ON THE ASSOCIATION OF THE PARETO AND THE YULE DISTRIBUTION*

E. XEKALAKI AND J. PANARETOS

1. Introduction. Engel and Zijlstra [3], investigating problems in the area of demand and supply analysis, showed that the distribution of the number of orders for a product during a lead time of such production quantities is negative binomial if and only if the lead time has a gamma distribution. They showed this by rederiving Feller's [4] interesting result that by means of the Poisson process each probability distribution on $[0, +\infty]$ is mapped on a discrete probability distribution defined on $\{0, 1, 2, \cdots\}$ and that the mapping is one-to-one. The result was then used to obtain characterizations of the exponential distribution corresponding to existing characterizations of the geometric distribution.

* Received by the editors June 6, 1986.

The analogy that exists between the studied characterizations of the exponential distribution and those of the geometric distribution is noteworthy because of the fact that the exponential distribution is the continuous analogue of the geometric distribution. The Pareto distribution and the Yule distribution (see [10]) comprise another pair of distributions that relate to each other in the same manner. In particular, the Pareto distribution is the continuous analogue of the Yule distribution. Both of these distributions have numerous applications. The Pareto distribution, for example, has been used extensively in economics literature to model distributions of income. The Yule distribution, on the other hand, has been employed in the context of problems in biology, linguistics, economics, and bibliographic research (see, e.g., [6], [9]). Furthermore, the Yule distribution was shown by Xekalaki [7] to arise as a demand distribution. This combined with the fact that the Yule distribution is an exponential mixture of the geometric distribution while the Pareto distribution is an exponential mixture of the exponential distribution makes one wonder whether it is possible to derive characterizations of the Pareto distribution corresponding to existing characterizations of the Yule distribution.

In this paper it is shown that indeed this is the case. In § 2, two characterization theorems are proved for the Pareto distribution. In § 3, it is pointed out that generalized Γ-convolutions (in Thorin's [5] sense) can be characterized by generalized negative binomial convolutions (in Bondesson's [2] sense). It is further shown that the results of § 2 can also be derived as a consequence of a general result associating the mixing distribution in the mixture representation of a generalized Γ-convolution to the generalized negative binomial convolution on which the former is mapped through a Poisson process.

2. Some characterizations of the Pareto distribution. Let us first introduce some notation and terminology. A non-negative, integer-valued random variable (r.v.) X is said to have the Yule distribution with parameter $a>0$ (Yule (a)) if its probability function (p.f.) is given by

$$\mathbf{P}(X=x)=ax!/(a+1)_{(x+1)}, \qquad x=0,1,2,\cdots, \tag{2.1}$$

where $\alpha_{(\beta)}=\Gamma(\alpha+\beta)/\Gamma(\alpha)$, $\alpha>0$, $\beta\in R$.

Xekalaki [8], dealing with a problem in reliability theory, showed that a non-negative, integer-valued r.v. X has the Yule distribution with p.f. given by (2.1) if and only if

$$\mathbf{P}(X=k|X\geqq k)=a/(k+a+1), \qquad k=0,1,2,\cdots. \tag{2.2}$$

We now prove a characterization theorem concerning the Pareto distribution.

THEOREM 2.1. *Let $Y, X_1, X_2, \cdots$ be independent nondegenerate r.v.'s with probability density functions (p.d.f.'s) defined on $[0,+\infty)$. Assume that X_i, $i=1,2,\cdots$, has the Pareto distribution with parameter 1 and distribution function (d.f.).*

$$F(x)=1-(1+x)^{-1} \tag{2.3}$$

and let $Y_r=\sum_{i=1}^{r} X_i$, $r=1,2,\cdots$, $Y_0=0$. Then Y has the Pareto distribution with parameter $a>0$ and d.f.

