Characterization Theorems for Some Univariate Probability Distributions

By G. P. Patil and V. Seshadri

McGill University

[Received November 1963. Revised March 1964]

SUMMARY

Suppose that one is dealing with a bivariate distribution of random variables $X$ and $Y$ where only the conditional distribution of $X$ given $X + Y$ is known. What can be said of the marginal distributions of $X$ and $Y$? This paper attempts to answer this question for situations involving certain univariate distributions known in the statistical literature.

1. INTRODUCTION

Characterization theorems for probability distributions have appeared from time to time in the literature, and most of them deal with characteristic functions or are based on the independence of a pair of suitable statistics. Lukacs (1956) has given an extensive bibliography of such theorems. Teicher (1961) has used the maximum-likelihood estimator to characterize the normal and the exponential distributions. Ferguson (1962a) has dealt with the characterization of distributions via location and scale parameters. He has also shown (1962b) that if $X$ and $Y$ are independent, non-degenerate and discrete random variables, then a necessary and sufficient condition for $X$ and $Y$ to have a geometric distribution is that $\min(X, Y)$ and $X - Y$ are stochastically independent. Parzen (1962, p. 123) has given two exercises which deal with quite another characterization of the exponential and the geometric distributions. Mauldin (1961) has studied the problem of finding out when the distribution of a sample statistic determines that of the population. A class of characterization problems treated by Laha (1959) considers the situation where, if a certain function of a pair of independent random variables follows a known distribution, then under suitable conditions the parent random variables have a specified distribution. Recently Patil and Seshadri (1963) have studied characterization problems for bivariate distributions by using the marginal and the conditional distributions of the same component.

The results obtained in this paper are motivated by a theorem of Moran (1951) characterizing the Poisson distribution which, for independent variates $X$ and $Y$ taking non-negative integral values, states that if the conditional distribution of $X$ given the total $X+Y$ is a binomial distribution with a common parameter $p$ for all given values of $X+Y$, and if there exists at least one integer $i$ so that $\text{Prob}(X = i)$ and $\text{Prob}(Y = i)$ are both positive, then $X$ and $Y$ are individually distributed in Poisson distributions. We investigate results of this nature to characterize the binomial, the negative binomial and the geometric distributions among the discrete distributions, and the normal, the negative exponential and Pearson’s Type III among the absolutely continuous distributions.
2. A GENERAL RESULT

Suppose that the following assumptions are made about the random variables \( X \) and \( Y \), whose probability distributions are denoted by \( f(x) \) and \( g(y) \).

(i) \( X \) and \( Y \) are independent.

(ii) \( X \) and \( Y \) are both discrete or both continuous.

(iii) The conditional distribution of \( X \) given \( (X + Y) \), denoted by

\[
c(x, x + y) = c(x, x + y, \alpha),
\]

where \( \alpha \) is a real parameter, is such that

\[
\frac{c(x + y, x + y)c(0, y)}{c(x, x + y)c(y, y)}
\]

is of the form \( h(x + y)h(x)h(y) \) where \( h(\cdot) \) is an arbitrary non-negative function.

(iv) \( f(x), g(y) > 0 \).

The following theorem can now be established to characterize several probability distributions.

**Theorem 1.** Under the assumptions (i)-(iv),

\[
f(x) = f(0) h(x) \exp(\alpha x),
\]

where \( \alpha \) is an arbitrary constant and \( f(0) \) a suitable normalizer which makes \( f(x) \) a probability function;

\[
g(y) = g(0) k(y) \exp(\alpha y)
\]

where

\[
k(y) = h(y) c(0, y)/c(y, y),
\]

and \( g(0) \) is the corresponding normalizer for \( g(y) \).

