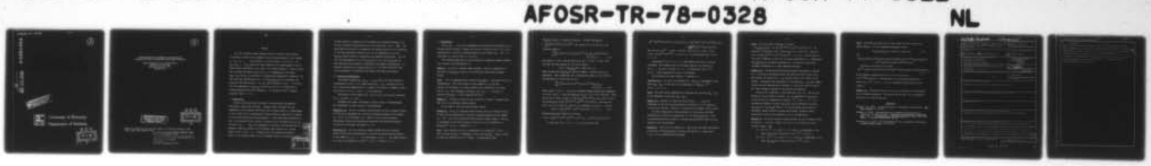
AD-A051 094 KENTUCKY UNIV LEXINGTON DEPT OF STATISTICS F/G 12/1
CHARACTERIZATIONS OF GEOMETRIC DISTRIBUTION AND DISCRETE IFR (D--ETC(U)
JAN 78 E EL-NEWEIHI, Z GOVINDARAJULU AFOSR-77-3322

UNCLASSIFIED AFOSR-TR-78-0328 NL

| OF |
AD
A051 094



END
DATE
FILMED
4-78
DDC

AFOSR-TR- 78-0328

(2)

University of Kentucky
Department of Statistics

(2)

CHARACTERIZATIONS OF GEOMETRIC DISTRIBUTION AND DISCRETE IFR (DFR) DISTRIBUTIONS USING ORDER STATISTICS

Emad El-Neweihi* and Z. Govindarajulu
University of Kentucky
Technical Report No. 119
January 1978

*Research sponsored by the Air Force Office of Scientific Research, AFSC, USAF, under Grant AFOSR-77-3322.

Abstract

Let X be a discrete random variable the set of possible values (finite or infinite) of which can be arranged as an increasing sequence of real numbers $a_1 < a_2 < a_3 < \dots$. In particular, a_i could be equal to i for all i. Let $X_{1n} \le X_{2n} \le \dots \le X_{nn}$ denote the order statistics in a random sample of size n drawn from the distribution of X, where n is a fixed integer ≥ 2. Then, we show that for some arbitrarily fixed $k(2 \le k \le n)$, independence of the event $\{X_{kn} = X_{1n}\}$ and X_{1n} is equivalent to X being either degenerate or geometric. We also show that the montonicity in i of $P\{X_{kn}=X_{1n}|X_{1n}=a_i\}$ is equivalent to X having the IFR (DFR) property. Let $a_i=i$ and $G(i)=P(X\ge i)$, i=1,2,.... We prove that the independence of $\{X_{2n}-X_{1n} \in B\}$ and X_{1n} for all i is equivalent to X being geometric, where $B=\{m\}$ $(B=\{m,m+1,\dots\})$, provided $G(i)=q^{i-1}$, $1\le i\le m+2$ $(1\le i\le m+1)$, where $0 < q < 1$.

1. Introduction.

Several contributions have been made to characterizing the geometric distribution using order statistics. Ferguson (1965) has shown that the independence of the smallest observation and the sample range in a random sample of size 2 drawn from a non-degenerate discrete population implies and is implied by the discrete distribution being geometric. If the underlying distribution is that of an unbounded lattice variate, Srivastava (1974) has shown that X_{1n} and the event $\{X_{1n} = \dots = X_{nn}\}$ are independent if and only if the distribution is geometric, where X_{in} denotes the ith smallest order statistic in a random sample of size n (i=1,...,n). Galambos (1975) has extended Srivastava's result to the situation where the set of possible values of the discrete random variable (finite or infinite) can be arranged in an increasing sequence (i.e. the set of possible values need not be of the form $\{\alpha+\beta i,\ i=1,2,\ldots,\beta\neq 0\}$. The main theme of our paper is to generalize the existing results in two directions: (i) For some arbitrarily fixed $k(2 \leq k \leq n)$ the independence of X_{1n} and $\{X_{kn}=X_{1n}\}$ should suffice to characterize the geometric distribution. (ii) For $a_i=i$, the independence of X_{1n} and $\{X_{2n}-X_{1n}=m\}$, or X_{1n} and $\{X_{2n}-X_{1n}\geq m\}$ for some fixed $m\geq 1$ should suffice to characterize the geometric distribution. In addition, monotonicity of $P(X_{kn}=X_{1n}|X_{1n}=a_i)$ in i for some arbitrarily fixed k can be employed to characterize the discrete IFR (DFR) distributions.

