Reprinted from Sankhyā : The Indian Journal of Statistics, Series A, Vol. 29, Part 3, 1967

ON CHARACTERIZING DISTRIBUTIONS BY PROPERTIES OF ORDER STATISTICS*

By THOMAS S. FERGUSON

University of California, Los Angeles and *Berkeley*

SUMMARY. In Section 2, those continuous distributions F are found for which the regression of $X_{(m)}$ on $X_{(m+1)}$ is linear, when $X_{(1)}, \ldots, X_{(n)}$ are the order statistics of a sample of size n from F, and $1 \leqslant m \leqslant n$. In Section 3, those discrete non-degenerate distributions of independent X, Y are found for which min (X, Y) and $|X-Y|$ are independent.

1. Introduction

An interesting property of the negative exponential distribution with location parameter α and scale parameter $\sigma > 0$, having density

$$f(x \,|\, \alpha, \sigma) = \begin{cases} \sigma^{-1} e^{-(x-\alpha)/\sigma} & \text{for} \quad x > \alpha \\ 0 & \text{for} \quad x < \alpha \end{cases} \qquad \ldots \quad (1)$$

is as follows (see Malmquist, 1950; Epstein and Sobel, 1954).

If $X_1, X_2, \ldots, X_n$ is a sample from the negative exponential distribution and if $Y_1 \leqslant Y_2 \leqslant \ldots \leqslant Y_n$ are the order statistics, then $Y_1, Y_2-Y_1, Y_3-Y_2, \ldots, Y_n-Y_{n-1}$ are mutually independent.

Various converses to this statement exist. Fisz (1958) proved the following theorem, a converse when $n = 2$. *If X_1 and X_2 are independent identically distributed with an absolutely continuous distribution, and if Y_1 and Y_2-Y_1 are independent, where $Y_1 = min\,(X_1, X_2)$ and $Y_2 = max\,(X_1, X_2)$, then X_1 and X_2 have a negative exponential distribution.* Further generalizations may be found in Rossberg (1966).

An extension of Fisz's theorem to arbitrary n has been made by Tanis (1964) as follows. *If $Y_1, \ldots, Y_n$ are the order statistics of a sample of size n from an absolutely continuous distribution, and if Y_1 and $\sum_{i=2}^{n} (Y_i-Y_1)$ are independent, then the distribution is negative exponential.*

Fisz's theorem has been strengthened and extended to arbitrary n by Rogers (1963) : *If $Y_1, \ldots, Y_n$ are the order statistics of a sample of size n from an absolutely continuous distribution, and if for some $m = 2, 3, \ldots, n$, the regression of $Y_{m+1}-Y_m$ on Y_m is constant, then the distribution is negative exponential.*

---

*The preparation of this paper was sponsored in part by N.S.F. Grant GP-5224.

Two comments are in order. First, Fisz and Rogers state their results for the exponentials of the negatives of the variables occurring here, assuming their variables are positive, using ratios instead of differences, and arriving at a characterization of the distribution with density

$$g(z|\beta, \theta) = \begin{cases} \theta\beta^{-\theta}z^{\theta-1} & \text{for } 0 < z < \beta \\ 0 & \text{otherwise} \end{cases} \quad \dots \quad (2)$$

where $\beta = e^{-\alpha}$, $\theta = \sigma^{-1}$, and $z = e^{-x}$. We have taken the negatives of the logarithms of the variables occurring in their papers to make comparisons easier and because it seems to be a more natural setting. Second, Rogers states that he assumes the independence of $Y_{m+1}-Y_m$ and Y_m, whereas in the proof he uses only the constancy of the regression of $Y_{m+1}-Y_m$ on Y_m, so that it is the above that we refer to as Roger's theorem.

In Section 2, we prove a theorem which generalizes Roger's theorem in two directions. First, the assumption of absolute continuity is replaced by the assumption of continuity. Second, instead of assuming the regression of $Y_{m+1}-Y_m$ on Y_m to be constant, we assume that the regression of Y_{m+1} on Y_m is linear. This second generalization admits two additional families of distributions in the conclusion of the theorem. One is a Beta type (as is the density (2)) with density

$$f(x|\alpha, \beta, \theta) = \begin{cases} \theta\beta^{-\theta}(\alpha-x)^{\theta-1} & \text{for } \alpha-\beta < x < \alpha \\ 0 & \text{otherwise} \end{cases} \quad \dots \quad (3)$$

where $\theta > 0$ and $\beta > 0$. The other is a reciprocal Beta type with density

$$f(x|\alpha, \beta, \theta) = \begin{cases} \theta\beta^{\theta}(x-\alpha)^{-(\theta+1)} & \text{for } x > \alpha+\beta \\ 0 & \text{for } x < \alpha+\beta \end{cases} \quad \dots \quad (4)$$

where $\theta > 0$ and $\beta > 0$. In both densities, α and β are location and scale parameters.

The first result along these lines appears to have been by Bartoo (1952) who treats this problem among others in his Ph.D. thesis. For his result, Bartoo assumes absolute continuity and that for fixed r and s the regression of Y_{n-r} on Y_s is linear for all sample sizes n $> r+s$ (whereas $n = r+s+1$ is all that is required). Furthermore, he assumes that the distribution is bounded and so arrives only at distributions (3), although he points out that (4) is also a possibility.