**Proof.** One has

\[
c(x, x + y) = f(x) g(y)/r(x + y)
\]

where

\[
r(z) = \Sigma f(x) g(z - x) \text{ or } \int f(x) g(z - x) dx
\]

as the case may be. Hence

\[
f(x) g(y) = r(x + y) c(x, x + y).
\]  \hspace{1cm} (1)

Put \( y = 0 \) in (1) and change \( x \) to \( x + y \) to obtain

\[
f(x + y) g(0) = r(x + y) c(x + y, x + y).
\]  \hspace{1cm} (2)

Thus

\[
\frac{f(x + y) g(0)}{f(x) g(y)} = \frac{c(x + y, x + y)}{c(x, x + y)}.
\]  \hspace{1cm} (3)

Setting \( x = 0 \) in (3) yields

\[
\frac{f(y) g(0)}{f(0) g(y)} = \frac{c(y, y)}{c(0, y)}.
\]  \hspace{1cm} (4)

From (3) and (4) it is clear that

\[
\frac{f(x + y) f(0)}{f(x) f(y)} = \frac{c(x + y, x + y) c(0, y)}{c(x, x + y) c(y, y)}.
\]
and by assumption (iii)

\[ \frac{f(x+y)h(0)}{f(x)h(y)} = \frac{h(x+y)}{h(x)h(y)}. \]  \hspace{1cm} (5)

The transformation \( \phi(x) = f(x)/f(0)h(x) \) reduces (5) to the well-known Cauchy's functional equation \( \phi(x+y) = \phi(x)\phi(y) \), the solution of which is \( \phi(x) = \exp(ax) \), \( a \) being an arbitrary constant. Thus \( f(x) = f(0)h(x)\exp(ax) \).

From (4),

\[ g(y) = g(0)f(y)c(0,y)/f(0)c(y,y). \]

Hence

\[ g(y) = g(0)k(y)\exp(ay), \]

where

\[ k(y) = h(y)c(0,y)/c(y,y). \]

If there are two functions \( h_1(\cdot) \) and \( h_2(\cdot) \) satisfying (iii) it is quite easy to show that \( h_1(x) = h_2(x)\exp(bx) \), \( b \) being an arbitrary constant. Thus the functional form of \( f(\cdot) \) and \( g(\cdot) \) remains unchanged.

3. Special Results for Discrete Distributions

As applications of Theorem 1 are given the following corollaries for non-negative integer-valued random variables \( X \) and \( Y \) for which the positivity of \( P(X = 0) \) and \( P(Y = 0) \) alone is required.

**Corollary 1.** If the conditional distribution of \( X \) given the total \( X + Y \) is the hypergeometric distribution with parameters \( m \) and \( n \), and \( z = x+y \) for all given values of \( X + Y \), then both \( X \) and \( Y \) have the binomial distribution with the same parameter \( p \), the other parameters being \( m \) and \( n \) respectively.

**Proof.** In Theorem 1, take

\[ c(x,x+y) = \binom{m}{x} \binom{n}{z-x} / \binom{m+n}{z}. \]

Then

\[ h(x) = \binom{m}{x} \text{ and } k(y) = \binom{n}{y}. \]

Thus

\[ f(x) = f(0) \binom{m}{x} \exp(ax) \]

and

\[ g(y) = g(0) \binom{n}{y} \exp(ay). \]

Let \( \exp(a) = \alpha \), and it is seen that \( f(0) = (1+\alpha)^{-m} \) and \( g(0) = (1+\alpha)^{-n} \), whence

\[ f(x) = \binom{m}{x} \theta^{x}(1-\theta)^{m-x} \]

and

\[ g(y) = \binom{n}{y} \theta^{y}(1-\theta)^{n-y} \]

where

\[ \theta = \alpha/(1+\alpha). \]
Corollary 2. If the conditional distribution of \( X \) given the total \( X + Y \) is a binomial distribution with a common parameter \( p \) for all given values of \( X + Y \), then both \( X \) and \( Y \) have Poisson distributions.

\[
c(x, x+y) = \binom{x+y}{x} p^x (1-p)^{y} = \binom{x+y}{x} \left( \frac{p}{1-p} \right)^x (1-p)^{y}.
\]

One can show easily that \( h(x) = 1/x! \) and \( k(y) = (1/(1-p)p)^y/y! \) Thus

\[
f(x) = f(0) \exp(ax)/x!
\]

and \( g(y) = g(0)((1-p)/p)^y \exp(ay)/y! \) Set \( \lambda = e^a \) and \( (1-p) e^{a}/p = \mu \). Then

\[
f(x) = e^{-\lambda} \frac{\lambda^x}{x!} \quad \text{and} \quad g(y) = e^{-\mu} \frac{\mu^y}{y!}.
\]

Corollary 3. If the conditional distribution of \( X \) given \( X + Y \) is the negative hypergeometric distribution with parameters \( m \) and \( n \), and \( z = x+y \) for all \( x+y \), then both \( X \) and \( Y \) have the negative binomial distribution with the same parameter \( \theta \), the other parameters being \( m \) and \( n \) respectively.

