Some Characterization Results on Generalized Pareto Distribution Based on Progressive Type-II Right Censoring

M. Hashemi, M. Asadi

Department of Statistics, University of Isfahan, Isfahan, Iran.

Abstract. The progressive censoring scheme is a method of data collecting in reliability and life testing which has been of intensified interest in recent years. In the present paper, we prove some characterization results on generalized Pareto distribution based upon the independency and expected values of some functions of progressive type-II right censored order statistics.

1 Introduction

In reliability and survival analysis there are several scenarios in which units that are subject to test are removed from the experiment before failure. The progressive censoring scheme is an important scheme of such scenarios. The progressive censoring allows the experimenter to remove surviving units from a life test at various stages during the experiment. A saving of costs and of time may be the consequence of such a sampling scheme. The idea of progressive censoring goes back almost half a century. However, researchers have shown considerable attention, on this scheme of sampling, in the past few years.

Key words and phrases: Mean residual life function, order statistics, survival function, truncated expectation.
We refer the reader to Balakrishnan and Aggarwala (2000) and references therein for a comprehensive discussion on progressive censoring scheme.

Let \( X_1, X_2, \ldots, X_n \) denote the failure times of \( n \) independent and identically distributed units which are placed on a lifetime test. Suppose further that the following censoring scheme \((R_1, R_2, \ldots, R_m)\) is fixed: Following the first failure time of one of the units, a number of \( R_1 \) surviving units are randomly selected and removed (censored) from the test; at the second failure time \( R_2 \) surviving units are selected at random and taken out of the experiment and so on; finally, at the time of the \( m \)th failure, the experiment is finished with \( R_m \) surviving units. The ordered observed failure times are denoted by \( X^{(R_1, R_2, \ldots, R_m)}_{1:m:n}, X^{(R_1, R_2, \ldots, R_m)}_{2:m:n}, \ldots, X^{(R_1, R_2, \ldots, R_m)}_{m:m:n} \) and called the progressive type II right censored order statistics of size \( m \) from a sample of size \( n \) with progressive censoring scheme \((R_1, R_2, \ldots, R_m)\). It is clear that \( n = m + R_1 + R_2 + \ldots + R_m \). If we assume that \( X_i \)'s have a common absolutely continuous distribution function (d.f) \( F \) with probability density function (p.d.f) \( f \), then the joint probability density function of the progressive type II right censored failure times \( X^{(R_1, R_2, \ldots, R_m)}_{1:m:n}, X^{(R_1, R_2, \ldots, R_m)}_{2:m:n}, \ldots, X^{(R_1, R_2, \ldots, R_m)}_{m:m:n} \) (which, in the sequel, we denote for simplicity by \( X_{1:m:n}, X_{2:m:n}, \ldots, X_{m:m:n} \)) is given by

\[
    f_{X_{1:m:n}, \ldots, X_{m:m:n}}(x_1, \ldots, x_m) = c \prod_{i=1}^{m} f(x_i)(1 - F(x_i))^{R_i}, \quad x_1 < x_2 < \ldots < x_m.
\]

where \( c = \prod_{j=1}^{m} \gamma_j \) with \( \gamma_j = n - \sum_{k=1}^{j-1} R_k - j + 1, \ \gamma_1 = n \) and \( n = m + \sum_{j=1}^{m} R_j \). See, for example, Balakrishnan and Aggarwala (2000).

In particular, the model of ordinary order statistics is contained in the above set-up by choosing \( m = n \) and \( R_1 = R_2 = \ldots = R_m = 0 \), where no withdrawals are made.

One can show that (see Kamps and Cramer (2001)) the marginal d.f and p.d.f of \( X_{r:m:n}, 1 \leq r \leq m \), are given, respectively, by

\[
    F_{X_{r:m:n}}(x_r) = 1 - c_{r-1} \sum_{i=1}^{r} \frac{a_i(r)}{\gamma_i} (1 - F(x_r))^{\gamma_i}, \quad 1 \leq r \leq m.
\]

and

\[
    f_{X_{r:m:n}}(x_r) = c_{r-1} f(x_r) \sum_{i=1}^{r} a_i(r)(1 - F(x_r))^{\gamma_i-1}, \quad 1 \leq r \leq m.
\]
where \( c_{r-1} = \prod_{j=1}^{r} \gamma_j \) and \( a_i(r) = \prod_{j=1}^{r} \frac{1}{\gamma_j - \gamma_i} \), \( 1 \leq i \leq r \leq m \), \( m \geq 2 \) and the empty product \( \prod_{\phi} \) is defined to be 1.