A result closely related to Fisz's theorem, dropping the assumption of identical distributions and absolute continuity, but requiring a slightly stronger independence assumption is due to Ferguson (1964, 1965) and Crawford (1966) : *If X_1 and X_2 are independent non-degenerate random variables, and if min (X_1, X_2) and X_1-X_2 are independent, then either* (i) *both X_1 and X_2 have negative exponential distributions with common location parameter but possibly different scale parameters, or* (ii) *both X_1 and X_2 have geometric distributions with common location-scale parameters but possibly different* *geometric parameters.* The geometric distribution is the discrete distribution with probability mass function

$$f(x\,|\,\alpha, \beta, r) = (1-r)r^{(x-\alpha)/\beta} \qquad x = \alpha, \alpha+\beta, \alpha+2\beta, \ldots, \qquad \ldots \quad (5)$$

where α and $\beta > 0$ are location and scale parameters, and $0 < r < 1$ is referred to as the geometric parameter.

In Section 3, this theorem is generalized in the discrete case to the weaker independence assumptions found in Fisz's theorem. Several families of distributions, in addition to the geometric distributions, appear. This study allows one to make a reasonable conjecture for the continuous case as well : that if X_1 and X_2 are independent with continuous distributions, and if min (X_1, X_2) and $|X_1 - X_2|$ are independent, then either (i) both X_1 and X_2 have negative exponential distributions with common location parameters but possibly different scale parameters, or (ii) X_1 and X_2 have densities of the form

$$f_{X_1}(x) = \frac{e^{-(x-\alpha)/\beta}}{\beta(1+\theta e^{-\gamma})} \cdot \begin{cases} 1 & \text{for} \quad \alpha < x < \alpha+\beta\gamma \\ 1+\theta & \text{for} \quad x > \alpha+\beta\gamma \end{cases}$$

$$\ldots \quad (6)$$

$$f_{X_2}(x) = \frac{e^{-(x-\alpha)/\beta}}{\beta(1-\theta e^{-\gamma})} \cdot \begin{cases} 1 & \text{for} \quad \alpha < x < \alpha+\beta\gamma \\ 1-\theta & \text{for} \quad x > \alpha+\beta\gamma \end{cases}$$

where $\beta > 0$, $-1 \leqslant \theta \leqslant +1$ and $\gamma > 0$. The distributions (6) are limits of the distributions of part (ii) of Theorem 2. The author has been unable to prove or disprove this conjecture. It is easy to check, however, that if X_1 and X_2 are independent and have the distributions under (6), then min (X_1, X_2) and $|X_1 - X_2|$ are independent.

Results related to those found in this paper have been obtained by Bolger and Harkness (1965), Govindarajulu (1966), Lukacs (1965), and Sethuraman (1965).

2. Linear regression of adjacent order statistics

In this section we assume that $Y_1, Y_2, \ldots, Y_n$ are the order statistics of a sample of size n from a distribution with continuous distribution function F. To simplify the notation used in the proofs of the following results, we deal with $E(Y_m\,|\,Y_{m+1})$ rather than $E(Y_{m+1}\,|\,Y_m)$. As a consequence of the assumed linearity of this regression, we obtain the distributions of the negatives of the variables whose densities are given in (1), (3), and (4). The method used works for rather arbitrary (non-linear) regression functions. However, it fails when the order statistics are not adjacent. It is unknown what new distributions arise if any when dealing with the regression of Y_m on Y_{m+2}.

Recall that a regression function $E(Y_m\,|\,Y_{m+1} = y)$ is determined only up to an equivalence, that is to say, it is only determined almost surely (a.s.) with respect to the distribution of Y_{m+1}. One possible and natural choice of this function is as

$R(y)$ below, the expectation of the maximum of a sample of size m from the distribution of F truncated at y.

$$R(y) = \begin{cases} \int_{-\infty}^{y} xdF(x)^m / F(y)^m & \text{if} \quad F(y) \neq 0 \\ y & \text{if} \quad F(y) = 0. \end{cases} \qquad \dots \quad (7)$$

It is easily checked that $R(y)$ satisfies the definition as a possible choice of $E(Y_m \mid Y_{m+1} = y)$. The following lemma gives the important properties of $R(y)$.

Lemma : *If F is continuous and if $R(y)$ defined by (7) exists, then $R(y)$ is continuous and non-decreasing. Furthermore, if $F(y_1) < F(y_2)$, then $R(y_1) < R(y_2)$.*

Proof : Since $F(y)$ and $\int_{-\infty}^{y} xdF(x)^m$ are continuous, it is clear that $R(y)$ is continuous at all points except perhaps at $c = \inf\{y : F(y) > 0\}$. But since for $y > c$, $R(y)$ is the expectation of a variable taking values between c and y, $R(y)$ must converge to c as y tends to c from above.

