Statistical Papers 45, 97-109 (2004)

Statistical Papers

Identification of models using failure rate and mean residual life of doubly truncated random variables

P. G. Sankaran and S. M. Sunoj

Department of Statistics, Cochin University of Science and Technology, Cochin 682 022, India

Received: January 2, 2002; revised version: August 6, 2002

In this paper, we study the relationship between the failure rate and the mean residual life of doubly truncated random variables. Accordingly, we develop characterizations for exponential, Pareto II and beta distributions. Further, we generalize the identities for the Pearson and the exponential family of distributions given respectively in Nair and Sankaran (1991) and Consul (1995). Applications of these measures in the context of length-biased models are also explored.

Key Words and Phrases: *Failure rate, Mean Residual Life, Length biased models*

1. Introduction

Let X be a random variable having absolutely continuous distribution function $F(x)$ in the support of $H = (a, b)$ where a can be $-\infty$ and b can be $+\infty$. Let $f(x)$ denote the probability density function (pdf) of X. When X represents the lifetime of a unit, the failure rate $h(x)$ is defined as

$$h(x) = \frac{f(x)}{1 - F(x)} \qquad (1.1)$$

and the mean residual life (MRL) of X is given by

$$r(x) = E(X - x | X > x). \tag{1.2}$$

It is well known that both failure rate and MRL uniquely determine the distribution. Using the vitality function (conditional mean function)

$$m(x) = E(X | X > x), \tag{1.3}$$

equation (1.2) will be $r(x) = m(x) - x$. Further, the failure rate is related to the MRL by

$$h(x) = \frac{1 + r'(x)}{r(x)} \tag{1.4}$$

where $r'(x)$ is the derivative of $r(x)$ with respect to x. The characterization results based on failure rate and MRL are available in many research papers and monographs, including those of Cox (1962), Galambos and Kotz (1978), Johnson, Kotz and Balakrishnan (1994), Kotz and Shanbhag (1980), Meilijson (1972), Nair and Sankaran (1991) and Navarro, Franco and Ruiz (1998).

Recently, Navarro and Ruiz (1996) defined the failure rate function and the conditional expectation for doubly truncated random variables that generalizes respectively (1.1) and (1.3). Further, it is shown that both the generalized failure rate (GFR) and the conditional expectation uniquely determine the distribution of X. For the properties and the estimates of GFR we refer to Navarro and Ruiz (1996). As the GFR and the conditional expectation (mean function) for doubly truncated random variables are generalizations over the left and right truncation of the values of the variables, in the physical sense it gives the instantaneous failure rate and the expected life of the component/system when it works between two time points. In human life length studies, it provides the rate of failure and expected life between two ages, which is important in studying the useful period of human life. Further, doubly truncated measures are also applicable to engineering systems when the observations are measured after it starts operating and before it fails.

In the present paper, we study the relationship between the GFR and the doubly truncated MRL function. Accordingly, we develop characterization results for the Pearson family and exponential family of distributions using the relationship between the GFR and the doubly truncated mean function. The results to be developed are very useful in providing characteristic properties of lifetime distributions which enable us to identify the particular models. They also form an essential tool for the estimation of parameters. Finally, some structural relationships are established for length-biased models, which is a generalization over the left and right truncation models proved in the literature.

2. Basic Concepts

For a continuous random variable X, Navarro and Ruiz (1996) defined the GFR function as

$$h_1(x_1,x_2)=\frac{f(x_1)}{F(x_2)-F(x_1)} \qquad (2.1)$$

and

$$h_2(x_1,x_2)=\frac{f(x_2)}{F(x_2)-F(x_1)} \qquad (2.2)$$

for all $(x_1,x_2)\in E=\{(x_1,x_2)\in R^2, F(x_1)<F(x_2)\}$.

