \geq 2$. From (ii),

$$Q_m^2(|x_2 - x_1| \geq 1) \geq [1 - Q_m^2(x_1 = x_2)]/2.$$

So,

(2A) $$Q_m^{n-2}(0 < \textstyle\sum_{i=2}^{n-2} |x_i - x_1| < 1/L_n) < \varepsilon \quad \text{for all} \quad m \geq 1.$$

Using a similar argument with the statistic $|x_4 - x_3|/|x_2 - x_1|$ it can be shown that

(2B) for all $\varepsilon > 0$ there exists $K > 0$ such that $Q_m(|x| > K) < \varepsilon$ for all $m \geq 1$.

Statement (2B) implies that the sequence $(Q_m)_{m=1,2,\cdots}$ is tight. So there exists some subsequence $Q_{m_i} \Rightarrow P$. It will be seen that P is the required distribution.

Now, $P^n T_n^{-1}(\{*_n\}) = P^n(x_1 = x_2 = \cdots = x_n) \geq \lim \sup_{i \geq 1} Q_{m_i}^n(x_1 = \cdots x_n) = R_n(\{*_n\})$. We can also write $P^n T_n^{-1}(\{*_n\}) = \lim_{L \to \infty} P^n(\sum_{i=2}^n |x_i - x_1| < 1/L) \leq \lim_{L \to \infty} \lim \inf_i Q_{m_i}^n(\sum_{i=2}^n |x_i - x_1| < 1/L) = R_n(\{*_n\})$. The last equality follows from (2A). Combining these results, $P^n T_n^{-1}(\{*_n\}) = R_n(\{*_n\})$, for all $n \geq 1$.

Let V be an open set in S^{n-2}. Then $T_n^{-1}(V)$ is open in $\mathbf{R}^n$. We write bdry(V) for the boundary of V in S^{n-2} and $\mathrm{Cl}(V)$ for its closure. Note that $T_n^{-1}(\mathrm{Cl}(V))$ is not closed in $\mathbf{R}^n$. However, $T_n^{-1}(\mathrm{Cl}(V) \cup \{*_n\})$ is closed. Suppose that $P^n T_n^{-1}(\mathrm{bdry}(V)) = 0$. Then $P^n T_n^{-1}(V) \leq \lim \inf_i Q^n_{m_i} T_n^{-1}(V) = R_n(V)$. However, in addition, $P^n T_n^{-1}(\mathrm{Cl}(V)) = P^n T_n^{-1}(\mathrm{Cl}(V) \cup \{*_n\}) - R_n(\{*_n\}) \geq \lim \sup_i Q^n_{m_i} T_n^{-1}(\mathrm{Cl}(V) \cup \{*_n\}) - R_n(\{*_n\}) = R_n(\mathrm{Cl}(V)) + R_n(\{*_n\}) - R_n(\{*_n\}) = R_n(\mathrm{Cl}(V))$. Combining these results, $P^n T_n^{-1}(V) \leq R_n(V) \leq R_n(\mathrm{Cl}(V)) \leq P^n T_n^{-1}(\mathrm{Cl}(V))$. This implies that $R_n(V) = P^n T_n^{-1}(V)$. Sets with $P^n T_n^{-1}$ – null boundaries form a separating class. Hence $R_n = P^n T_n^{-1}$ for all $n \geq 1$. So P is the required distribution. □

The Consistency Theorem of Daniell and Kolmogorov, in the form presented in (9; pages 137–140), shows that the sequence $(P^n T_n^{-1})$ can be identified with a distribution induced on the inverse limit of the sequence $(S^{n-2} \cup \{*_n\}, \beta_n)$. This inverse limit is endowed with a σ-algebra generated by the σ-algebras on the finite dimensional spheres of the sequence with the points $\{*_n\}$. Khintchine's Convergence of Types Theorem establishes that the σ-algebra on the inverse limit is sufficiently large as to permit characterization of $[P]$ from the induced distribution on that inverse limit.

3. Counterexamples to characterization. In this section, counterexamples to characterization are given. It is shown that for all $n \geq 1$ there exists $P, Q \in \mathscr{P}$ such that $P^n T_n^{-1} = Q^n T_n^{-1}$ and $[P] \neq [Q]$. The result follows from Kovalenko's construction of counterexamples for affine type. Because these are not readily accessible in the literature, an explicit construction is given here.

To construct such counterexamples, it is sufficient to find P, Q of different affine type which generate the same distribution on the maximal translation invariant statistic $(x_2 - x_1, x_3 - x_1, \cdots, x_n - x_1)$. As $T_n(x_1, x_2, \cdots, x_n)$ is a function of $(x_2 - x_1, \cdots, x_n - x_1)$ P and Q will generate the same distribution on T_n.