Clearly, $R(y)$ is increasing for $y \leqslant c$, and $R(y) > R(c)$ if $y > c$. Suppose that $c < y_1 < y_2$. Then

$$\begin{aligned} &[R(y_2) - R(y_1)]\, F(y_2)^m F(y_1)^m \\ &\qquad = F(y_1)^m \int_{-\infty}^{y_2} xdF(x)^m - F(y_2)^m \int_{-\infty}^{y_1} xdF(x)^m, \\ &\qquad = F(y_1)^m \int_{y_1}^{y_2} xdF(x)^m - [F(y_2)^m - F(y_1)^m] \int_{-\infty}^{y_1} xdF(x)^m \\ &\qquad \geqslant F(y_1)^m y_1 [F(y_2)^m - F(y_1)^m] - [F(y_2)^m - F(y_1)^m] y_1 F(y_1)^m = 0, \end{aligned}$$

with strict inequality if $F(y_1) < F(y_2)$, completing the proof.

One of the uses we make of this lemma is to note that if $E(Y_m \mid Y_{m+1} = y)$ is linear, say equal to $ay - b$ a.s., then $a > 0$.

Theorem 1 : *If $Y_1, Y_2, \dots, Y_n$ are the order statistics of a sample of size n from a distribution with continuous distribution function F, and if for some positive integer m less than n, $E(Y_m \mid Y_{m+1} = y) = ay - b$ a.s. for some numbers a and b, then the distribution function is except for change of location and scale*

(i) $F(x) = e^x$ *for* $x < 0$, *if* $a = 1$

(ii) $F(x) = x^\theta$ *for* $0 < x < 1$, *if* $0 < a < 1$

(iii) $F(x) = (-x)^\theta$ *for* $x < -1$, *if* $a > 1$

where $\theta = a/(m(1-a))$.

Proof : Since one way of writing the regression $E(Y_m \mid Y_{m+1} = y)$ is (7), we have

$$\int_{-\infty}^{y} xdF(x)^m = (ay - b)\, F(y)^m \quad \text{a.s.} \quad (F). \qquad \dots \quad (8)$$

Find numbers c and d, possibly infinite, such that $\{x : 0 < F(x) < 1\}$ is the interval (c, d). There does not exist a subinterval (c_1, d_1), $c < c_1 < d_1 < d$, over which F is constant since the left side of (8) is constant in such an interval and the right side is increasing, while both sides are continuous, so that they could not possibly be equal at the next points of increase of F. Thus (8) is valid for all $y \in (c, d)$. The left side may be written

$$\int_{-\infty}^{y} x dF(x)^m = yF(y)^m - \int_{-\infty}^{y} F(x)^m dx. \quad \ldots \quad (9)$$

The existence of the integral on the left implies the existence of the integral on the right. Let $H(y) = \int_{-\infty}^{y} F(x)^m dx$; then $H'(y)$ exists for all y and is equal to $F(y)^m$. Equation (8) may be rewritten as

$$\frac{d}{dy} \log H(y) = ((1-a)y+b)^{-1} \quad \text{for all } y \in (c, d). \quad \ldots \quad (10)$$

We may solve this differential equation separately for the three cases mentioned in the theorem.

Case (i) : $a = 1$. Since $\log H(y)$ is increasing in (c, d), b must be positive. By integrating (10) we find $H(y) = Ke^{y/b}$, and by differentiating, we find $F(y)^m = Kb^{-1} e^{y/b}$. Clearly $c = -\infty$ and $d < \infty$; and since $F(d) = 1$, $F(y) = e^{(y-d)/(mb)}$ for $y < d$. This is the distribution under (i) with change of location and scale.

Case (ii) : $0 < a \lessdot 1$. Integrating (10), we find $H(y) = K((1-a)y+b)^{1/(1-a)}$ for $y \in (c, d)$. In this case, c and d are finite and the restrictions $F(c) = 0$ and $F(d) = 1$ give

$$F(y) = \left(\frac{y-c}{d-c}\right)^{\theta} \quad \text{for } c < y < d,$$

where $\theta = a/((1-a)m)$ and $c = -b/(1-a)$.

Case (iii) : $a > 1$. Integrating (10) we find $H(y) = K(b-(a-1)y)^{-1/(a-1)}$ for $y \in (c, d)$. In this case $c = -\infty$ and $d < \infty$. The restriction $F(d) = 1$ gives

$$F(y) = \left(\frac{\gamma-y}{\gamma-d}\right)^{\theta} \quad \text{for } y < d$$

where $\theta = -a/((a-1)m) < 0$, and $\gamma = b/(a-1) > d$. This completes the proof.

3. Independence of min (X, Y) and $|X-Y|$ in the discrete case

The following theorem may be considered in two ways—as a generalization of the Ferguson-Crawford Theorem in the discrete case by weakening the hypotheses of independence, or as an extension of Fisz's theorem to the discrete case dropping the assumption of identical distributions.

We seek all possible pairs of distributions of independent discrete random variables X and Y for which $U = \min(X, Y)$ and $W = |X-Y|$ are independent. If either X or Y is degenerate, the problem is rather trivial. One can easily show that if X, for example, is degenerate, then in order that U and W be independent it is necessary and sufficient that U or W be degenerate. Thus, we restrict attention to non-degenerate X and Y.

In this theorem, a simultaneous change of location and scale means of change of (X, Y) into $(aX+b, aY+b)$, where $a > 0$.

