Applied Statistical Science, I
ISBN 1-56072-430-7

ON CHARACTERIZING DISTRIBUTIONS BASED ON LINEARITY OF MAXIMUM ON MINIMUM OF THREE OBSERVATIONS

M. Ahsanullah
Department of Management Sciences
Rider University, Lawrenceville, NJ 08648-3099 USA
and
J. Wesolowski
Mathematical Institute
Warsaw University of Technology 00-661 Warsaw, Poland

ABSTRACT

Let X_1, X_2 and X_3 be a random sample from an absolutely continuous distribution with $W = \max(X_1, X_2, X_3)$ and $V = \min(X_1, X_2, X_3)$. A complete solution of the problem of determining the distribution by linearity of regression of W with respect to V is given. The only possible distributions are of exponential, power function and Pareto type.

Key Words and Phrases: linearity of regression, characterization, exponential distribution, power function and Pareto distribution.

1991 Americal Mathematical subject classification: 60E05, 62E10, 62E15.

1. INTRODUCTION

Fisz (1958) characterized the exponential distribution based on two independent identically distributed random variables X_1 and X_2 by the independence of max (X_1, X_2) -

min (X_1, X_2) and min (X_1, X_2). Since then characterizations of probability distributions based on order statistics are widely investigated. Excellent reviews are given in Azlarov and Volodin (1986) and Galambos and Kotz (1978). Many of the results know in the literature follow from solutions of the integrated Cauchy equation. In this paper we are interested to characterize distributions based on the linearity of regression of $W = \max(X_1, X_2, X_3)$ on $V = \min(X_1, X_2, X_3)$. This leads to new characterization of the exponential, power function and Pareto distributions. The result presented in this paper cannot be obtained using the integrated Cauchy equation.

We will denote $X \in E(\lambda,\gamma)$, shifted exponential distribution. if the pdf of X is of the form

$$f(x) = \lambda \exp(-\lambda (x-\gamma)) \text{ for } x > \gamma, \tag{1.1}$$

where γ and λ (> 0) are real constants.

Denote $X \in PAR(\theta, \mu, \delta)$, Pareto distribution, if the pdf of X is given by

$$f(x) = \frac{\theta(\mu+\delta)^{\theta}}{(x+\delta)^{\theta+1}}, \quad x \geq \mu, \tag{1.2}$$

where θ, μ, and δ are some real constants such that $\theta > 0$ and $\mu + \delta > 0$.

We will denote $X \in POW(\theta, \mu, \nu)$, power function distribution, if the pdf of X is of the following form

$$f(x) = \frac{\theta(\nu - x)^{\theta-1}}{(\nu-\mu)^{\theta}}, \quad \mu \leq x \leq \nu, \tag{1.3}$$

where $\theta > 0$, $-\infty < \mu < \nu < \infty$ are some real constants.

2. MAIN RESULTS

To prove the main theorem, we need the following result which gives a new characterization of the exponential distribution being of independent interest.

Theorem 2.1

Suppose $E(\exp(cX_i)) < \infty$, $i = 1,2,3$ and $E(\exp(cW)|V) = a\exp(cV)$ a.s. for some real constants a and c. Then X's are exponentially distributed with

where $a \in (0,1)$ if $c < 0$ and $a > 1$ if $c > 0$.

Proof:

Let (γ, δ) be the support of F. For any $x \in (\gamma,\delta)$, $E(\exp(cW)|V) = a\exp(cV)$ implies

$$\int_x^d e^{cy}\left[\bar{F}(x) - \bar{F}(y)\right] f(y)\, dy = \frac{a(\bar{F}(x))^2 e^{cx}}{2} \tag{2.1}$$

where $\bar{F} = 1 - F$ and f is the density of F. Since the left hand side of (2.1) is differentiable then we can obtain from (2.1)

$$\int_x^\delta e^{cy} f(y)\, dy = a\bar{F}(x)e^{cx} - \frac{ac(\bar{F}(x))^2 e^{cx}}{2} \tag{2.2}$$