Obviously when $x_2\to\infty$, (2.1) reduces to the failure rate (1.1). The MRL for a doubly truncated random variable is defined by

$$r_d(x_1,x_2)=E(X-x_1|x_1<X<x_2)$$

which can be represented as

$$r_d(x_1,x_2)=\frac{1}{F(x_2)-F(x_1)}\int_{x_1}^{x_2}(t-x_1)dF(t). \qquad (2.3)$$

Equation (2.3) measures the expected remaining life of a unit that has the age x_1 and is broken before age x_2. It is easy to show that

$$h_1(x_1,x_2)r_d(x_1,x_2)=1+\frac{dr_d(x_1,x_2)}{dx_1} \qquad (2.4)$$

Also from (2.2) and (2.3), we get

$$h_2(x_1,x_2)=\frac{r_d'(x_1,x_2)}{x_2-x_1-r_d(x_1,x_2)} \qquad (2.5)$$

where the prime denote the derivative with respect to x_2.

From (2.3) it is easy to see that

$$r_d(x_1,x_2)=m(x_1,x_2)-x_1 \tag{2.6}$$

where $m(x_1,x_2)$ is doubly truncated mean function given in Ruiz and Navarro (1996) and the distribution is completely specified whenever $m(x_1,x_2)$ is known.

The following section provides characterization theorems for various lifetime models using the relationship among the GFR, $r_d(x_1,x_2)$ and $m(x_1,x_2)$.

3. Characterization Results

Now we present a characterization result for the exponential distribution, based on GFR.

Theorem 3.1: A necessary and sufficient condition for the distribution of X is exponential with mean θ^{-1} is that

$$h_1(x_1,x_2)-h_2(x_1,x_2)=\theta \tag{3.1}$$

for all $(x_1,x_2)\in E$.

Proof: Suppose that (3.1) holds. Then from (2.1) and (2.2) we have

$$f(x_1)-f(x_2)=\theta(F(x_2)-F(x_1))$$

which gives

$$f(x_1)+\theta F(x_1)=f(x_2)+\theta F(x_2). \tag{3.2}$$

The equation (3.2) is satisfied only when

$$f(x)+\theta F(x)=p\text{, a constant.}$$

Or

$$\frac{dF(x)}{dx} + \theta F(x) = p \tag{3.3}$$

Using the boundary condition $\lim_{x \to b} F(x) = 1$, equation (3.3) yields

$$F(x) = 1 - e^{-\theta x} \, .$$

The proof of the converse is direct.

Remark 3.1: When $x_2 \to b, h_2(x_1, x_2) \to 0$ and $h_1(x_1, x_2) \to h(x_1)$, which reduces to $h(x_1) = \theta$, a well-known characteristic property of the exponential distribution.

Remark 3.2: Navarro and Ruiz (1996) provide non-parametric product limit estimatoıs for GFR functions. The relationship (3.1) provides an estimate for θ under censored situations.

Theorem 3.2: Assume that the derivative of $f(x)$ w.r.to x exists. The relationship

$$r_d(x_1, x_2) = \theta^{-1}\left[(x_1 - x_2)h_2(x_1, x_2) + 1\right] \tag{3.4}$$

is satisfied for all $(x_1, x_2) \in E$ and $\theta > 0$ if and only if the distribution of X is exponential with mean θ^{-1}.

Proof: Suppose that (3.4) is satisfied. Using (2.2) and (2.3), (3.4) becomes

$$\theta \int_{x_1}^{x_2} (t - x_1) dF(t) = (x_1 - x_2) f(x_2) + F(x_2) - F(x_1)$$

or

$$\theta\left[(x_2 - x_1)F(x_2) - \int_{x_1}^{x_2} F(t)dt\right] = (x_2 - x_1) f(x_2) + F(x_2) - F(x_1). \tag{3.5}$$

Differentiating (3.5) with respect to x_2, we get

$$\theta(x_2 - x_1)f(x_2) = (x_1 - x_2)f'(x_2)$$

or

$$\frac{f'(x_2)}{f(x_2)} = -\theta$$

where prime denote the derivative with respect to x_2. Thus the distribution of X is exponential. Conversely, for the exponential distribution with mean θ^{-1}, we have

$$r_d(x_1, x_2) = \frac{1}{\theta} + \frac{(x_1 - x_2)e^{-\theta x_2}}{\left(e^{-\theta x_1} - e^{-\theta x_2}\right)}$$

and

$$h_2(x_1, x_2) = \frac{\theta e^{-\theta x_2}}{e^{-\theta x_1} - e^{-\theta x_2}}$$

which satisfies (3.4).