Theorem 2 : *Let X and Y be independent, discrete, non-degenerate random variables. Then, $U = \min(X, Y)$ and $W = |X-Y|$ are independent if, and only if, the distributions of X and Y may, by a simultaneous change of location and scale, be put into one of the following four forms.*

(i) *For some* $0 < r_1 < 1$ *and* $0 < r_2 < 1$,

$$P\{X = k\} = (1-r_1)r_1^k \qquad \textit{for} \quad k = 0, 1, \ldots$$

$$P\{Y = k\} = (1-r_2)r_2^k \qquad \textit{for} \quad k = 0, 1, \ldots$$

(ii) *For some integer* $n \geqslant 1$, $0 < r < 1$ *and* $-1 \leqslant \theta \leqslant +1$,

$$P\{X = k\} = \frac{(1-r)r^k}{(1+\theta r^n)} \cdot \begin{cases} 1 & \textit{for} \quad k = 0, 1, \ldots, n-1 \\ 1+\theta & \textit{for} \quad k = n, n+1, \ldots \end{cases}$$

$$P\{Y = k\} = \frac{(1-r)r^k}{(1-\theta r^n)} \cdot \begin{cases} 1 & \textit{for} \quad k = 0, 1, \ldots, n-1 \\ 1-\theta & \textit{for} \quad k = n, n+1, \ldots \end{cases}$$

(iii) *For some* $0 < r < 1$ *and* $0 < \theta < \infty$, *either*

$$P\{X = k\} = \frac{(1-r)r^k}{(1+\theta r^2)} \cdot \begin{cases} 1 & \textit{for} \quad k = 0, 1 \\ 1+\theta & \textit{for} \quad k = 2, 3, \ldots \end{cases}$$

$$P\{Y = k\} = \begin{cases} (1+\theta r)^{-1} & \textit{for} \quad k = 0 \\ 1-(1+\theta r)^{-1} & \textit{for} \quad k = 1 \end{cases}$$

or the same with X and Y interchanged.

(iv) *For some* $0 < r < 1$ *and* $0 \leqslant \theta \leqslant \infty$,

$$P\{X = k\} = (1-r^2)r^k \cdot \begin{cases} (1+\theta r)^{-1} & \textit{for} \quad k = 0, 2, 4, \ldots \\ \theta(1+\theta r)^{-1} & \textit{for} \quad k = 1, 3, 5, \ldots \end{cases}$$

$$P\{Y = k\} = (1-r^2)r^k \cdot \begin{cases} \theta(\theta+r)^{-1} & \textit{for} \quad k = 0, 2, 4, \ldots \\ (\theta+r)^{-1} & \textit{for} \quad k = 1, 3, 5, \ldots \end{cases}$$

Remarks : Each of the four forms above is a two-parameter family of distributions (in addition to location and scale parameters). The distributions under (i) are the geometric distributions of equation (5). In (ii), if equality holds in the inequality involving θ, then the distribution of X or of Y is concentrated on exactly n

points. In (iii) the distribution of Y (or X) is concentrated at two points. Although this can occur in (ii) with $n = 2$, (iii) is not a special case, since $P\{Y = 1\}$ is not necessarily equal to $rP\{Y = 0\}$. In (iv), the values of the formulae at $\theta = +\infty$ are to be interpreted as the values of the limits as $\theta \to +\infty$. The distribution of Y is obtained from the distribution of X by replacing θ by θ^{-1}. If $\theta = 0$ or ∞, then the probability of successive integers is zero alternately for X and Y.

Proof : We consider two main cases, and in the second case we consider five subcases.

Case I : $P(W = 0) = 0$. This states that no possible value of X is a possible value of Y. Since X and Y are non-degenerate, there exist either two possible values of X less than some possible value of Y or two possible values of Y less than some possible value of X. We shall suppose that the former is true, since the distributions in the latter case may be obtained by symmetry. Thus, we suppose that $x_1 < x_2$ are such that $P(X = x_1) > 0$, $P(X = x_2) > 0$, and $P(Y > x_2) > 0$. Then, for all $w > 0$,

$$P\{W = w\} = P\{W = w \mid U = x_i\} = P\{Y = x_i + w \mid X = x_i,\ Y > x_i\}$$

$$= P\{Y = x_i + w \mid Y > x_i\} = \frac{P\{Y = x_i + w\}}{P\{Y > x_i\}} \quad \dots \quad (11)$$

for $i = 1, 2$. Combining these two equations gives

$$P\{Y = x_2 + w\} = r_1 P\{Y = x_1 + w\} \quad \dots \quad (12)$$

where

$$r_1 = P\{Y > x_2\}/P\{Y > x_1\}. \quad \dots \quad (13)$$

Now let y be any possible value of Y, $y > x_1$, and put $w = y - x_1$ in equation (12).

$$P\{Y = (x_2 - x_1) + y\} = r_1 P\{Y = y\}.$$

Thus, inductively, $k(x_2 - x_1) + y$ are possible values of Y for $k = 1, 2, \dots$, and

$$P\{Y = k(x_2 - x_1) + y\} = r_1^k\, P\{Y = y\}. \quad \dots \quad (14)$$

Since the sum of these probabilities cannot be greater than one, $r_1 < 1$, or from (13), $P\{Y > x_1\} > P\{Y > x_2\}$. This says that between any two possible values of X less than some value of Y, there must exist some possible value of Y. Formula (14) also shows that for any possible value, y_2, of Y, $y_2 > x_2$, there exists a possible value y_1 of Y, $x_1 < y_1 < x_2$, for which $P(Y = y_1) > P(Y = y_2)$. This immediately shows that there are at most a finite number of possible values of X less than x_2, since if there were an infinite number, there would also have to be an infinite number of possible values of Y, all with a larger probability than that of y_2, clearly an impossibility.