Now again the left hand side of (2.2) is differentiable and f'(x) exists in (γ,δ). Hence upon differentiation of (2.2) for any $x \in (\gamma,\delta)$, we obtain

$$-f(x)e^{cx} = -af(x)e^{cx} + ac\bar{F}(x)e^{cx} - \frac{ac\left\{-2f^2(x)\bar{F}(x)e^{cx} + cf(x)(\bar{F}(x))^2 e^{cx} - f'(x)(\bar{F}(x))^2 e^{cx}\right\}}{2f^2(x)} \tag{2.3}$$

A slight rearrangement allows to write (2.3) as follows:

$$acf'(x)(\bar{F}(x))^2 - ac^2 f(x)(\bar{F}(x))^2 + 4ac\, f^2(x)\bar{F}(x) + 2(1-a)f^3(x) = 0 \tag{2.4}$$

for any $x \in (\gamma,\delta)$. Now denoting $y = \bar{F}$ (i.e. $f = 0\ y'$, $f' = -\ y''$), we get from (2.4) the following second order differential equation in (γ,δ),

$$-y''y^2 + cy'y^2 + 4y'^2 y - \Delta y'^2 = 0 \tag{2.5}$$

where $\Delta = 2\frac{1-a}{ac}$. Substituting u(y) = y' in (2.5) we have

$$-u'y^2 + cy^2 + 4uy - \Delta u^2 = 0. \tag{2.6}$$

Let u(y) = v(y) - βy, where β is a real constant (to be decided). Then (2.6) leads to

$$-v'y^2 + \beta y^2 + cy^2 + 4vy - 4\beta y^2 - \Delta v^2 + 2\Delta\, yv - \Delta\, \beta^2 y^2 = 0 \tag{2.7}$$

We will choose β such that

$$\frac{2(1-a)}{ac}\beta^2 + 3\beta - c = 0.$$

Hence β has to be equal to one of the numbers

$$\beta_{1,2} = c\frac{3a \pm \sqrt{a^2 + 8a}}{4(a-1)}$$

Then (2.7) takes the form

$$-y^2 v' + 2(2 + \Delta\beta) y v - \Delta v^2 = 0 \tag{2.8}$$

which is the Bernoulli equation.

Take $\beta = \beta_1$.

Consider first the trivial solution $v \equiv 0$. Consequently we have a solution of (2. 5) as

$$y = k\ e^{-\beta_1 x}\ ,\ x \in (\gamma, \delta) \tag{2.9}$$

where k is a constant.

Now consider the case of v not equal to zero identically. Then using the standard techniques for the Bernoulli equation (2.8), we obtain

$$-y' = y\frac{\beta_1 Dy^B + C}{Dy^B + A} \tag{2.10}$$

where

$$A = \frac{2(1-a)}{3ac + 4\beta_1(1-1)} = \frac{1}{\beta_1 - \beta_2}$$

$$B = 1 - 4\left(1 + \frac{2\beta_1(1-a)}{ac}\right) = \sqrt{1 + \frac{8}{a}}$$

$$C = \beta_1 A - 1 = \frac{\beta_2}{\beta_1 - \beta_2}$$ and D is a constant.

Note (2.10) implies

$$f(x) = \bar{F}(x)\,\frac{\beta_1 D(\bar{F}(x))^B + C}{D(\bar{F}(x))^B + A}, \quad x \in (\gamma,\delta). \tag{2.11}$$

Consider now two possible cases:

(i) Observe that for $c < 0$, we have $a \in (0,1)$ consequently $\beta_2 < 0 < \beta_1$. Now for $v \equiv 0$, we have $\bar{F}(x) = k\, e^{-\beta_1 x}$, $x \in (\gamma,\delta)$. By the properties of a distribution function, it follows that $\delta = \infty$ and

$$\bar{F}(x) = \exp(-\beta_1(x-\gamma)), \quad x \in (\gamma,\infty) \tag{2.12}$$

If v is not identically equal to zero, then consider (2.11). We will have $A > 0$, $B > 0$ and $C < 0$. Observe that by taking x's less than but close to δ in (2.11), its right hand side becomes negative, which is impossible since on the left hand side we have a density function. In this case the solution of (2.5) is not a tail of a df. Hence the only solution for $c < 0$ is the one given by (2.12).

