CHARACTERIZATION OF DISTRIBUTIONS BY THE EXPECTED VALUES OF THE ORDER STATISTICS*

J. S. HUANG

(Received Oct. 5, 1972; revised Aug. 26, 1974)

Introduction

Several parametric families of distributions have been characterized by the properties of their order statistics. See, for example, Ferguson [4] and David ([3], p. 19). An interesting non-parametric result is due to Chan [2] and Konheim [6]: Let $X_1, X_2, \cdots$ be independent random variables with common distribution function F. Let $X_{1,n} \leqq \cdots \leqq X_{n,n}$ be the order statistics of $X_1, \cdots, X_n$. If $\mathrm{E}\,|X_1|$ is finite then F is determined by the sequence $\{\mathrm{E}\,(X_{1,n}): n=1, 2, \cdots\}$ (and likewise, by $\{\mathrm{E}\,(X_{n,n}): n=1, 2, \cdots\}$). Wang [12] showed that for any $k \geqq 1$, the sequence $\{\mathrm{E}\,(X_{k,n}): n=k, k+1, \cdots\}$ (and likewise, $\{\mathrm{E}\,(X_{n-k+1,n}): n=k, k+1, \cdots\}$) determines F, also assuming the finiteness of $\mathrm{E}\,|X_1|$. In this paper we show that under lesser restrictions on the moments of X_1, the distribution F is characterizable by more general sequences.

1. Characterization by a tail sequence

Throughout this paper we shall let $\{k(n): n=1, 2, \cdots\}$ denote a sequence of integers with $1 \leqq k(n) \leqq n$. $\{k(n)\}$ is said to satisfy property $(A\text{-}m)$ if furthermore

$$k(m) \leqq k(n) \leqq k(m)+n-m \qquad \text{for all } n \geqq m . \tag{1}$$

We note that each $\{k(n)\}$ automatically satisfies property $(A\text{-}1)$.

LEMMA 1. *If* $\{k(n)\}$ *satisfies* $(A\text{-}m)$ *for some* m, *then*

$$\int_0^1 f(x) x^{k(n)-1}(1-x)^{n-k(n)} dx = 0 , \qquad n=m, m+1, \cdots \tag{2}$$

implies $f(x)=0$ *a.e.* $(0, 1)$.

PROOF. Letting $g(x)=f(x)x^{k(m)-1}(1-x)^{m-k(m)}$ and $P_i(x)=x^{k(m+i)-k(m)} \cdot (1-x)^{i-k(m+i)+k(m)}$, $i=0, 1, \cdots$, we may write (2) in the form

* Research partially supported by NRC of Canada grant number A8057.

(3) $$\int_0^1 g(x)P_i(x)dx=0\ , \qquad i=0,1,\cdots .$$

It follows from (*A*-*m*) that $P_i(x)$ is a polynomial (of degree i) and therefore (3) is equivalent to $\int_0^1 g(x)x^i dx=0$, $i=0,1,\cdots$. It is well known (see Sz.-Nagy [11], p. 331) that

(4) $$\int_0^1 g(x)x^i dx=0\ , \qquad i=0,1,\cdots \Rightarrow g(x)=0 \text{ a.e. } [0,1]\ .$$

Hence $f(x)=0$ a.e. [0, 1].

In other words, the lemma asserts the completeness of a certain subfamily of the beta distributions. Since $\mathrm{E}(X_{k,n})=\mathrm{E}[F^{-1}(Y)]$, where Y is the beta random variable with parameters $(k, n-k+1)$, F^{-1} is the inverse function of F (see Moriguti [7] and Hájek and Šidák [5]) defined by

$$F^{-1}(s)=\inf\{x: F(x)\geqq s\} \qquad 0<s<1\ ,$$

it leads to the following application.

THEOREM 1. *Let* $X_{1,n}\leqq\cdots\leqq X_{n,n}$ *be the order statistics from the distribution function* F. *If the sequence* $\{k(n)\}$ *satisfies the property* (*A*-*m*) *for some* m *and* $\mathrm{E}\,|X_{k(m),m}|<\infty$, *then* F *is uniquely determined by the sequence of numbers*