The following equation may be derived analogously to formula (11),

$$P\{X = y_1 + w\} = P\{W = w\}\, P\{X > y_1\}, \quad \dots \quad (15)$$

where y_1 is any possible value of Y less than x_2. Now, formula (14) shows that the distribution of Y, and hence of W, is not bounded above, so that by (15) the distribution of X is not bounded above. Thus, there are two possible values of Y less than some possible value of X, which implies that the preceding analysis may be carried out symmetrically with respect to X and Y : that between any two possible values of Y there is a possible value of X, that there is only a finite number of possible values of Y less than a given number, and that, analogously to formula (14), for $y_1 < y_2$,

$$P\{X = k(y_2-y_1)+x\} = r_2^k P\{X = x\}. \quad \dots \quad (16)$$

We may suppose in formulas (14) and (16), that x_1 and x_2 are the smallest two possible values of X and that y_1 and y_2 are the smallest two possible values of Y.

Suppose, first, that $y_1 < x_1$. Then we know that $w_1 = x_1-y_1$ is a possible value of W, and that there is no possible value of W smaller than w_1, since from formula (15) X would then have a possible value smaller than x_1. Now, formulas (11) and (15) imply that $y_2-x_1 = w_1$, and, inductively, that the only possible values of X are $x_{k+1}=2kw_1+x_1$ for $k=0, 1, 2, \ldots$, and the only possible values of Y are $y_{k+1}=2kw_1+y_1$, for $k = 0, 1, 2, \ldots$. Formulas (14) and (16) imply that X and Y both have geometric distributions on these equally spaced and alternating possible values, and thus that $P\{X = x_1\} = 1-r_2$ and $P\{Y = y_1\} = 1-r_1$. Formula (15) with $w = w_1$ becomes

$$1-r_2 = P\{X = x_1\} = P\{W = w_1\} = P\{|X-Y| = w_1\}$$

$$= \sum_{j=1}^{\infty} P\{X = x_j\}P\{Y = y_j\}+\sum_{j=1}^{\infty} P\{X = x_j\}\, P\{Y = y_{j+1}\}$$

$$= \frac{(1-r_2)(1-r_1)(1+r_1)}{1-r_1r_2}.$$

This implies that $r_1 = r_2$. With a change of location and scale which puts $y_1 = 0$ and $x_1 = 1$, these distributions are found in the statement of Theorem 2 under (iv) with $\theta = \infty$. By symmetry, the distributions of X and Y when x_1 is assumed less than y_1 is found under (iv) with $\theta = 0$.

The fact that if X and Y are given these distributions, then U and W are independent, is easily checked directly and the details are omitted.

Case II : $P\{W = 0\} \neq 0$. If $P(X = x)$ and $P\{Y > x\}$ are positive, then $P\{Y = x\}$ must be positive, since, if not, then $0 < P\{W = 0\} = P\{W = 0 \,|\, U = x\}$ $= P\{X = Y \,|\, X = x,\, Y > x\} = 0$, which is a contradiction. Thus, every possible value of X ess than some possible value of Y is a possible value of Y, and, symmetrically, every possible value of Y less than some possible value of X is a possible value of X. This and non-degeneracy imply that there are at least two distinct numbers $x_0 < x_1$ which are possible values simultaneously for X and Y.

The independence of U and W, and that of X and Y imply that

$$P\{U=u\}\,P\{W=w\}=P\{\min(X,Y)=u,\ |X-Y|=w\}$$

$$=\begin{cases} P\{X=u\}\,P\{Y=u\} & \text{if } w=0 \\ P\{X=u\}\,P\{Y=u+w\}+P\{X=u+w\}P\{Y=u\} & \text{if } w>0, \end{cases} \quad \dots \quad (17)$$

for all u. Since $P\{W=0\}\neq 0$, we may solve for $P\{U=u\}$, substitute the solution back into equations (17), and arrive at the equation

$$P\{X=u\}\,P\{Y=u\}\,\frac{P\{W=w\}}{P\{W=0\}}=P\{X=u\}\,P\{Y=u+w\}+P\{X=u+w\}\,P\{Y=u\}, \quad \dots \quad (18)$$

for all u and all $w>0$. Letting u assume alternately the values x_0 and x_1 will yield two equations valid for all $w>0$.

$$\frac{P\{X=x_0+w\}}{P\{X=x_0\}}+\frac{P\{Y=x_0+w\}}{P\{Y=x_0\}}=\frac{P\{W=w\}}{P\{W=0\}}=\frac{P\{X=x_1+w\}}{P\{X=x_1\}}+\frac{P\{Y=x_1+w\}}{P\{Y=x_1\}}. \quad \dots \quad (19)$$