(ii) Consider now $c > 0$. Then $a > 1$ and $\beta_1 > \beta_2 > 0$. Now similarly as in the first case for $v \equiv 0$, we have the solution (2.12). The integrability condition $(E(\exp(cX) < \infty)$ holds since $\beta_1 > c$. In the case v is not equal to zero identically, consider (2.10). This time $A > 0$, $B > 0$ and $C > 0$. Hence solving (2.10), we obtain

$$y = \left(\frac{\beta_1 D Y^B + c}{\beta_1 D + c}\right)^{\beta_2/c} e^{-\beta_2(x-\gamma)}, \tag{2.13}$$

for $x \in (\gamma,\delta)$. Observe that (2.13) implies $\delta = \infty$. It is sufficient to $x \uparrow \delta$. Further it follows from (2.13) that

$$\bar{F}(x) \geq K e^{-\beta_2 x}, \; x \geq \gamma,$$

for some positive constant K. On the other hand the integrability condition implies that for large x, we have $\bar{F}(x) \leq m e^{-cx}$, where $m = E(\exp(cX))$. If both the above inequalities are to be satisfied then we must have $\beta_2 > c$. Hence $a \leq 4$ and $3a - 4(a-1)\}^2 > a^2 + 8a$, i.e.

$a < 1$ and the last inequality is contradictory. Thus the second solution is impossible and therefore the only solution for $c > 0$ is given by (2.12). □

Now we give the main result characterizing all the distributions that have linearity of regression of maximum on minimum of three observations. The famility coincides with the class obtained by Ferguson (1967).

Theorem 2.2

Let $E(|X_1|) < \infty$ and $E(W|V) = aV + b$ a.s. for some real constants a and b. Then only the following three cases are possible:

1. If $a < 1$, then $X \underline{\underline{d}}$ POW(θ,μ,ν), where $\theta = \frac{3a + \sqrt{a^2 + 8a}}{4(1-a)}$, $\nu = \frac{b}{1-a}$ and $\mu < \nu$ is a real number.

2. If $a > 1$, then $X \underline{\underline{d}}$ PAR (θ,μ,δ), where $\theta = \frac{3a + \sqrt{a^2 + 8a}}{4(a-1)}$, $\delta = \frac{b}{a-1}$ and μ is a real number and $\delta + \mu > 0$.

3. If $a = 1$, then $b > 0$ and $X \underline{\underline{d}}$ $E(\lambda,\gamma)$, where $\lambda = \frac{3}{2b}$ and γ is a real number.

Proof:

First consider the case $a < 1$. Let (μ,ν) be the support of X. Assume that $\nu = \infty$, then by the condition $E(W|V) = aV + b$, we must have $x \le a x + b$ for all $x \in R$, which is impossible, since $a < 1$. Consequently $\nu = \frac{b}{1-a} < \infty$. Define a new rv Z by the relation $Z = -\ln(\nu - X)$, and $W^* = \max(Z_1,Z_2,Z_3)$ and $V^* = \min(Z_1,Z_2,Z_3)$. Then we have

$$E(\exp(-W^*|V^*)) = a\exp(-V^*).$$

Hence by Lemma 1, it follows that Z is an exponential rv $E(\lambda,\gamma)$. Consequently X is POW(θ,μ,ν) with $\theta = \lambda$, ν as defined above and $\mu = \nu - \exp(-\gamma)$.

Consider now a > 1 and let (μ,ν) be the support of X. The condition $E(W|V) = a\,V + b$ implies $\mu > b/(1-a)$. For $\delta = b/(a-1)$ define $Z = \ln(\delta + X)$ and consider W^* and V^*. Then we have $E(\exp(W^*)|V^*) = a\exp(V^*)$. Consequently by lemma 1, Z is exponential $E(\lambda,\gamma)$. Thus X is $PAR(\theta,\mu,\delta)$ with $\theta = \lambda$, δ as defined above and $\mu = \exp(\lambda\gamma) - \delta$.