$$F_Y(y)=1-(1+y)^{-a} \tag{2.4}$$

if and only if

$$\mathbf{P}(Y>Y_{k+1}|Y>Y_k)=k+1/(k+a+1), \qquad k=0,1,2,\cdots. \tag{2.5}$$

Proof. Relationship (2.5) is obviously equivalent to

$$\mathbf{P}(Y>Y_{k+1})-((k+1)/(k+a+1))\mathbf{P}(Y>Y_k)=0, \qquad k=0,1,2,\cdots.$$

This is a first-order difference equation in $\mathbf{P}(Y>Y_k)$ with a unique solution (under the condition $\mathbf{P}(Y>0)=1$) given by

$$\mathbf{P}(Y>Y_k)=k!/(a+1)_{(k)}, \qquad k=0,1,2,\cdots. \tag{2.6}$$

But, for $k=0,1,2,\cdots$, $\mathbf{P}(Y>Y_k)=\int_0^\infty \mathbf{P}(Y>y)f_{Y_k}(y)\,dy$ where $f_{Y_k}(y)$ denotes the p.d.f. of the distribution of Y_k. This is a Pearson type VI distribution (beta of the second kind) with parameters k and 1. Therefore, (2.5) holds if and only if

$$k\int_0^\infty (1-F_Y(y))y^{k-1}(1+y)^{-(k+1)}\,dy=\frac{k!}{(a+1)_{(k)}}, \qquad k=1,2,\cdots. \tag{2.7}$$

One solution to this equation is given by (2.4) and it is unique by the completeness of the Pearson type VI family of distributions. This establishes the result.

Note 1. Relationships of type (2.5) may be of interest in problems in the area of reliability theory. The r.v. Y, for example, can be considered as representing the strength of a component subjected to an accumulation of stress from r independent environmental factors denoted by $Y_r = X_1 + \cdots + X_r$ (X_i represents the stress administered by the ith factor). Then, according to Birnbaum's [1] definition of the reliability function, $\mathbf{P}(Y > Y_{k+1} | Y > Y_k)$ denotes the reliability of the component to survive the cumulative stress Y_{k+1} given it has survived Y_k.

Note 2. The characterization of the Pareto distribution by Theorem 2.1 is a variant of Xekalaki's [8] characterization of the Pareto distribution by the form of the hazard rate.

In the context of a problem of income underreporting, Xekalaki [7] showed that a non-negative, integer-valued r.v. X with p.f. $p_X = \mathbf{P}(X = x)$ has a Yule distribution if and only if

$$\frac{1}{1-p_0} \sum_{x=r+1}^{\infty} \frac{p_x}{x} = p_r, \qquad r = 0, 1, 2, \cdots. \tag{2.8}$$

The following theorem can now be shown concerning the Pareto distribution.

THEOREM 2.2. *Let $Y, X_1, X_2, \cdots$ and Y_r be r.v.'s defined as in Theorem 2.1. Then Y has the Pareto distribution with parameter $(1-c)/c$ if and only if*

$$\sum_{x=k+1}^{\infty} \frac{\mathbf{P}(Y_x < Y < Y_{x+1})}{x} = c\mathbf{P}(Y_k < Y < Y_{k+1}), \qquad k = 0, 1, 2, \cdots, \tag{2.9}$$

where $c = \mathbf{P}(Y > Y_1)$.

Proof. "If" part. Denote $\mathbf{P}(Y_k < Y < Y_{k+1})$ by P_k, $k = 0, 1, 2, \cdots$. Specializing (2.9) for $k = r$ and $k = r+1$ and subtracting the resulting equations we obtain

$$P_{r+1}/r+1 = c(P_r - P_{r+1}), \qquad r = 0, 1, 2, \cdots,$$

which is equivalent to

$$P_{r+1} - (c(r+1)/(c(r+1)+1))P_r - 0, \qquad r = 0, 1, 2, \cdots.$$

This is a first-order difference equation in P_r with a unique solution given for $r = 0, 1, 2, \cdots$ by