The following result provides characterizations for exponential, Pareto II and beta distributions using the relationship between GFR and MRL.

Theorem 3.3: Assume that the derivative of $f(x)$ w.r.to x exists. The identity

$$r_d(x_1, x_2) = \lambda\left[(r(x_1))^2 h_1(x_1, x_2) - (r(x_2))^2 h_2(x_1, x_2)\right] + \lambda(x_1 - x_2)r(x_2)h_2(x_1, x_2) \tag{3.6}$$

characterize

(i) the exponential distribution for $\lambda = 1$

(ii) the Pareto distribution with pdf

$$f(x) = ab(1 + ax)^{-b-1}, a > 0, b > 1, x > 0 \tag{3.7}$$

for $0 < \lambda < 1$, and

(iii) the beta distribution with pdf

$$f(x) = cd(1 - cx)^{d-1}, c, d > 0, 0 < x < c^{-1} \tag{3.8}$$

for $\lambda > 1$.

Proof: Assume that the relationship (3.6) holds. Substituting (2.1), (2.2) and (2.3) in (3.6), we get

$$\int_{x_1}^{x_2} (t - x_1) dF(t) = \lambda\left[(r(x_1))^2 f(x_1) - (r(x_2))^2 f(x_2)\right] + \lambda(x_1 - x_2) r(x_2) f(x_2)$$

or

$$(x_2 - x_1)F(x_2) + \int_{x_1}^{x_2} [1 - F(t)]dt = \lambda\left[(r(x_1))^2 f(x_1) - (r(x_2))^2 f(x_2)\right] + \lambda(x_1 - x_2) r(x_2) f(x_2) \quad (3.9)$$

Differentiating (3.9) with respect to x_2, we get

$$(x_2 - x_1) f(x_2) = \lambda\left[-2r(x_2)r'(x_2)f(x_2) - (r(x_2))^2 f'(x_2)\right] + \lambda\left[(x_1 - x_2) f'(x_2) r(x_2) - f(x_2) r(x_2) + (x_1 - x_2) f(x_2) r'(x_2)\right] \quad (3.10)$$

where prime represent the derivative with respect to x_2. Equating the coefficients of either x_1 or x_2 on both sides of (3.10), we have

$$-f(x_2) = \lambda\left[f'(x_2) r(x_2) + f(x_2) r'(x_2)\right]$$

or

$$-f(x_2) = \lambda \frac{d\left[f(x_2) r(x_2)\right]}{dx_2}. \quad (3.11)$$

On integration (3.11) provides

$$1 - F(x) = \lambda f(x) r(x)$$

or

$$h(x) r(x) = \frac{1}{\lambda}. \quad (3.12)$$

From Theorem 1 in Ruiz and Navarro (1994), (3.12) provides characterization for exponential distribution when $\lambda = 1$, for Pareto model (3.7) when $0 < \lambda < 1$ and for Beta distribution (3.8) when $\lambda > 1$.

The converse part can be easily verified by direct calculations.

Remark 3.3: As $x_2 \to b$, the Theorem 3.3 reduces to the result of Mukherjee and Roy (1986).

We next prove a general characterization theorem for the Pearson family of distributions based on GFR and $m(x_1, x_2)$.

Definition 3.1: Let X be a continuous random variable as defined in Section 1. Then the distribution of X belongs to the Pearson family of distributions if the pdf $f(x)$ of X is differentiable and $f(x)$ satisfies the differential equation

$$\frac{d \log f(x)}{dx} = -\frac{x + d_1}{b_0 + b_1 x + b_2 x^2} \tag{3.14}$$

where b_0, b_1, b_2 and d_1 are real constants.

Theorem 3.4: The distribution of X belongs to the Pearson family if and only if

$$m(x_1, x_2) = \mu - (a_0 + a_1 x_2 + a_2 x_2^2) h_2(x_1, x_2) + (a_0 + a_1 x_1 + a_2 x_1^2) h_1(x_1, x_2)$$

where

$$a_i = \frac{b_i}{1 - 2b_2}, i = 0,1,2$$

and

$$\mu = E(X).$$

Proof: The theorem is proved by proceeding the similar steps in Nair and Sankaran (1991).