The right side of equation (19) is positive for $w=x_1-x_0$, so that, from the left side, not both of $P\{X=x_0+2(x_1-x_0)\}$ and $P\{Y=x_0+2(x_1-x_0)\}$ can be zero. By induction, each of the points $x_0+n(x_1-x_0)$ for $n=0,1,2,\dots$, must be a possible value either of X or of Y. Now let

$$C=\min\ \{P(X=x_1)/P(X=x_0),\ \ P(Y=x_1)/P(Y=x_0)\}.$$

Then, from equation (19),

$$\frac{P\{X=x_0+n(x_1-x_0)\}}{P\{X=x_0\}}+\frac{P\{Y=x_0+n(x_1-x_0)\}}{P\{Y=x_0\}}$$

$$\leqslant\frac{1}{C}\left(\frac{P\{X=x_0+(n+1)(x_1-x_0)\}}{P\{X=x_0\}}+\frac{P\{Y=x_0+(n+1)(x_1-x_0)\}}{P\{Y=x_0\}}\right)$$

for all n, so that by induction

$$\frac{P\{X=x_0+n(x_1-x_0)\}}{P\{X=x_0\}}+\frac{P\{Y=x_0+n(x_1-x_0)\}}{P\{Y=x_0\}}\geqslant 2C^n$$

for all n. This implies that $C<1$. In other words, either $P\{X=x_0\}>P\{X=x_1\}$ or $P\{Y=x_0\}>P\{Y=x_1\}$. Since x_0 was an arbitrary possible value of X and Y less than x_1, there can be at most a finite number of such possible values of either X or Y less than x_1 if both X and Y are to have finite probability mass.

We now suppose that x_0 is the smallest possible value of X and Y and that x_1 is the next smallest. From equation (19), $P\{W=w\}$ must be zero for $0<w<x_1-x_0$, and by induction $x_0+n(x_1-x_0)$ for $n = 0, 1, 2, \ldots$ are the only possible values of X and Y. We change the location and scale of the distributions if necessary so that $x_0 = 0$ and $x_1 = 1$. Now, 0 and 1 are possible values of both X and Y, and the only possible values of X and Y are the non-negative integers.

To simplify the notation, we shall let $a_k = P(X = k)$ and $b_k = P(Y = k)$. Then equation (18) after the elimination of $P(W = w)$ becomes

$$a_k b_k\left(\frac{a_n}{a_0}+\frac{b_n}{b_0}\right) = a_k b_{n+k}+a_{n+k}b_k, \qquad \ldots \quad (20)$$

for all non-negative integers k and n. From this we may derive the simultaneous equations,

$$a_k b_{k+1}+b_k a_{k+1} = a_k b_k\left(\frac{a_1}{a_0}+\frac{b_1}{b_0}\right),$$

$$a_1 b_{k+1}+b_1 a_{k+1} = a_1 b_1\left(\frac{a_k}{a_0}+\frac{b_k}{b_0}\right). \qquad \ldots \quad (21)$$

If the determinant

$$\Delta_k = a_k b_1 - b_k a_1 \qquad \ldots \quad (22)$$

is not equal to zero, we may solve equations (21) for a_{k+1} and b_{k+1}. Elementary manipulations yield

$$a_{k+1} = \left(\frac{a_1}{a_0}\right) a_k$$

$$b_{k+1} = \left(\frac{b_1}{b_0}\right) b_k. \qquad \ldots \quad (23)$$

Automatically, $a_0 \neq 0$, $b_0 \neq 0$, $a_1 \neq 0$, $b_1 \neq 0$, and not both of a_k and b_k can be zero for any positive integer k.

Case IIA : $a_1 b_0 = b_1 a_0$. Let $r = \dfrac{a_1}{a_0} = \dfrac{b_1}{b_0}$, and let n be the smallest positive integer for which $\Delta_n \neq 0$, or equivalently $a_n b_0 \neq a_0 b_n$. Then $n \geqslant 2$, and $a_k b_0 = a_0 b_k$ for $k < n$, so that

$$\left.\begin{aligned} a_k &= r^k a_0 \\ b_k &= r^k b_0 \end{aligned}\right\} \quad \text{for } k < n. \qquad \ldots \quad (24)$$

Suppose that $n < \infty$. Since $\Delta_n \neq 0$, equations (23) hold for $k = n$. Then $\Delta_{n+1} = r\Delta_n \neq 0$, and equations (23) hold for $k = n+1$. By induction, equations (23) hold for all $k \geqslant n$. Thus,

$$\left.\begin{aligned} a_k &= r^{k-n} a_n \\ b_k &= r^{k-n} b_n \end{aligned}\right\} \quad \text{for } k \geqslant n. \qquad \ldots \quad (25)$$

In addition, equation (20) reduces to

$$\frac{a_n}{a_0}+\frac{b_n}{b_0} = 2r^n \qquad \dots \quad (26)$$

which allows us to put $a_n = r^n(1+\theta)a_0$ and $b_n = r^n(1-\theta)b_0$. Then $\Sigma a_k = 1$ and $\Sigma\, b_k = 1$ imply $a_0 = (1-r)/(1+\theta r^n)$ and $b_0 = (1-r)/(1-\theta r^n)$, yielding the distributions found under (ii) with $n \geqslant 2$.

If $n = \infty$, then equations (24) hold for all values of k, giving geometric distributions with the same geometric parameter, a special case of (i).