Proof. Take

\[
c(x, x+y) = \binom{x+y}{x} B(m+x, n+y)/B(m, n),
\]

where \( B(m, n) \) is the beta function with arguments \( m \) and \( n \). It can be verified that

\[
h(x) = \Gamma(m+x)/x! \quad \text{and} \quad k(y) = \Gamma(n+y)/y!.
\]

Then \( f(0) = (1-\lambda)^m/\Gamma(m) \) and

\[
g(0) = (1-\lambda)^n/\Gamma(n),
\]

where \( \lambda = \exp(a) \). Thus

\[
f(x) = \binom{x+m-1}{m-1} \lambda^x (1-\lambda)^m
\]

and

\[
g(y) = \binom{y+n-1}{n-1} \lambda^y (1-\lambda)^n.
\]

Corollary 4. If the conditional distribution of \( X \) given the total \( X + Y \) has the discrete uniform distribution for all given values of the total \( X + Y \), then both \( X \) and \( Y \) have identical geometric distributions.

This is a special case of Corollary 3 with \( m = n = 1 \) yielding \( h(x) = k(y) = 1 \). Therefore \( f(x) = \lambda^x (1-\lambda) \) and \( g(y) = \lambda^y (1-\lambda) \).

4. Special Results for Continuous Distributions

As applications of Theorem 1 are given the following corollaries for absolutely continuous random variables \( X \) and \( Y \).

Corollary 5. If the conditional distribution of \( X \) given \( X + Y \) is the normal distribution with mean \( (X+Y)/2 \), and if \( f(0) \) and \( g(0) \neq 0 \), then both \( X \) and \( Y \) are normal with the same location and scale parameters.

Proof.

\[
c(x, x+y) = (1/\sqrt{(2\pi)}) \exp\{-((x-y)^2)/2\sigma^2\},
\]
where $a > 0$. It can be verified that $h(x) = \exp(-x^2/4a^2)$ and $k(y) = \exp(-y^2/4a^2)$. From the theorem

$$f(x) = f(0) \exp\left\{ (-x^2/4a^2) + ax \right\},$$

whence

$$f(0) = (1/\sqrt{2\pi}) \exp(-a^2/4a^2).$$

Thus

$$f(x) = (1/\sqrt{2\pi}) \exp\left\{ -(x-2ax)^2/4a^2 \right\}$$

and

$$g(y) = (1/\sqrt{2\pi}) \exp\left\{ -(y-2ay)^2/4a^2 \right\}.$$

**Corollary 6.** If the conditional distribution of $X$ given $X + Y = Z$ is the uniform density over $(0, Z)$, where $X$ and $Y$ are independent non-negative random variables, then both $X$ and $Y$ have the negative exponential distribution with the same scale parameter.

**Proof.** $c(x, x+y) = 1/z (0 < x < z = x+y)$. Therefore $h(x) = 1$ and $k(y) = 1$ yielding $f(x) = f(0) e^{ax} = \theta e^{-\theta x}$, where $\theta = a$.

Similarly $g(y) = \theta e^{-\theta y}$.

If $c(x, x+y)$ is such that $c(x, x)$ and $c(0, x)$ are both zero, as in the case when

$$c(x, x+y = z) = (1/z) (x/z)^n (1-x/z)^n / B(m+1, n+1),$$

then a characterization of Pearson’s Type III distribution is available from the following theorem.

**Theorem 2.** Let $X$ and $Y$ be two independent non-negative random variables with absolutely continuous distributions $f(x)$ and $g(y)$ where (i) $f(x) = O(x^m)$ and (ii) $g(y) = O(y^n)$ as $x \rightarrow 0$. If the conditional distribution of $X$ given the total $Z = X + Y$ is a beta distribution of the first kind on $(0, Z)$, then both $X$ and $Y$ have Pearson’s Type III distribution with the same scale parameter.