$$P_r = (1-c)r!/(1+1/c)_{(r)}, \qquad r = 0, 1, 2, \cdots. \tag{2.10}$$

Recalling that $P_r = \mathbf{P}(Y_r < Y < Y_{r+1}) = \mathbf{P}(Y > Y_r) - \mathbf{P}(Y > Y_{r+1})$ we obtain

$$\mathbf{P}(Y > Y_r) = 1 - (1-c) \sum_{i=0}^{r-1} \frac{i!}{(1+1/c)_{(i)}}, \qquad r = 1, 2, \cdots. \tag{2.11}$$

But

$$\sum_{i=0}^{r-1} \frac{i!}{(1+1/c)_{(i)}} = \frac{\Gamma(1+1/c)\Gamma(-1+1/c)}{\Gamma(1/c)\Gamma(1/c)} - \frac{r!}{(1+1/c)_{(r)}} \sum_{i=0}^{+\infty} \frac{(r+1)_{(i)}}{(r+1+1/c)_{(i)}}$$

$$= \frac{1/c}{1/c-1} - \frac{r!}{(1+1/c)_{(r)}} \frac{\Gamma(r+1+1/c)\Gamma(-1+1/c)}{\Gamma(1/c)\Gamma(r+1/c)}$$

$$= \frac{1}{1-c} - \frac{c}{1-c} \frac{r!}{(1+1/c)_{(r-1)}}.$$

Hence (2.11) reduces to

$$\mathbf{P}(Y > Y_r) = cr!/(1+1/c)_{(r-1)}, \qquad r = 1, 2, \cdots. \tag{2.12}$$

Combining (2.10) and (2.12) we get

$$\mathbf{P}(Y_r < Y < Y_{r+1}) = ((1-c)/c(r+1/c))\mathbf{P}(Y > Y_r), \qquad r = 0, 1, 2, \cdots,$$

which is equivalent to relationship (2.5) of Theorem 2.1. Hence Y has a Pareto distribution with parameter $(1-c)/c$.

"Only if" part. Suppose that Y has a Pareto distribution with parameter $b > 0$. Then it follows from (2.7) that

$$\mathbf{P}(Y > Y_k) = k!/(b+1)_{(k)}, \qquad k = 0, 1, 2, \cdots.$$

This, in turn, implies that

$$\mathbf{P}(Y_k < Y < Y_{k+1}) = bk!/(b+1)_{(k+1)}, \qquad k = 0, 1, 2, \cdots. \tag{2.13}$$

But the right-hand side of (2.13) represents for $k=0, 1, 2, \cdots$ the p.f. of the Yule distribution. Hence by Xekalaki's result [7], relationship (2.9) is satisfied. This completes the proof of the theorem.

3. Characterization of probability distributions by generalized negative binomial convolutions. In this section it will be shown that our results of § 2 can alternatively be obtained by using a unifying result that connects two general families of distributions. These are the family of generalized negative binomial convolutions (see [2]) and the family of generalized Γ-convolutions (see [5]).

It is interesting that Bondesson [2] obtained a result concerning these two families of distributions which can be considered as a special case of Feller's [4] and Engel and Zijlstra's [3] results on the identifiability of compound Poisson distributions. This, employing Engel and Zijlstra's [3] terminology, states that for a homogeneous Poisson process $\{N(t), t \geqq 0\}$ with parameter $\lambda = 1$, the distribution of the number $N(Z)$ of points in the time interval $[0, Z)$, with Z as a non-negative r.v. independent of $\{N(t), t \geqq 0\}$, is a generalized negative binomial convolution if and only if the distribution of Z is a generalized Γ-convolution. Furthermore, it was shown that every generalized Γ-convolution is a scale mixture of the gamma distribution with p.d.f.

$$h(x) = \int_0^\infty \frac{a^\beta}{\Gamma(\beta)} x^{\beta-1} e^{-ax}\, dF(a), \qquad x > 0, \quad \beta > 0,$$

where F is a distribution function. This result is of central importance to our discussion because as indicated by the theorem that follows it leads to a one-to-one correspondence between the mixing distribution F and the distribution of $N(Z)$.