(5) $$\{\mathrm{E}(X_{k(n),n}):\ n=m, m+1,\cdots\}\ .$$

PROOF. We first show that each term in (5) is finite. By (*A*-*m*) it follows that for $n\geqq m$, $k(n)\geqq k(m)$ and $n-k(n)\geqq m-k(m)$. Thus

$$\begin{aligned}\mathrm{E}\,|X_{k(n),n}| &= C_n\int_0^1 |F^{-1}(s)|\,s^{k(n)-1}(1-s)^{n-k(n)}ds\\ &\leqq C_n\int_0^1 |F^{-1}(s)|\,s^{k(m)-1}(1-s)^{m-k(m)}ds\\ &= C_nC_m^{-1}\,\mathrm{E}\,|X_{k(m),m}|<\infty\ ,\end{aligned}$$

where $C_i=k(i)\binom{i}{k(i)}$. Now suppose G is a distribution function whose order statistics $Y_{k,n}$ satisfy

$$\mathrm{E}(X_{k(n),n})=\mathrm{E}(Y_{k(n),n})\ , \qquad n=m, m+1,\cdots .$$

It is equivalent to

$$\int_0^1 [F^{-1}(s)-G^{-1}(s)]s^{k(n)-1}(1-s)^{n-k(n)}ds=0\ , \qquad n=m, m+1,\cdots .$$

From Lemma 1 it follows that $F^{-1}=G^{-1}$ a.e. Finally, it follows from the left continuity of F^{-1} that $F=G$. This completes the proof.

We wish to point out that, in order to characterize a distribution F via the sequence (5) it is necessary that each term in (5) is finite. Clearly, there are infinitely many distributions having, say,

$$\mathrm{E}\,|X_{n,n}|=\infty\,, \qquad n=1,2,\cdots .$$

Thus the sequence $\{\mathrm{E}\,(X_{n,n})\}$ will not characterize, for example, the Cauchy distribution nor the Pareto distribution defined by

$$(6) \qquad F(x)=1-x^{-1}\,, \qquad x>1\,.$$

Of course, one notices that $\mathrm{E}\,(X_{1,2})$ is finite for the case of (6) and thus Theorem 1 is applicable with the choice $m=2$, $k(2)=1$.

It is also possible to characterize a distribution for which none of its order statistics has a finite mean. For example, let X be distributed by $F(x)=1-(\log x)^{-1}$, $x>e$. Stoops and Barr [10] showed that not only is $\mathrm{E}\,|X|^{\delta}=\infty$ for all $\delta>0$ but also $\mathrm{E}\,|X_{k,n}|^{\delta}=\infty$ for all (k, n) and all $\delta>0$. Consider the random variable $Y=\log(\log X)$. We see that Y is exponentially distributed and $\mathrm{E}\,(Y_{k,n})$ is finite for all k, n. Thus one can characterize the distribution of Y via $\{\mathrm{E}\,(Y_{k,n})\}$, which in turn characterizes the distribution of X.

In general, if $X\sim F$ and ϕ is a measurable 1-1 function on the support of F, to characterize X it is equivalent to characterize $Y=\phi(X)$. If furthermore ϕ is monotone increasing then $Y_{k,n}=(\phi(X))_{k,n}=\phi(X_{k,n})$ and thus $\mathrm{E}\,|\phi(X_{k(m),m})|<\infty \Rightarrow \mathrm{E}\,|\phi(X_{k(n),n})|<\infty$, $n\geqq m$ for those $\{k(n)\}$ satisfying $(A\text{-}m)$. For monotone decreasing ϕ the situation is analogous except for interchanging $\phi(X_{k,n})$ and $\phi(X_{n-k+1,n})$. A simple application of this idea leads to the following result:

COROLLARY 1. *Let r be a positive odd integer and let $\mathrm{E}\,|X_1|^{\delta}$ be finite for some $\delta>0$. If $\{k(n)\}$ satisfies $(A\text{-}m)$ for some $m\geqq 1$, and if $r\delta^{-1}\leqq k(m)\leqq m+1-r\delta^{-1}$, then F is determined by*

$$(7) \qquad \{\mathrm{E}\,(X^{r}_{k(n),n}):\ n=m, m+1,\cdots\}\,.$$

PROOF. Since $\phi(x)=x^{r}$ is a monotone increasing function there is no ambiguity about the meaning of (7). Clearly, (7) determines F as long as each term there is finite. It remains to show that this is implied by the finiteness of $\mathrm{E}\,|X_1|^{\delta}$. That this is true is the consequence of the following lemma which was proved by Sen [9] for absolutely continuous distribution F. His proof continues to hold for arbitrary distribution function F with appropriate modifications.

LEMMA 2 (Sen). *If* $\mathrm{E}\,|X_1|^\delta<\infty$ *for some* $\delta>0$ *then* $\mathrm{E}\,|X_{k,n}|^r<\infty$ *for all* n *and* k *satisfying* $r\delta^{-1}\leqq k\leqq n+1-r\delta^{-1}$.

COROLLARY 2. *Let* $\mathrm{E}\,|X_1|<\infty$. *If for some* m, *either* $\{k(n): n=m, m+1, \cdots\}$ *or* $\{n-k(n): n=m, m+1, \cdots\}$ *is non-decreasing, then* F *is determined by* (5).