Case IIB : $a_0b_1 \neq b_0a_1$ and $a_2b_1 = b_2a_1$. Since not both of a_2 and b_2 can be zero, $a_2b_1 = b_2a_1$ implies that neither can be zero. Thus we are as in Case IIA with the origin shifted one unit, so that

$$\left.\begin{aligned} a_k &= r^{k-1}\, a_1 \\ b_k &= r^{k-1}\, b_1 \end{aligned}\right\} \quad \text{for} \quad 1 \leqslant k < n+1 \qquad \dots \quad (27)$$

and

$$\left.\begin{aligned} a_k &= r^{n+k-1}\, a_{n+1} \\ b_k &= r^{n+k-1}\, a_{n+1} \end{aligned}\right\} \quad \text{for} \quad k \geqslant n+1 \qquad \dots \quad (28)$$

where n is determined so that $\Delta_k = 0$ for $1 \leqslant k \leqslant n$ and $\Delta_{n+1} \neq 0$, $(n \geqslant 2)$.

Suppose first that $n < \infty$; then, since $\Delta_{n+1} \neq 0$, equations (23) give

$$a_{n+2} = \left(\frac{a_1}{a_0}\right) a_{n+1}$$

$$b_{n+2} = \left(\frac{b_1}{b_0}\right) b_{n+1}.$$

Combined with (28), we see that $r = a_1/a_0 = b_1/b_0$, contradicting $a_0b_1 \neq b_0a_1$. Thus $n = \infty$, and equations (27) hold for all k. In addition, equation (20) with $n = k = 1$ implies that

$$\frac{a_1}{a_0} + \frac{b_1}{b_0} = 2r.$$

As in Case IIA, this reduces to the distributions found under (ii) with $n = 1$.

Case IIC : $a_0b_1 \neq b_0a_1$, $a_2b_1 \neq b_2a_1$ and $a_2b_0 = b_2a_0$. Let $r^2 = a_2/a_0 = b_2/b_0$. Then $r \neq 0$, since not both of a_2 and b_2 can be zero. Since Δ_2 is assumed to be different from zero, equation (23) with $k = 2$ becomes

$$\begin{aligned} a_3 &= r^2 a_1 \\ b_3 &= r^2\, b_1. \end{aligned} \qquad \dots \quad (29)$$

Furthermore, equation (20) with $k = n = 1$ becomes

$$a_1b_1 = r^2\, a_0\, b_0. \qquad \dots \quad (30)$$

The following simultaneous equations are also special cases of equation (20).

$$a_1 b_{n+1} + b_1 a_{n+1} = a_1 b_1 \left(\frac{a_n}{a_0} + \frac{b_n}{b_0} \right) \quad \ldots \quad (31)$$

$$a_2 b_{n+1} + b_2 a_{n+1} = a_2 b_2 \left(\frac{a_{n-1}}{a_0} + \frac{b_{n-1}}{b_0} \right).$$

We shall prove by induction that

$$\begin{matrix} a_{2k} = r^{2k} a_0 & \text{and} & a_{2k+1} = r^{2k} a_1 \\ b_{2k} = r^{2k} b_0 & & b_{2k+1} = r^{2k} b_1. \end{matrix} \quad \ldots \quad (32)$$

These four equations are already known to be valid for $k = 0$ and $k = 1$. We suppose that equations (32) are valid for $k < m$ and proceed to prove their validity for $k = m$. Since $\Delta_2 \neq 0$, equation (31) may be solved for a_{2m} and b_{2m} (when $n = 2m-1$). Elementary manipulations involving (30) and the induction hypotheses yield $a_{2m} = r^{2m} a_0$ and $b_{2m} = r^{2m} b_0$, the first two equations of (32) when $k = m$. But now $\Delta_{2m} = r^{2m}(a_0 b_1 - a_1 b_0) \neq 0$, so that from equations (23) $a_{2m+1} = a_1 a_{2m}/a_0 = r^{2m} a_1$, and $b_{2m+1} = r^{2m} b_1$. This proves (32) for $k = m$ completing the induction.

Equation (30) allows us to set $a_1 = \theta r a_0$ and $b_1 = r b_0/\theta$. Then $\Sigma a_k = 1$ and $\Sigma b_k = 1$ imply $a_0 = (1-r^2)/(1+r\theta)$ and $b_0 = (1-r^2)\theta/(\theta+r)$, yielding the distributions contained under (iv) when the inequalities involving θ are strict.

Case IID : $a_0 b_1 \neq b_0 a_1$, $a_2 b_1 \neq b_2 a_1$, $a_2 b_0 \neq b_2 a_0$ and $a_2 b_2 = 0$. Since not both a_2 and b_2 can be zero, either $a_2 \neq 0$ and $b_2 = 0$, or $a_2 = 0$ and $b_2 \neq 0$. We assume that $a_2 \neq 0$ and $b_2 = 0$, the other situation being completely analogous. Then, since $\Delta_2 \neq 0$, equation (23) gives $a_3 = (a_1/a_0)a_2$ and $b_3 = 0$ which shows that $\Delta_3 \neq 0$. Obviously this may be continued indefinitely, so that

$$\begin{matrix} a_k = r^{k-2} a_2 & \\ & \text{for} \quad k \geqslant 2 \\ b_k = 0 & \end{matrix}$$

where $r = a_1/a_0$. Furthermore, equation (20) with $k = n = 1$ entails $a_2 = a_1 \left(\frac{a_1}{a_0} + \frac{b_1}{b_0} \right)$ which allows us to set $a_2 = r^2(1+\theta)a_0$ and $b_1 = r\theta b_0$. Then, $\Sigma\, a_k = 1$ and $\Sigma\, b_k = 1$ imply $a_0 = (1-r)/(1+\theta r^2)$ and $b_0 = 1/(1+\theta r)$, yielding the distributions found under (iii). Of course, if we had assumed that $a_2 = 0$ and $b_2 \neq 0$, then the distributions would be as under (iii) with X and Y interchanged.