Now take a = 1. Let (μ,ν) be the support of F.

We have

$$\int_x^{n} (\bar{F}(x) - \bar{F}(y))\, f(y)dy = \frac{(\bar{F}(x))^2\,(x+b)}{2}. \tag{2.14}$$

Differentiating (2.14) w.r.t. x and simplifying we obtain

$$\int_x^{\gamma} y\, f(y)\, dy = \bar{F}(x)\,(x+b) + \frac{(\bar{F}(x))^2}{2\, f(x)}. \tag{2.15}$$

Differentiating both sides of (2.15) w.r.t. x and simplifying we have

$$f'(x)(\bar{F}(x))^2 + 4\, f^2(x)\ \bar{F}(x) - 2\, b\, f^3(x) = 0 \tag{2.16}$$

Let $y - \bar{F}$ (i,e,, $f = -y'$, $f' = -y''$), then (2.16) yields

$$-y''\, y^2 + 4\, y'^2\, y + 2\, b\, y'^3 = 0 \tag{2.17}$$

Substituting $u(y) = y'$ in (2.17), we have

$$-u'\, y^2 + 4\, u\, y + 2\, b\, u^2 = 0, \tag{2.18}$$

since $y' \equiv 0$ is impossible (it yields $f \equiv 0$).

Observe that (2.18) is a kind of the Bernoulli equation, which can be solved and consequently

$$y' = \frac{y^4}{D - \frac{2b}{3} y^3} \tag{2.19}$$

Hence

$$f(x) = \frac{(\bar{F}(x))^4}{\frac{2b}{3}(\bar{F}(x))^3 - D} \tag{2.20}$$

for $x \in (\mu,\nu)$. Observe that the left hand side of (2.20) is always non negative in (μ,ν). However if we tend x to ν, then the right hand side of (2.20) becomes negative unless

$D \leq 0$. Obviously $b \leq 0$ is impossible.

Now we solve (2.19) and obtain

$$-\frac{D}{3y^3} - \frac{2b}{3}\ln(y) = x + d \tag{2.21}$$

for $x \in (\mu, \nu)$, where d is a real constant.

Since $\lim_{x \to \mu} y(x) = 1$, then by (2.21), it follows that $\mu > -\infty$ and

$$-\frac{D}{3} = \mu + d \geq 0. \tag{2.22}$$

Then

$$\bar{F}(x) = \exp\left[\frac{3(x+d)}{2b} + \frac{3(\mu+d)}{2b(\bar{F}(x))^3}\right] \tag{2.23}$$

for $x \in (\mu, \nu)$. Take first $\nu < \infty$. Then upon taking limit for $x \to \nu^-$ in (2.21), we observe that its left hand side tends to infinity, while the right hand side remains finite. Hence $\nu = \infty$. Observe now that by the Markov inquality

$$\bar{F}(x) \leq \frac{m}{x}$$

for sufficiently large positive x's, where $m = E(X)$. Consequently (2.23) implies

$$\bar{F}(x) \geq \exp\left\{-\frac{3(x+d)}{2b} + \frac{3(\mu+d)}{2b\,m^3}x^3\right\} \tag{2.24}$$

for large $x \geq \mu$. Observe that for $\mu + d > 0$, the right hand side of (2.24) is unbounded. Hence $\mu + d = 0$ and now the result follows by (2.23). □

REFERENCES

Azlarov, T.A. and Volodin, N.A. (1986). Characterization Problems Associated with the Exponential Distribution. Springer Verlag, Berlin.

Ferguson, T.S. (1967). On characterizations of distributions by properties of order statistics. Sankhya, A, 29, 265-278.

Fisz, M. (1958). Characterizations of some probability distributions. Skand. Aktuar. 41, 65-70.

Galambos, J. and Kotz, S. (1978). Characterizations of Probability Distributions. Lecture Notes in Mathematics. 675, Springer, Berlin.