PROOF. It suffices to show that $\{k(n)\}$ satisfies (A-m_0) for some m_0. Suppose $k(n)$ is non-decreasing. There exists $m_0\geqq m$ such that $m_0-k(m_0)\leqq n-k(n)$ for all $n\geqq m_0$. Thus (A-m_0) is satisfied. It thus follows from Theorem 1 that F is determined by the subset of (5):

$$\{\mathrm{E}\,(X_{k(n),n}): n=m_0, m_0+1, \cdots\}\ .$$

The case with non-decreasing $n-k(n)$ is analogous.

Wang's result (and, a fortiori, Chan's) is a special case of our next corollary, which is itself a special case of Corollary 2.

COROLLARY 3. *If* $\mathrm{E}\,|X_1|<\infty$ *then* F *is determined by* $\{\mathrm{E}\,(X_{k,n}): n=m, m+1, \cdots\}$ *for any* k *and any* $m\geqq k$ (*and, likewise, by* $\{\mathrm{E}\,(X_{n-k+1,n}): n=m, m+1, \cdots\}$).

Chan's result is a special case of our next corollary.

COROLLARY 4. *If* $\mathrm{E}\,|X_1|<\infty$ *then* F *is determined by* $\{\mathrm{E}\,(X_{k(n),n}): n=1, 2, \cdots\}$ *for any* $\{k(n)\}$.

PROOF. This is a special case ($m=1$) of Theorem 1. As was mentioned before, the constraint (A-m) in Theorem 1 is vacuous when $m=1$ since all $\{k(n)\}$ are supposed to satisfy $1\leqq k(n)\leqq n$.

2. Characterization by a subsequence

It seems natural to ask: Is F determined if more than finitely many terms are removed from the sequence $\{\mathrm{E}\,(X_{k(n),n}): n=1, 2, \cdots\}$? Under some conditions the answer is affirmative. For instance, by using (4) alone it can be seen that for any fixed positive integers k, a and b, the equally spaced subsequence $\{\mathrm{E}\,(X_{k,n}): n=a, a+b, a+2b, \cdots\}$ does determine F. This is a special case of our next theorem, whose proof depends on a generalization of the uniqueness theorem (4). The following lemma is proved in Boas ([1], Theorem 12.4.4, p. 235):

LEMMA 3 (Müntz). *Let* $f\in L_1(0, 1)$. *If*

$$\int_0^1 f(x)x^{n_i}dx=0\ , \qquad i=1, 2, \cdots, \tag{8}$$

where n_i *are distinct positive real numbers with* $\sum_{i=1}^{\infty} n_i^{-1}=\infty$ *then* $f(x)=0$ *a.e.* $(0, 1)$.

THEOREM 2. *Let* $\mathrm{E}\,|X_{k,n}|<\infty$ *for some* k *and* n *(say,* $=n_1\geqq k$*). Then* F *is uniquely determined by*

$$\{\mathrm{E}\,(X_{k,n}):\ n=n_1, n_2, \cdots\}\,, \tag{9}$$

for any sequence of integers $n_2, n_3, \cdots$ *such that* $n_1<n_2<n_3<\cdots$ *and* $\sum_{i=1}^{\infty} n_i^{-1}=\infty$ *(and, by* $\{\mathrm{E}\,(X_{n-k+1,n}):\ n=n_1, n_2, \cdots\}$ *provided* $\mathrm{E}\,|X_{n_1-k+1,n_1}|<\infty$*).*

PROOF. The proof is essentially the same as in Theorem 1. Instead of (4), we now have the stronger Lemma 3. It suffices to point out that $\sum_{i=2}^{\infty}(n_i-n_1)^{-1}$ diverges if and only if $\sum_{i=2}^{\infty} n_i^{-1}$ does. The parenthetical remark is proved by interchanging x and $1-x$.

We note that Corollary 3 is also a special case of Theorem 2.