Case IIE : $a_0 b_1 \neq b_0 a_1$, $a_2 b_1 \neq b_2 a_1$, $a_2 b_0 \neq b_2 a_0$ and $a_2 b_2 \neq 0$. Since $\Delta_2 \neq 0$, equations (23) with $k = 2$

$$\begin{matrix} a_3 = r_1 a_2 \\ b_3 = r_2 b_2 \end{matrix} \quad \ldots \quad (33)$$

are valid, where r_1 and r_2 are defined by

$$a_1 = r_1 a_0 \qquad \ldots \quad (34)$$
$$b_1 = r_2 b_0.$$

Equations (33) and (34) imply that

$$\Delta_3 = r_1 r_2 (a_2 b_0 - b_2 a_0) \neq 0,$$

so that

$$a_4 = r_1^2 a_2 \qquad \ldots \quad (35)$$
$$b_4 = r_2^2 b_2.$$

Equation (20) with $k = n = 2$ may be transformed with the help of equations (35) and the hypothesis $a_2 b_2 \neq 0$ into the equation $a_0 b_2 + b_0 a_2 = (r_1^2 + r_2^2) a_0 b_0$. This and equation (20) with $k = n = 1$ form two simultaneous linear equations in a_2 and b_2 whose determinant is $a_0 b_1 - a_1 b_0 \neq 0$. Solving these equations gives

$$a_2 = r_1^2 a_0$$
$$b_2 = r_2^2 b_0.$$

Now, we may show by induction that $\Delta_n = a_n b_1 - a_1 b_n = a_1 b_1 [r_1^{n-1} - r_2^{n-2}] \neq 0$, so that equations (23) become

$$a_k = r_1^k a_0 \qquad \text{for all } k$$
$$b_k = r_2^k b_0.$$

This obviously implies that both X and Y have the geometric distributions found under (i).

In order to show that each of the four forms of the distributions of X and Y, given by (i), (ii), (iii), and (iv) in the statement of Theorem 2, lead for independent X and Y to independent U and W, it is sufficient in Case II merely to check that equation (20) is valid for all integers k and n. In each of the cases, this is a tedious but straightforward task. In the interest of brevity these calculations are omitted.

We finish with the following corollary which is an immediate consequence of Theorem 2.

Corollary : *Let X and Y be independent, identically distributed, non-degenerate, discrete random variables. Then $U = \min (X, Y)$ and $W = |X - Y|$ are independent if, and only if, X and Y have geometric distributions.*

References

Bartoo, J. B. (1952) : *Certain Theorems on Order Statistics*, unpublished Ph.D. thesis, State University of Iowa.

Bolger, E. M. and Harkness, W. L. (1965) : Characterizations of some distributions by conditional moments. *Ann. Math. Stat.*, **36**, 703-705.

Crawford, G. B. (1966) : Characterizations of geometric and exponential distributions. *Ann. Math. Stat.*, **37**, 1790-1795.

Epstein, B. and Sobel, M. (1954) : Some theorems relevant to life testing from an exponential distribution. *Ann. Math. Stat.*, **25**, 373-381.

Ferguson, T. S. (1964) : A characterization of the negative exponential distribution. *Ann. Math. Stat.*, **35**, 1199-1207.

——— (1965) : A characterization of the geometric distribution. *Amer. Math. Monthly*, **72**, 256-260.

Fisz, M. (1958) : Characterization of some probability distributions. *Skand. Aktuarietidskr*, **41**, 65-70.

Govindarajulu, Z. (1966) : Characterization of normal and generalized truncated normal distributions using order statistics. *Ann. Math. Stat.*, **37**, 1011-1015.

Lukacs, E. (1965) : Characterization problems for discrete distributions. *Sankhyā*, Series A, **27**, 231-240.

Malmquist, S. (1950) : On a property of order statistics from a rectangular distribution. *Skand. Aktuarietidskr*, **33**, 214-222.

Rogers, G. S. (1963) : An alternative proof of the characterization of the density Ax^B. *Amer. Math. Monthly*, **70**, 857-858.

Rossberg, H. J. (1966) : Charakterisierungs probleme, die sich aus der von A. Rényi in die Theorie der Ranggrossen eingeführten Methode ergeben. Monatsb. Deutsch. Akad. Wiss, Berlin, **8**, 561-572.

Sethuraman, J. (1965) : On a characterization of the three limiting types of the extreme. *Sankhyā*, Series A, **27**, 357-364.

Tanis, E. (1964) : Linear forms in the order statistics from an exponential distribution. *Ann. Math. Stat.*, **35**, 270-276.

Paper received : January, 1967.