Remark 3.4: When $x_2 \to b$, Theorem 3.4 reduces to the Theorem 2.1 given in Nair and Sankaran (1991).

We now prove a characterization result for exponential family of distributions.

Theorem 3.5: A necessary and sufficient condition for the distribution of X belong to the one parameter exponential family is that

$$m(x_1, x_2) = \mu + \frac{1}{g'(\theta)} \frac{\partial \log[F(x_2) - F(x_1)]}{\partial \theta} \tag{3.19}$$

where $\mu = E(X)$.

Proof: The proof is analogous to that of Consul (1995).

Remark 3.5: Theorem 3.5 reduces to the Theorem 1 of Consul (1995) when $x_2 \to b$.

In the following section, we examine the use of GFR and $m(x_1, x_2)$ in the context of length biased models

4. Length biased models

The statistical interpretation of the length biased distribution was originally identified by Cox (1962) in the context of renewal theory. Assume that the population of failure times is distributed according to $f(x)$, the probability of selection of any individual in the population proportional to its life length x, then the density function of the life length for the sampled component of random variable Y has the form

$$g(x) = \frac{xf(x)}{\mu}, x > 0 \tag{4.1}$$

where $\mu = E(X) < \infty$, which is the length biased form (see also Blumenthal (1967), Scheaffer (1972)). A detailed survey of literature on applications of length biased models we refer to Rao (1965), Patil and Rao (1977) and Gupta and Kirmani (1990).

In the present section, we examine the structural relationships between the random variables of X and Y in the context of GFR and $m(x_1, x_2)$ defined in Navarro and Ruiz (1996) for doubly truncated distributions. By definition, probability that the sampled component Y lies between x_1 and x_2 is given by

$$P(x_1 < Y < x_2) = \int_{x_1}^{x_2} g(t)dt$$

$$= \frac{1}{\mu}\int_{x_1}^{x_2} t\,dF(t)$$

i.e.,
$$P(x_1 < Y < x_2) = \frac{m(x_1,x_2)}{\mu} P(x_1 < X < x_2). \quad (4.2)$$

Taking logarithm on both sides and differentiating partially with respect to x_1 and x_2 respectively and using relationships

$$\frac{\partial m(x_1,x_2)}{\partial x_1} = \big(m(x_1,x_2) - x_1\big)h_1(x_1,x_2) \qquad (4.3)$$

and

$$\frac{\partial m(x_1,x_2)}{\partial x_2} = \big(x_2 - m(x_1,x_2)\big)h_2(x_1,x_2) \qquad (4.4)$$

we obtain the GFR functions of Y as

$$k_i(x_1,x_2) = \frac{x_i h_i(x_1,x_2)}{m(x_1,x_2)}, i = 1,2. \qquad (4.5)$$

where $k_i(x_1,x_2), i = 1,2$ are defined respectively

$$k_1(x_1,x_2) = -\frac{\partial}{\partial x_1}\log P(x_1 < Y < x_2)$$

and

$$k_2(x_1,x_2) = \frac{\partial}{\partial x_2}\log P(x_1 < Y < x_2).$$

From (4.5), it is easy to find the ratio

$$\frac{k_1(x_1,x_2)}{k_2(x_1,x_2)} = \frac{x_1 h_1(x_1,x_2)}{x_2 h_2(x_1,x_2)}. \qquad (4.6)$$

Remark 4.1: Equation (4.2) is a more general expression, which can be applied to both left and right truncation when the truncation values $x_1 \to a$ and $x_2 \to b$ respectively. Further, since $m(x_1, x_2)$ uniquely determines the distribution function (Ruiz and Navarro (1996)), the ratios $\frac{P(x_1 < Y < x_2)}{P(x_1 < X < x_2)}$ and $\frac{k_t(x_1, x_2)}{h_t(x_1, x_2)}$ also characterizes the distribution function. These characterizing expressions purely depend on the structural forms of $m(x_1, x_2)$ for different distributions (see Table 1 of Ruiz and Navarro (1996)).