THEOREM 3.1. *Let $\{N(t), t > 0\}$ be a homogeneous Poisson process with parameter $\lambda = 1$. Let Z_1, Z_2 be two independent, non-negative r.v.'s that are distributed independently of $\{N(t), t \geqq 0\}$ with p.d.f.'s*

$$h_{Z_i}(z) = \int_0^\infty \frac{a_i^\beta}{\Gamma(\beta)} z^{\beta-1} e^{-a_i z}\, dF_i(a_i), \qquad z, \beta > 0, \quad i = 1, 2. \tag{3.1}$$

Then F_1 is equal to F_2 if and only if $N(Z_1)$ and $N(Z_2)$ are identically distributed.

Proof. The "only if" part is straightforward. For the "if" part, observe that by Feller's [4] result, $N(Z_1)$ and $N(Z_2)$ are identically distributed if and only if Z_1, Z_2 are identically distributed, i.e., if and only if $h_{Z_1}(x) = h_{Z_2}(x)$. The result then follows by the uniqueness of the Laplace transform.

The theorem that has just been proved implies that when Z has a distribution of the form (3.1) i.e., $Z \sim$ gamma $(a, \beta) \wedge F(a)$, the distribution of $N(Z)$ uniquely determines the mixing distribution F. (Here $\wedge$ denotes mixing with respect to a.) Therefore the form of a generalized negative binomial convolution characterizes the mixing distribution involved in the representation of the associated generalized Γ-convolution as a mixture of gamma distributions. As an illustration, consider the following characterization of Pearson type VI distributions by the generalized Waring distribution. (The generalized Waring distribution is a generalized negative binomial convolution that has applications in many diverse fields (see, e.g., [6].)

COROLLARY 3.1 (Characterization of the Pearson type VI distribution). *Let $\{N(t), t \geqq 0\}$ and Z be defined as in Theorem 3.1. Then the distribution of $N(Z)$ is the generalized Waring distribution with p.f.*

$$\frac{a_{(c)}}{(a+b)_{(c)}} \frac{b_{(x)} c_{(x)}}{(a+b+c)_{(x)}} \frac{1}{x!}, \qquad a, b, c > 0, \quad x = 0, 1, 2, \cdots \tag{3.2}$$

if and only if F is the distribution function of the Pearson type VI distribution with parameters a and b as defined by the p.d.f.

$$f(u) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} u^{a-1}(1+u)^{-(a+b)}, \qquad u > 0.$$

Proof. The proof of this corollary follows immediately from Theorem 3.1 if one observes that the generalized Waring distribution can result from a Poisson (λ) distribution whose λ has a gamma distribution with a scale parameter that is itself a r.v. having the Pearson type VI distribution.

COROLLARY 3.2 (Characterization of the Pareto distribution). *The distribution of $N(Z)$ is the Yule with parameter a if and only if F is the distribution function of the Pareto distribution with parameter a.*

Proof. The result follows from Corollary 3.1 since the Yule distribution is a special case of the generalized Waring distribution as given by (3.2) for $b=c=1$.

Note. Obviously, for a generalized negative binomial convolution (g.n.b.c.) we have g.n.b.c. $\sim$ Poisson $\wedge$ generalized Γ-convolution, g.n.b.c. $\sim$ Poisson $\wedge$ gamma $\wedge$ F, g.n.b.c. $\sim$ negative binomial $\wedge$ F. Therefore, Corollaries 3.1 and 3.2 illustrate the fact that by means of the negative binomial distribution, Theorem 3.1 establishes a one-to-one mapping of probability on $[0, +\infty)$ on discrete probability distributions on $\{0, 1, 2, \cdots\}$.