It will be convenient to to use characteristic functions. Let $\psi(t)$ and $\xi(t)$ be the "saw tooth" characteristic functions

$$\psi(t) = 1 - |t - 2j| \quad \text{for} \quad t \in [2j-1, 2j+1], j \text{ an integer}$$
$$\xi(t) = 2\psi(t/2) - 1.$$

See [8; page 123]. We define $\phi_1^{(m)}(t)$ and $\phi_2^{(m)}(t)$ inductively.

$$\phi_1^{(1)}(t) = \left[\frac{\psi(t) + \xi(t)}{2}\right]\psi\left(\frac{t}{2}\right), \quad \phi_2^{(1)}(t) = \left[\frac{\psi(t) + \xi(t)}{2}\right]\xi\left(\frac{t}{2}\right)$$
$$\phi_1^{(m+1)}(t) = \left[\frac{\phi_1^{(m)}(t) + \phi_2^{(m)}(t)}{2}\right]\psi\left(\frac{t}{2^{m+1}}\right)$$
$$\phi_2^{(m+1)}(t) = \left[\frac{\phi_1^{(m)}(t) + \phi_2^{(m)}(t)}{2}\right]\xi\left(\frac{t}{2^{m+1}}\right)$$

If $x_1, x_2, \cdots, x_{2^{m+1}}$ are 2^{m+1} independent points whose distributions have characteristic function $\phi_i^{(m)}(t)$, $i = 1, 2$ then $(x_2 - x_1, \cdots, x_{2^{m+1}} - x_1)$ has characteristic function

$$\phi_i^{(m)}[-\textstyle\sum_{k=1}^{2^{m+1}-2} t_k] \prod_{k=1}^{2^{m+1}-2} \phi_i^{(m)}(t_k).$$

we must prove that

$$\text{(3A)} \qquad \phi_1^{(m)}[-\textstyle\sum_{k=1}^{2^{m+1}-2} t_k] \prod_{k=1}^{2^{m+1}-2} \phi_1^{(m)}(t_k) = \phi_2^{(m)}[-\sum_{k=1}^{2^{m+1}-2} t_k] \prod_{k=1}^{2^{m+1}-2} \phi_2^{(m)}(t_k).$$

Let $D_m = \{t : \phi_1^{(m)}(t) \neq 0\}$. Then $D_m = \cup_j (2^{m+1}j - 1, 2^{m+1}j + 1)$ where j is an integer. Define on D_m, $\chi^{(m)}(t) = \phi_2^{(m)}(t)/\phi_1^{(m)}(t)$. It can be shown that $\chi^{(m)}(t) = (-1)^j$ for $t \in (2^{m+1}j - 1, 2^{m+1}j + 1)$. Elsewhere $\chi^{(m)}$ is undefined. Now $|t_k - 2^{m+1}j_k| < 1$ for some j_k, and $|\sum_{k=1}^{2^{m+1}-2} t_k - 2^{m+1}j| < 1$ for some j, imply that $\sum_{k=1}^{2^{m+1}-2} j_k = j$. So the equation

$$\text{(3B)} \qquad \chi^{(m)}(\textstyle\sum_{k=1}^{2^{m+1}-2} t_k) = \prod_{k=1}^{2^{m+1}-2} \chi^{(m)}(t_k)$$

holds whenever defined (namely where $t_k \in D_m$ for all k and $\sum_{k=1}^{2^{m+1}-2} t_k \in D_m$). From (3B) equation (3A) follows.

If P is a distribution with characteristic function $\phi_1^{(m)}(t)$ and Q has characteristic function $\phi_2^{(m)}(t)$, then $[P] \neq [Q]$. However (3A) implies that P, Q generate the same distribution on $(x_2 - x_1, \cdots, x_n - x_1)$ for all $n \leq 2^{m+1} - 1$. So $P^n T_n^{-1} = Q^n T_n^{-1}$ for all $n \leq 2^{m+1} - 1$.

Acknowledgments. The author would like to thank D. G. Kendall for his helpful comments. This research was partially supported by the Natural Sciences and Engineering Research Council Canada.

REFERENCES

[1] BONDESSON, L. (1974). Characterizations of probability laws through constant regression. *Z. Wahrsch. verw. Gebiete* **30** 93–115.

[2] BROWN, G. W. (1940). Reduction of a certain class of composite statistical hypotheses. *Ann. Math. Statist.* **11** 254-270.

[3] FRASER, D. A. S. (1968). *The Structure of Inference*. Wiley, New York.

[4] HOTELLING, H. (1936). Relations between two sets of variates. *Biometrika* **28** 321–377.

[5] KENDALL, D. G. (1984). Foundations of a theory of random shape. Unpublished manuscript.

[6] LOÈVE, M. (1955). *Probability Theory* I. Van Nostrand, New York. 201–205.

[7] KOVALENKO, I. N. (1958). On the reconstruction of an additive type of distribution based upon a sequence of independent trials. *Memoirs of the All-Union Conference on Probability Theory and Mathematical Statistics*, Erevan.

[8] LUKACS, E. (1970). *Characteristic Functions*. Griffin, London.

[9] PARTHASARATHY, K. R. (1967). *Probability Measures on Metric Spaces*. Academic, New York.

[10] PITMAN, E. J. G. (1938). The estimation of the location and scale parameters of a continuous population of any given form. *Biometrika* **30** 391–421.

[11] PITMAN, E. J. G. (1939). Tests of hypotheses concerning location and scale parameters. *Biometrika* **31** 200–215.

[12] PROKHOROV, YU. V. (1965). On a characterization of a class of probability distributions by distribution of some statistics. *Theor. Probability Appl.* **10** 438–445.

[13] SMALL, C. G. (1983). Characterization results for maximal invariant statistics. *Z. Wahrsch. verw. Gebiete*, to appear.

[14] ZINGER, A. A. and LINNIK, YU. V. (1964). On characterizations of the normal distribution. *Theor. Probability Appl.* **9** 624–626.

DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCE
UNIVERSITY OF WATERLOO
WATERLOO, ONTARIO
CANADA N2L 3G1