Remark 4.2: When $x_2 \to b$, equation (4.2) reduces to

$$\frac{P(Y > x_1)}{P(X > x_1)} = \frac{m(x_1)}{\mu} \tag{4.7}$$

where $m(x) = E(X \mid X \geq x)$ called the vitality function. Following Remark 4.1, (4.7) also uniquely determines different models with respect to various forms of $m(x)$. In particular, (4.7) provides characterizations to exponential, Pareto II, beta models when the ratio $\frac{P(Y > x_1)}{P(X > x_1)} = 1 + cx, c > 0$ and according as $\mu c - 1 \gtreqless 0$ (Gupta and Keating (1986), Navarro et. al (2001)). Further, when $m(x) = k + q(x)h(x)$ where k is a constant and $q(x)$ is a real function in (a, b), equation (4.7) characterizes gamma, generalized beta and Pearson family of distributions (see Ruiz and Navarro (1994)).

Remark 4.3: When $x_2 \to b$, equation (4.5) becomes

$$\frac{k(x_1)}{h(x_1)} = \frac{x_1}{m(x_1)} \tag{4.8}$$

where $k(x)$ is the failure rate of Y in the right truncation case. Equation (4.8) also provides characterizations to different models which depends on the functional forms of $m(x)$ (see Gupta and Keating (1986), Ruiz and Navarro (1994) and Navarro et. al (2001)).

References

1. Blumenthal S (1967) Proportional sampling in life length studies. Technometrics, **9**, 205-218.
2. Consul PC (1995) Some characterizations for the exponential class of distributions. IEEE Trans. Reliability, **44**, 403-407.
3. Cox DR (1962) Renewal Theory. Frome and Long, Butler and Tanner Ltd.
4. Galambos J, Kotz S (1978) Characterization of probability distributions. Lecturer notes in mathematics 675. Springer Verlag, New York.
5. Gupta RC, Keating JP (1986) Relations for reliability measures under length biased sampling. Scand. J. Statist., **13**, 49-56.
6. Gupta RC, Kirmani SNUA (1990) The role of weighted distributions in stochastic modeling. Commun. Statist. – Theor. Meth., **19**(9), 3147-3162.
7. Johnson NL, Kotz S, Balakrishnan N (1994) Continuous Univariate Distributions. John Wiley and Sons, **1**, New York.
8. Kotz S, Shanbhag DN (1980) Some new approaches to probability distributions. Adv. Applied Prob., **12**, 903-921.
9. Meilijson, I (1972) Limiting properties of the mean residual life function. Ann. Math. Statist., **43**, 354-357.
10. Nair NU, Sankaran PG (1991) Characterization of Pearson family of distributions. IEEE Trans. Reliability, **40**, 75-77.
11. Patil GP, Rao CR (1977) The weighted distribution: A survey of their applications, In Applications of Statistics. P.R.Krishnaiah (ed.), North Holland Publ. Co., 383-405.
12. Navarro J, Franco M, Ruiz JM (1998) Characterization through moments of the residual life and conditional spacing. Sankhya A, **60**, 36-48.
13. Navarro J, Ruiz JM (1996) Failure rate functions for doubly-truncated random variables. IEEE Trans. Reliability, **45**, 685-690.
14. Navarro J, del Aguila Y, Ruiz JM (2001) Characterization through reliability measures from weighted distributions. Statistical Papers, **42**, 395-402.
15. Rao CR (1965) On discrete distributions arising out of methods of ascertainments. In Classical and Contagious Discrete Distributions, G.P.Patil (ed.), Pergamon Press and Statistical Publishing Society, Calcutta, 320-332, Also reprinted in Sankhya A, **27**, 311-324.
16. Ruiz JM, Navarro, J (1994) Characterization of distributions by relationships between failure rate and mean residual life. IEEE Trans. Reliability, **43** (4), 640-644.

17. Ruiz JM, Navarro J (1996) Characterizations based on conditional expectations of the doubled truncated distribution. Ann. Inst. Statist. Math., **48**, 563-572.
18. Scheaffer RL (1972) Size-biased sampling. Technometrics, **14**, 635-644.