Also the joint density function of \( X_{r:m:n} \) and \( X_{s:m:n} \), \( 1 \leq r < s \leq m \), is given as

\[
f_{X_{r:m:n},X_{s:m:n}}(x_r, x_s) = c_{s-1} \left\{ \sum_{i=r+1}^{s} a_i^{(r)}(s) \left( \frac{1 - F(x_s)}{1 - F(x_r)} \right)^{\gamma_i} \right\} \left\{ \sum_{i=1}^{r} a_i(r) (1 - F(x_r))^{\gamma_i} \right\} \times f(x_r) \frac{f(x_s)}{1 - F(x_r) - F(x_s)}, \quad x_r \leq x_s, \quad 1 \leq r \leq m. \tag{3}\]

where

\[
a_i^{(r)}(s) = \prod_{j=r+1}^{s} \frac{1}{\gamma_j - \gamma_i}, \quad r + 1 \leq i \leq s.
\]

In this paper, we give some characterization results on generalized Pareto distribution (GPD) based on progressive type-II censoring scheme. The GPD is a flexible statistical model which includes the exponential distribution, the Pareto distribution and the power distribution. We say that a distribution function \( F \) is GPD if its survival function \( \bar{F} = 1 - F \) is given by

\[
\bar{F}(x) = \left( \frac{b}{ax + b} \right)^{\frac{1}{a} + 1}, \quad x \geq 0, \tag{4}\]

where \( a > -1 \) and \( b > 0 \). The GPD includes, depending on the values of \( a \), the exponential distribution (when \( a \to 0 \)), the Pareto distribution (when \( a > 0 \)) and the power distribution (when \(-1 < a < 0\)). In particular when \( a = -\frac{1}{2} \) the distribution is uniform. (Note that for \(-1 < a < 0\) the distribution is bounded above.)

Hall and Wellner (1981) proved that the GPD is the only family of distributions that has a linear mean residual function, where the mean residual life function is defined as \( m(x) = E(X - x | X > x) \). Oakes and Dasu (1990) obtained a characterization result of GPD of the type of lack of memory property of the exponential distribution as follows: Let \( \tilde{F}(x) \) be an absolutely continuous survival function with \( \tilde{F}(0) = 1 \) and finite mean. Suppose that the survival function \( \tilde{F}(x) \) satisfies the following equation

\[
\tilde{F}(\theta(x)y + x) = \tilde{F}(y)\tilde{F}(x), \quad x, y \geq 0, \tag{5}\]
where \( \theta(x) = \frac{m(x)}{m(0)} \). Then \( \theta \) is linear and hence \( \bar{F} \) is GPD. Using the result of Oakes and Dasu (1990), Asadi et al. (2001) extended many characterization results of exponential distribution, based on order statistics and record values, to the case of GPD. Recently Marohn (2002) gave some characterization results on GPD based on the independence properties of spacings of progressive type-II censored order statistics.

The aim of the present note is to give some new characterizations results on GPD based on progressive type-II right censored order statistics. The results are given in the following section.

2 Main Results

In this section we give our main results. First we give the following theorem.

**Theorem 2.1.** Let \( X_{1:m:n}, X_{2:m:n}, \ldots, X_{m:m:n} \) be non-negative progressively type-II right censored order statistics from a population with absolutely continuous distribution function \( F \). Furthermore, assume that \( \theta(x) = \frac{m(x)}{m(0)} \), where \( m \) denotes the mean residual life function of \( F \). Then the random variables

\[
\frac{X_{s:m:n} - X_{r:m:n}}{\theta(X_{r:m:n})}
\]

and \( X_{r:m:n}, 1 \leq r < s \leq n \), are independent if and only if \( F \) is GPD.

**Proof.** First we prove the ‘only if’ part of theorem. Note that for \( u, t \geq 0 \) we have

\[
P\left\{ \frac{X_{s:m:n} - X_{r:m:n}}{\theta(X_{r:m:n})} \leq u | X_{r:m:n} > t \right\} =
\]

\[
P(t < X_{r:m:n} < X_{s:m:n} \leq X_{r:m:n} + u\theta(X_{r:m:n}))
\]

\[
P(X_{r:m:n} > t)
\]