2. Notation and Definitions.

The random variables $X_{1n} \leq X_{2n} \leq \ldots \leq X_{nn}$ denote the order statistics corresponding to the n i.i.d. random variable $X_1,\ldots,X_n$. We denote by N the set of the natural numbers and by I a segment of N, where by a segment we mean that either $I = N$, or $I = \{i\varepsilon N:\ i \leq r\}$ for some $r\varepsilon N$.

The sequence of real numbers $\{a_i:\ i\varepsilon I\}$ is said to be strictly increasing if $a_i < a_j$ when $i < j$, $i,j\varepsilon I$.

Throughout this paper "increasing" is used in place of "nondecreasing" and "decreasing" is used in place of "nonincreasing".

Definition 2.1. Let X be a discrete random variable the set of possible values of which can be represented by a strictly increasing sequence of real numbers $\{a_i:\ i\varepsilon I\}$. Let $G(i)=P(X\geq a_i)$, $i\varepsilon I$. Then X is said to have increasing (decreasing) failure rate distribution (denoted by IFR (DFR) distribution), if $G(i+1)/G(i)$ decreases (increases) in $i \varepsilon I$.

Definition 2.2. Let X be a discrete random variable the set of possible values of which can be represented by a strictly increasing infinite sequence of real number $a_1 < a_2 < \ldots$. Let $G(i)=P(X\geq a_i)$, $i=1,2,\ldots$. The random variable X is said to be geometric if $G(i)=q^{i-1}$, $i=1,2,\ldots$, where $0 < q < 1$.

3. Main Results.

Let $X_1, X_2, \ldots, X_n$, $n \geq 2$ be independent and identically distributed (i.i.d.) discrete random variables. Assume that the set of possible values of X_1 can be represented by a strictly increasing sequence of real numbers $\{a_i: i \varepsilon I\}$. In particular, a_i could be equal to i for all i.

The following Lemma gives a characterization of degenerate random variables and is useful in proving Theorem 3.1.

Lemma 3.1. Let X_1 be a discrete random variable. Then X_1 is degenerate if and only if $P(X_{1n}=X_{kn})=1$, where k is an arbitrarily fixed postive integer $(2 \leq k \leq n)$.

Proof. If X_1 is degenerate then trivially $P(X_{1n}=X_{kn})=1$. Now assume that X_1 is non-degenerate. Then there exists two real numbers b_1, b_2 such that $P(X_1=b_1) > 0$ and $P(X_1=b_2) > 0$, where, without loss of generality, we assume that $b_1 < b_2$. Now, $P\{X_{1n} \neq X_{kn}\} \geq P\{X_{1n}=b_1, X_{2n} = X_{3n} = \ldots = X_{nn}=b_2\} > 0$, therefore $P\{X_{1n}=X_{kn}\} < 1$ which completes the proof.

Remark 3.1. It should be noted that the conclusion of Lemma 3.1 remains valid even if X_1 is an arbitrary random variable.

We are ready to state and prove the main results.

Theorem 3.1. Let X_1 be a discrete random variable the set of possible values of which can be represented by a strictly increasing sequence of real numbers $\{a_i: i \varepsilon I\}$. Let k be an arbitrarily fixed positive integer $(2 \leq k \leq n)$. Then X_{1n} is independent of the event $\{X_{1n}=X_{kn}\}$ if and only if X_1 is degenerate or $P(X_1 \geq a_i)=q^{i-1}$, $i=1,2,\ldots$, where $0 < q < 1$.