Let us now give a brief description of the way in which Theorem 3.1 along with Corollary 3.2 unifies the derivation of the results of Theorems 2.1 and 2.2 by connecting them to the characterizations of the Yule distribution mentioned in § 2.

The r.v. X in relationships (2.3) and (2.9) can be thought of as representing the image $N(Z)$ of a non-negative r.v. Z whose distribution is a scale mixture of a gamma distribution ($Z \sim$ gamma $(a, b) \wedge F(a)$). Theorem 3.1 implies that the mixing distribution F can be uniquely determined while Corollary 3.2 specifies that it is of a Pareto form. It has been shown [8] that for a Yule (a) r.v. X, $\mathbf{P}(X \geqq k) = k!/(a+1)_{(k)}$. It can now be checked easily that the right-hand side of this equation is equal to $\mathbf{P}(Y > X_1 + \cdots + X_k)$ where Y is a Pareto (a) r.v. and X_i, $i = 1, 2, \cdots k$, are mutually independent Pareto (1) r.v.'s independent of Y. This is a consequence of the fact that

$$\mathbf{P}(Y > X_1 + \cdots + X_k) = k \int_0^\infty y^{k-1}(1+y)^{-(k+a+1)}\, dy.$$

It follows then that relationships (2.2) and (2.8) are equivalent to relationships (2.5) and (2.9), respectively, and hence the results of Theorems 2.1 and 2.2 can be derived through the corresponding characterizations of the Yule distribution.

It is obvious from the previous argument that if $X, Y, X_1, X_2, \cdots$ are defined as above then

$$\mathbf{P}(X \geqq k) = \mathbf{P}(Y > X_1 + \cdots + X_k), \qquad k = 1, 2, \cdots. \tag{3.3}$$

The same relationship also holds when Y is an exponential r.v. and $X_1, X_2, \cdots$ are independent r.v.'s that are exponentially distributed independently of Y with parameter 1 while X is geometric. This implies that mixing each of the r.v.'s involved with the same exponential distribution does not affect their interrelationship.

REFERENCES

[1] Z. W. BIRNBAUM, *On the use of the Mann-Whitney statistics*, in Proc. of the Third Berkeley Symposium on Mathematical Statistics and Probability, J. Neyman, ed., University of California Press, Berkeley, 1956, pp. 13-17.

[2] L. BONDESSON, *On generalized gamma and generalized negative binomial convolutions*, Scand. Actuar. J., 2 (1979), pp. 125-166.

[3] J. ENGEL AND M. ZIJLSTRA, *A characterization of the gamma distribution by the negative binomial distribution*, J. Appl. Probab., 17 (1980), pp. 1138-1144.

[4] W. FELLER, *On a general class of contagious distributions*, Ann. Math. Statist., 14 (1943), pp. 389-399.

[5] O. THORIN, *On the infinite divisibility of the Pareto distribution*, Scand. Actuar. J., 1 (1977), pp. 31-40.

[6] E. XEKALAKI, *Chance mechanisms for the univariate generalized Waring distribution and related characterizations*, in Statistical Distributions in Scientific Work, Vol. 4, C. Taillie, G. P. Patil, B. Balderassi, eds., D. Reidel, Dordrecht, the Netherlands, 1981, pp. 157-171.

[7] ———, *A property of the Yule distribution and its applications*, Commun. Statis. Ser. A., pp. 1181-1189.

[8] ———, *Hazard functions and life distributions in discrete time*, Commun. Statis. Ser. A., pp. 2503-2509.

[9] ———, *Linear regression and the Yule distribution*, J. Econometrics, 24 (1984), pp. 397-403.

[10] G. W. YULE, *A mathematical theory of evolution based on the conclusions of J. C. Willis F.R.S.*, Phil. Trans. Roy. Soc. London, Ser. B, vol. 213, pp. 21-87.