From (2) and (3) we get

\[
P(X_{r:m:n} > t) = c_{r-1} \int_{t}^{\infty} \sum_{i=1}^{r} a_i(r)(\bar{F}(x))^{\gamma_i-1}dF(x),
\]

\[
(6)
\]

\[
(7)
\]
and

\[ P(t < X_{r:m:n} < X_{s:m:n} \leq X_{r:m:n} + u\theta(X_{r:m:n})) = c_{s-1} \int_t^\infty \int_x^{x+u\theta(x)} \left\{ \sum_{i=r+1}^s a_i^{(r)}(s) \left( \frac{\bar{F}(y)}{\bar{F}(x)} \right)^{\gamma_i} \right\} dF(y) dF(x) \]

\[ = c_{s-1} \int_t^\infty \sum_{i=1}^r a_i(r)(\bar{F}(x))^{\gamma_i} \int_x^{x+u\theta(x)} \left( \frac{\bar{F}(y)}{\bar{F}(x)} \right)^{\gamma_i} dF(y) dF(x) \]

The inside integral in (8) is equal to

\[ \sum_{i=r+1}^s a_i^{(r)}(s) \int_x^{x+u\theta(x)} \left( \frac{\bar{F}(y)}{\bar{F}(x)} \right)^{\gamma_i} dF(y) = \sum_{i=r+1}^s a_i^{(r)}(s) \int_{\bar{F}(x)}^{\bar{F}(x+u\theta(x))} \nu^{\gamma_i-1} d\nu \]

\[ = \sum_{i=r+1}^s a_i^{(r)}(s) \gamma_i \left( 1 - \left( \frac{\bar{F}(x+u\theta(x))}{\bar{F}(x)} \right)^{\gamma_i} \right) \]

Hence the right hand side of (8) is equal to

\[ c_{s-1} \int_t^\infty \sum_{i=1}^r a_i(r)(\bar{F}(x))^{\gamma_i-1} H(x, u) dF(x). \] (10)

Based on the independence assumption of theorem, the conditional probability in (6) is only a function of u, say, \( G(u) \). Hence, by substituting (7) and (10) in (6), we have

\[ G(u) = \frac{c_{s-1}}{c_{r-1}} \int_t^\infty \sum_{i=1}^r a_i(r)(\bar{F}(x))^{\gamma_i-1} H(x, u) dF(x) \cdot \frac{dF(x)}{dF(x)} \]

On differentiating the last equality with respect to \( t \), we obtain

\[ G(u) = \frac{c_{s-1}}{c_{r-1}} \sum_{i=r+1}^s a_i^{(r)}(s) \gamma_i \left( 1 - \left( \frac{\bar{F}(t+u\theta(t))}{\bar{F}(t)} \right)^{\gamma_i} \right). \] (11)
Differentiating again equation (11), with respect to $t$, gives

$$\frac{d}{dt} r(t + u\theta(t)) = r(t), \quad \text{(12)}$$

where $\theta'(t)$ is the derivative of $\theta(t)$ with respect to $t$ and $r(t) = \frac{f(t)}{F(t)}$ denotes the hazard rate corresponding to $F$.

Now, on taking the integration of both sides of (12) on the interval $(0, x)$, we get that

$$\int_0^x (1 + u\theta'(t)) r(t + u\theta(t)) \, dt = \int_0^x r(t) \, dt,$$

or equivalently

$$\bar{F}(u)\bar{F}(x) = \bar{F}(x + u\theta(x)).$$

Hence the result follows from the result of Oakes and Dasu (1990). That is, the underlying distribution is GPD. The ‘if’ part of the theorem follows easily from equation (11) and the fact that the mean residual life function of GPD is linear. This completes the proof of the theorem.

**Corollary 2.1.** Let $X_1, X_2, \ldots, X_n$ be independent non-negative random variables from a continuous distribution function $F$. Let also $X_{1:n}, X_{2:n}, \ldots, X_{n:n}$ denote the corresponding order statistics. (That is, in the progressively type-II right censored order statistics model $m = n$ and $R_1 = R_2 = \ldots = R_m = 0$). Then random variables

$$\frac{X_{s:n} - X_{r:n}}{\theta(X_{r:n})}$$

and $X_{r:n}, 1 \leq r < s \leq n$, are independent if and only if $F$ is GPD.