Proof. First observe that if X_1 is degenerate or if $P(X \geq a_i)=q^{i-1}$, $i=1,2,\ldots$, then in either case X_{1n} is independent of the event $\{X_{1n}=X_{kn}\}$. Next, in order to prove the converse, let $G(i) = P(X \geq a_i)$. By hypothesis we have

$P(X_{kn}=X_{1n}, X_{1n}=a_i) = P(X_{kn}=X_{1n})\ P(X_{1n}=a_i)$. Writing $P(X_{1n}=X_{kn}=a_i)$ $= \sum_{j=k}^{n} \binom{n}{j}[G(i)-G(i+1)]^j [G(i+1)]^{n-j}$, and setting $j'=n-j$ we are led to the following equation:

$$\sum_{j'=0}^{n-k} \binom{n}{j'}[G(i+1)]^{j'}[G(i)-G(i+1)]^{n-j'} = P(X_{1n}=X_{kn})[G^n(i)-G^n(i+1)],$$
$$\text{for all } i\varepsilon I. \qquad (3.1)$$

Now either $I = \{i\varepsilon N:\ i\leq r\}$ for some $r\varepsilon N$ or $I=N$. In case $I = \{i\varepsilon N:\ i\leq r\}$ for some $r\varepsilon N$, then setting $i=r$ in (3.1) we obtain

$$G^n(r) = P(X_{1n}=X_{kn})G^n(r) \quad \text{where } G(r) > 0.$$

Hence we must have $P(X_{1n}=X_{kn})=1$, which by Lemma 3.1 implies that X_1 is degenerate. Next, assume that $I=N$. Dividing both sides in (3.1) by $G^n(i)$ and letting $q(i) = G(i+1)/G(i)$ we have

$$\{\sum_{j=0}^{n-k} \binom{n}{j}[q(i)]^j[1-q(i)]^{n-j}\}(1-[q(i)]^n)^{-1} = P(X_{1n}=X_{kn}),$$
$$\text{for } i=1,2,\ldots. \qquad (3.2)$$

Notice that $0 < q(i) < 1$. Let Y_i be a binomial random variable with parameters $(n, q(i))$, $i=1,2,\ldots$, then the numerator of L.H.S. of (3.2) is $P(Y_i\leq n-k)$. Since $P(Y_i \leq n-k) = 1-P(Y_i \geq n-k+1) = k\binom{n}{k}\int_{q(i)}^{1} u^{n-k}(1-u)^{k-1}du$, the L.H.S. of (3.2) can be written as $\{k\binom{n}{k}\int_0^{1-q(i)} t^{k-1}(1-t)^{n-k}\,dt\}/(1-q^n(i))$. Now since the R.H.S. of (3.2) is free of i the L.H.S. is constant in $i=1,2,3,\ldots$. Now let

$$f(x) = \{k\binom{n}{k}\int_0^{1-x} t^{k-1}(1-t)^{n-k}\,dt\}/(1-x^n), \quad 0 < x < 1. \qquad (3.3)$$

Differentiating with respect to x we have

$$f'(x) = \{k\binom{n}{k}x^{n-k}[nx^{k-1}\int_0^{1-x} t^{k-1}(1-t)^{n-k}\,dt - (1-x)^{k-1}(1-x^n)]\}(1-x^n)^{-2}.$$

To show that $f'(x) < 0$, $0 < x < 1$, we first observe that

$$nx^{k-1}\int_0^{1-x} t^{k-1}(1-t)^{n-k}dt - (1-x)^{k-1}(1-x^n) \leq (1-x)^{k-1}[nx^{k-1}\int_0^{1-x}(1-t)^{n-k}dt-(1-x^n)]$$
$$= \frac{(1-x)^{k-1}}{n-k+1}[nx^{k-1}-(n-k+1)-(k-1)x^n].$$

Now, let $g(x) = nx^{k-1} - (n-k+1) - (k-1)x^n$. Since $g(0) < 0$, $g(1) = 0$ and $g'(x) = n(k-1)x^{k-2}(1-x^{n-k+1}) > 0$ for $0 < x < 1$ it follows that $g(x) < 0$ for $0 < x < 1$.

Consequently $f'(x) < 0$, $0 < x < 1$ which implies that $f(x)$ is strictly decreasing. This together with (3.2) implies that $q(i)$ is constant for $i=1,2,\ldots$. Let $q(i) = q$ where $0 < q < 1$. It follows that $G(i) = q^{i-1}$, $i=1,2,\ldots$, which completes the proof of the theorem.

The following is an easy corollary 2 Theorem 3.1:

Corollary 3.1.1. Let X_1 be as in Theorem 3.1. The X_{1n} is independent of the event $\{X_{kn} > X_{1n}\}$ if and only if X_1 is degenerate or $P(X_1 \geq a_i) = q^{i-1}$, $i=1,2,\ldots$, $0 < q < 1$.