**Proof.** Let

$$h(x/z) = (1/z) (x/z)^m (1-x/z)^n / B(m+1, n+1),$$

$$p(x) = f(x)/x^m, \quad q(z-x) = g(z-x)/(z-x)^n.$$  

Then, following the steps in the proof of Theorem 1, one finds that $f(x) = \theta \Gamma(n+1)$, and for $f(x)$ to be a density $k$ must be negative, say $-\theta$, and hence

$$p(0) = e^{-\theta x}.$$ 

Similarly it can be established that $g(y) = \theta^{n+1} y^n e^{-\theta y} / \Gamma(n+1)$.

It may not be surprising that both $f(\cdot)$ and $g(\cdot)$ are members of the exponential family $c(\theta) \exp\{ \theta f(x) \}$ where, in Theorem 1, $f(x)$ is linear. The results of Koopman and Pitman (1936) on sufficiency confirm Theorem 1. Besides establishing a special case of their results in a simpler manner, Theorem 1 shows that it is possible to obtain the marginal distributions of the independent random variables $X$ and $Y$, when their conditional distribution alone is known relative to the specified partition of the sample space.

5. A General Result for Discrete Sample Space

Suppose that the conditional probabilities of $X$ and $Y$ corresponding to the partition $X + Y = Z$ are given in the form of a table of constants, as for instance when $X$ and $Y$ are discrete variables. The following theorem can be of help in the determination of the marginal distributions of $X$ and $Y$. 

Theorem 3. Let $X$ and $Y$ be independent random variables defined over the non-negative integers $a \leq x \leq a+k$ and $b \leq y \leq b+m$ respectively. If $P(X = a) > 0$, $P(Y = b) > 0$, and $c(x, x+y)$ is the conditional distribution of $X$ given $X+Y$, then

$$f(x) = P(X = x) = A(x) \theta^x \sum_{z=a}^{a+k} A(x) \theta^z$$

and

$$g(y) = P(Y = y) = B(y) \lambda^y \sum_{u=b}^{b+m} B(y) \lambda^u,$$

where

$$\theta = g(b+1)/g(b), \quad \lambda = f(a+1)/f(a),$$

$$A(x) = \prod_{i=1}^{x-a} \{c(a+i, a+b+i)/c(a+i-1, a+b+i)\},$$

and

$$B(y) = \prod_{j=1}^{y-b} \{c(b+j, a+b+j)/c(b+j-1, a+b+j)\}.$$

Proof. Since $f(x)g(z-x) = r(z)c(x,z)$, one has for all $z$

$$f(a+1)g(z-a-1) = \frac{c(a+1, z)}{c(a, z)} f(a)g(z-a). \quad (6)$$

Putting $z = a+b+1$, one gets

$$f(a+1)g(b) = \frac{c(a+1, a+b+1)}{c(a, a+b+1)} f(a)g(b+1).$$

Thus

$$f(a+1) = f(a) \frac{g(b+1) c(a+1, a+b+1)}{g(b) c(a, a+b+1)}.$$ 

In general, it can be easily shown by induction from (6) that

$$f(a+n) = f(a) \left( \frac{g(b+1)}{g(b)} \right)^n \prod_{m=0}^{n} \{c(a+m, a+b+m)/c(a+m-1, a+b+m)\}.$$ 

Now, since

$$\sum_{n=0}^{k} f(a+n) = 1, \quad \sum_{n=0}^{k} f(a) \theta^n A(a+n) = 1,$$

we obtain

$$f(a) = 1 / \sum_{n=0}^{k} \theta^n A(a+n),$$

where $\theta$ and $A(x)$ are as defined. It follows that

$$f(x) = A(x) \theta^x \sum_{z=a}^{a+k} A(x) \theta^z.$$ 

Similarly one gets

$$g(y) = B(y) \lambda^y \sum_{u=b}^{b+m} B(y) \lambda^u.$$
ACKNOWLEDGEMENT

Part of the work of this paper was carried out while the authors held the Fellowship of the Summer Research Institute of the Canadian Mathematical Congress.

REFERENCES