The result of Lemma 3 appears self-strengthening in that if $k(n)$ keeps taking on a fixed value "often" enough, then

$$\int_0^1 f(x)x^{k(n)-1}(1-x)^{n-k(n)}dx=0\,, \qquad n=n_1, n_2, \cdots \tag{10}$$

implies $f=0$ a.e. For instance, let $\{k(n_i)\}$ be bounded and let $\sum_{i=1}^{\infty} n_i^{-1}=\infty$. It follows that there exists a subsequence $\{m_i\}\subset\{n_i\}$ such that $k(m_i)$ is constant (say, $=k^*$) and $\sum m_i^{-1}=\infty$. By (10) we have $\int g(x)x^n dx=0$, $n=m_1, m_2, \cdots$, where $g(x)\equiv f(1-x)x^{-k^*}(1-x)^{k^*-1}$. Thus we see by Lemma 3 that $g\equiv 0$, and so is f. Indeed, the boundedness of $\{k(n)\}$ is not necessary. It suffices to have $\{k(n)\}$ bounded "sufficiently often" in the sense that the sum of n_i^{-1} over $\{i\,|\,k(n_i)\leqq c$ or $n_i-k(n_i)\leqq c\}$ for some constant c diverges. In terms of application, this means that F is then characterizable by sequences of form

$$\{\mathrm{E}\,(X_{k(n),n}):\ n=n_1, n_2, \cdots\}\,. \tag{11}$$

Inasmuch as $k(n)$ need not be a constant function (or even bounded), (11) appears to be more general than (9).

This "strengthening", however, is illusory. It amounts to the detection of a subsequence of (11) to which Theorem 2 is applicable (and simply ignore the rest of the terms). In other words, the question is whether or not F is already determined by some subset of (11). Since (10) demands more than (8), it is thus anything but a strengthening.

COROLLARY 5. *Let* $\{\min\{k(n), n-k(n)\}\}$ *be bounded on* $\{n_i\}$ *with* $\sum n_i^{-1}=\infty$. *Then* F *is determined by* (11).

Remark 1. Without the constraint of $\min\{k(n), n-k(n)\}$ being bounded "sufficiently often", the sequence (11) will not characterize F. Take, for example, two symmetric distributions F and G with the same mean (say, $=0$). It is clear that the expected values of their medians are all identical, namely, $\mathrm{E}(X_{(n+1)/2,n})=\mathrm{E}(Y_{(n+1)/2,n})=0$ for $n=1, 3, 5, \cdots$. Thus $\{\mathrm{E}(X_{(n+1)/2,n}): n=1, 3, 5, \cdots\}$ does not characterize F.

Remark 2. We wish to conclude this paper with a question: Is Theorem 1 true without the extra constraint (A-m) on the sequence $\{k(n)\}$? Let $\{k(n)\,|\,n\geqq m\}$ be such that for each $m_0\geqq m$ the property (A-m_0) is violated. This means that there exists at least one value of n ($>m_0$) such that either $k(n)<k(m_0)$ or $n-k(n)<m_0-k(m_0)$. If there are "sufficiently many" such n's, F will be determined by virtue of Corollary 5. It is when there are, for each m_0, *some* but *not* "many enough" such n's that we do not know the answer. Our conjecture is that (A-m) is not really needed in Theorem 1.

Acknowledgement

I am indebted to Dr. P. Fischer, University of Guelph, for many helpful discussions throughout the development of this paper, and in particular, for communicating to me the existence and a proof to Lemma 3. I would also like to thank Professor Q. Rahman, University of Montreal, for calling my attention to Boas' works.

Addendum

After this paper was accepted for publication it came to our attention that Corollary 4 was obtained also by Pollak [8] via a different argument.

UNIVERSITY OF GUELPH

REFERENCES

[1] Boas, R. P., Jr. (1954). *Entire Functions*, Academic Press.
[2] Chan, L. K. (1967). On a characterization of distributions by expected values of extreme order statistics, *Amer. Math. Monthly*, **74**, 950-951.
[3] David, H. A. (1970). *Order Statistics*, Wiley, New York.
[4] Ferguson, T. S. (1967). On characterizing distributions by properties of order statistics, *Sankhyā* A, **29**, 265-278.
[5] Hájek, J. and Šidák, Z. (1967). *Theory of Rank Tests*, Academia, Prague.
[6] Konheim, A. G. (1971). A note on order statistics, *Amer. Math. Monthly*, **78**, 524.

[7] Moriguti, Sigeiti (1951). Extremal properties of extreme value distributions, *Ann. Math. Statist.*, **22**, 523-536.

[8] Pollak, Moshe (1973). On equal distributions, *Ann. Statist.*, **1**, 180-182.

[9] Sen, Pranab Kumar (1959). On the moments of the sample quantiles, *Calcutta Statist. Ass. Bull.*, **9**, 1-19.

[10] Stoops, Glenn and Barr, Donald (1971). Moments of certain Cauchy order statistics, *Amer. Statistician*, **25**, No. 5, 51.

[11] Sz.-Nagy, Bela (1965). *Introduction to Real Functions and Orthogonal Expansions*, Oxford Univ. Press, New York.

[12] Wang, Y. H. (1971). On a characterization of the distribution functions by the expected values of the order statistics, *Tech. Report*, No. 71-10, Division of Statistics, The Ohio State University.