Proof. The proof follows immediately by observing that the event $\{X_{kn} > X_{1n}\}$ is the complement of the event $\{X_{1n} = X_{kn}\}$.

Remark 3.1.1. Theorem 3.1 states that X_{1n} and $\{X_{1n} = \ldots = X_{kn}\}$ are independent if and only if X_1 has geometric distribution or X is degenerate. In particular, when k=n. Theorem 3.1 conincides with Galambos' (1975) result.

Our next theorem gives a characterization of the discrete IFR (DFR) distributions in terms of the montonicity in i of $P\{X_{1n}=X_{kn}|X_{1n}=a_i\}$. Such a characterization will be useful in constructing statistical tests for such classes of life distributions.

Theorem 3.2. Let X_1 be as in Theorem 3.1. Then X_1 has IFR (DFR) distribution if and only if $P\{X_{1n}=X_{kn}|X_{1n}=a_i\}$ increases (decreases) in i, where again $2 \leq k \leq n$ is an arbitrarily fixed integer.

Proof. As in the proof of Theorem 3.1 we have

$$P\{X_{1n}=X_{kn}|X_{1n}=a_i\} = \{k \binom{n}{k} \int_0^{1-q(i)} t^{k-1} (1-t)^{n-k} dt\}(1-q^n(i))^{-1}, \quad i\varepsilon I$$

where $q(i) = G(i+1)/G(i)$ $i\varepsilon I$. {Notice that $G(i) > 0$ for $i\varepsilon I$}. Again let $f(x) = \{k \binom{n}{k} \int_0^{1-x} t^{k-1} (1-t)^{n-k} dt\}(1-x^n)^{-1}$, $0 \leq x \leq 1$. We have shown in the proof of Theorem 3.1 that $f(x)$ is strictly decreasing in x. Consequently $P\{X_{1n}=X_{kn}|X_{1n}=a_i\}$ increases (decreases) in i if and only if $G(i+1)/G(i)$ decreases (increases) in i, which completes the proof.

Remark 3.2.1. One may give the following intuitive explanation of Theorem 3.2. If X_1 has an increasing failure rate then as the given value of X_{1n} gets larger, the values of $X_1,\ldots X_n$ are more likely to be "close" to one another. Consequently the probability of ties among $X_{1n},\ldots,X_{nn}$ gets higher. Similar intuitive explanations of Theorem 3.1 can be given that is based on the "lack of memory" property of the geometric distribution.

Let X_1 be as in Theorem 3.1, and assume that $a_i=i$, $i\varepsilon I$. Then for $k=2$, Theorem 3.1 can be stated as follows: X_{1n} is independent of $\{X_{2n}-X_{1n}=0\}$ if and only if X_1 is degenerate or $P(X_1 \geq i) = q^{i-1}$, $i=1,2,\ldots$, $0 < q < 1$. One might ask whether the event $\{X_{2n}-X_{1n}=0\}$ can be replaced by the event $\{X_{2n}-X_{1n}=m\}$ or $\{X_{2n}-X_{1n} \geq m\}$ where $m > 0$? The following theorem gives an affirmative answer provided we assume some boundary conditions (which automatically rule out the possibility of X_1 being degenerate).

Theorem 3.3. Let X_1 be a discrete random variable the set of possible values of which is I. Let $G(i) = P(X \geq i)$, $i\varepsilon I$, and $m \geq 1$ be arbitrarily fixed positive integer. Then

(i) $G(i) = q^{i-1}$ $1 \leq i \leq m+2$ $0 < q < 1$ and X_{1n} is independent of the event $\{X_{2n}-X_{1n}=m\}$ if and only if $G(i) = q^{i-1}$, $i=1,2,3,\ldots$.

(ii) $G(i) = q^{i-1}$, $i \leq 1 \leq m+1$, $0 < q < 1$, and X_{1n} is independent of the event $\{X_{2n}-X_{1n} \geq m\}$ if and only if $G(i) = q^{i-1}$, $i=1,2,\ldots$.

Proof. We provide the proof for (ii) only, since (i) can be proved in a similar fashion. By the independence assumption we have

$$P(X_{2n}-X_{1n}\geq m|X_{1n}=i) \text{ is free of } i, \text{ where } i\varepsilon I. \tag{3.4}$$

Now

$$P(X_{2n}-X_{1n}\geq m|X_{1n}=i)=[P(X_{2n}\geq m+i,X_{1n}\geq i)-P(X_{2n}\geq m+i,X_{1n}\geq i+1)]/[P(X_{1n}\geq i-P(X_{1n}\geq i+1)]$$
$$= (nG^{n-1}(i+m)[G(i)-G(i+1)])/(G^{n}(i)-G^{n}(i+1)).$$

Setting i=1 and using (3.4) we have

$$(nG^{n-1}(1+m)[G(1)-G(2)])/(G^{n}(1)-G^{n}(2))=(nG^{n-1}(i+m)[G(i)-G(i+1)])/(G^{n}(i)-G^{n}(i+1)) \tag{3.5}$$

By the boundary conditions the L.H.S. of (3.5) is equal to $(n\, q^{(n-1)m}[1-q])/(1-q^{n})$. Substituting in (3.5) and using induction we obtain $G(i) = q^{i-1}$, $i=1,2,\ldots$, i.e. X_1 is geometric and the proof is now complete .

Remark 3.3.1. Notice that results (i) and (ii) in Theorem 3.3 have different sets of boundary conditions. Also notice that for m=1, (ii) is subsumed by Corollary 3.1.1 with k=2 and a_i=i.

References

Ferguson, T.S. (1965). A characterization of the geometric distribution. Amer. Math. Monthly 72, 256-260.

Galambos, J. (1975). Characterizations of probability distributions by properties of order statistics. Statistical Distributions in Scientific Work. Vol. 2: Characterizations and Applications (ed. Patil, Kotz, ord). Boston: D. Reidel Publishing Co. 89-101.

Srivastava, R.C. (1974). Two characterizations of the geometric distribution. J. Amer. Statist. Assoc. 69, 267-269.

(18) AFOSR-TR- 78-0328 (19) TR-78-0328

4. TITLE (and Subtitle) CHARACTERIZATIONS OF GEOMETRIC DISTRIBUTION AND DISCRETE IFR (DFR) DISTRIBUTIONS USING ORDER STATISTICS (6)	5. TYPE OF REPORT & PERIOD COVERED (9) Interim rept.,
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) (10) Emad El-Neweihi and Z. Govindarajulu	8. CONTRACT OR GRANT NUMBER(s) (15) AFOSR 77-3322
9. PERFORMING ORGANIZATION NAME AND ADDRESS University of Kentucky Department of Statistics Lexington, Kentucky 40506	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2304/A5 (16) (17)
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research (NM) Bolling AFB Washington, D.C. 20332	12. REPORT DATE (11) January 1978
	13. NUMBER OF PAGES 8 (12) 11p.
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report) UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release: distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Characterization, Order Statistics, Geometric, IFR (DFR) Distribution

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Let X be a discrete random variable the set of possible values (finite or infinite) of which can be arranged as an increasing sequence of real numbers $a_1 < a_2 < a_3 < \ldots$. In particular, a_i could be equal to i for all i. Let $X_{1n} \le X_{2n} \le \ldots \le X_{nn}$ denote the order statistics in a random sample of size n drawn from the distribution of X, where n is a fixed integer ≥ 2. Then, we show it is shown

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

408 091

20. Abstract

that for some arbitrarily fixed $k(2 \leq k \leq n)$, independence of the event $\{X_{kn} = X_{1n}\}$ and X_{1n} is equivalent to X being either degenerate or geometric. We also show that the monotonicity in i of $P\{X_{kn}=X_{1n}|X_{1n}=a_i\}$ is equivalent to X having the IFR (DFR) property. Let $a_i=i$ and $G(i)=P(X\geq i)$, $i=1,2,\ldots$. We prove that the independence of $\{X_{2n}-X_{1n} \in B\}$ and X_{1n} for all i is equivalent to X being geometric, where $B=\{m\}$ $(B=\{m,m+1,\ldots\})$, provided $G(i)=q^{i-1}$, $1\leq i\leq m+2$ $(1\leq i\leq m+1)$, where $0 < q < 1$.