**Theorem 2.2.** Let $X_{1:n}, X_{2:n}, \ldots, X_{m:n}$ be non-negative progressively type-II right censored order statistics from a population with absolutely continuous distribution function $F$. Furthermore, assume that $\theta(x) = \frac{m(x)}{m(0)}$, where $m$ denotes the mean residual life function of $F$. Then the distribution function $F$ is GPD if and only if for $t > 0$ and $s = 1, 2, \ldots, m$,

$$\frac{X_{s:n} - t}{\theta(t)} | X_{1:n} > t \overset{d}{=} X_{s:n} \quad \text{(13)}$$

where $d$ stands for distribution.
Proof. First we prove the ‘if’ part of theorem. To this end, note that for $u > 0$, 
\[ P(X_{1:m:n} > u) = (\bar{F}(u))^n. \] (14)
If $s = 1$, then we have (13)
\[ P\left(\frac{X_{1:m:n} - t}{\theta(t)} > u | X_{1:m:n} > t\right) = P(X_{1:m:n} > u), \]
or
\[ P(X_{1:m:n} > t + u\theta(t)) = P(X_{1:m:n} > t)P(X_{1:m:n} > u). \]
This implies, using (14), that for all $t > 0$ and $u > 0$,
\[ \bar{F}(t + u\theta(t)) = \bar{F}(t)\bar{F}(u). \]
That is, $F$ is GPD. Now suppose that $2 \leq s \leq m$. Then we have
\[ P\left(\frac{X_{s:m:n} - t}{\theta(t)} < u | X_{1:m:n} > t\right) = P(X_{s,m:n} < u). \] (15)
The right side of (15) is
\[ P(X_{s,m:n} < u) = c_{s-1} \int_0^u \sum_{i=1}^S a_i(s)(\bar{F}(x))^{\gamma_i - 1}dF(x) \]
\[ = c_{s-1} \int_0^{1-\bar{F}(u)} \sum_{i=1}^S a_i(s)(1 - \nu)^{\gamma_i - 1}d\nu, \] (16)
and the left side of (15) is
\[ P\left(\frac{X_{s,m:n} - t}{\theta(t)} < u | X_{1:m:n} > t\right) = \frac{P(t < X_{1:m:n} < X_{s,m:n} \leq t + u\theta(t))}{P(X_{1:m:n} > t)} \]
\[ = \frac{1}{(\bar{F}(t))^n} \int_t^{t+u\theta(t)} \int_t^y c_{s-1} \]
\[ \bar{F}(y) \left\{ \sum_{i=2}^S a_i^{(1)}(s)(\frac{\bar{F}(y)}{\bar{F}(x)})^{\gamma_i} (\bar{F}(x))^{n-1}dF(x) \right\} d\bar{F}(y) \]
\[ = \frac{1}{(\bar{F}(t))^n} \int_t^{t+u\theta(t)} \frac{c_{s-1}}{\bar{F}(y)} \]
\[ \left\{ \int_t^y \sum_{i=2}^S a_i^{(1)}(s)(\frac{\bar{F}(y)}{\bar{F}(x)})^{\gamma_i} (\bar{F}(x))^{n-1}dF(x) \right\} dF(y). \] (17)
The inside integral in (17) is equal to
\[ \sum_{i=2}^{s} \frac{a_i^{(1)}(s)}{n - \gamma_i} [(\bar{F}(t))^n - (\bar{F}(y))^n]. \]

If we substitute this in (17), we get
\[
\begin{align*}
&c_{s-1} \int_{t}^{t+u\theta(t)} \frac{\bar{F}(t)}{F(y)} \sum_{i=2}^{s} \frac{a_i^{(1)}(s)}{n - \gamma_i} [(\bar{F}(y))^n - (\bar{F}(y))^n] dF(y) \\
&= c_{s-1} \int_{0}^{1} \frac{F(t+u\theta(t))}{F(t)} (\sum_{i=2}^{s} \frac{a_i^{(1)}(s)}{n - \gamma_i} [(1 - \nu)^{\gamma_i-1} - (1 - \nu)^{n-1}] d\nu).
\end{align*}
\]

On the other hand, for \( i = 2, 3, ..., s \), we have
\[ \frac{a_i^{(1)}(s)}{n - \gamma_i} = a_i(s). \]

Hence
\[
\begin{align*}
&c_{s-1} \int_{0}^{1} \frac{F(t+u\theta(t))}{F(t)} (\sum_{i=2}^{s} a_i(s)(1 - \nu)^{\gamma_i-1} d\nu) \\
&= c_{s-1} \int_{0}^{1} \frac{F(t+u\theta(t))}{F(t)} (\sum_{i=2}^{s} a_i(s)(1 - \nu)^{\gamma_i-1} d\nu) \\
&+ c_{s-1} \int_{0}^{1} \frac{F(t+u\theta(t))}{F(t)} (-\sum_{i=2}^{s} a_i(s)(1 - \nu)^{n-1} d\nu).
\end{align*}
\]

By taking \( a_1(s) = -\sum_{i=2}^{s} a_i(s) \) and \( \gamma_1 = n \), we have
\[
P\left( \frac{X_{s:n:m:n} - t}{\theta(t)} < u \mid X_{1:n:m:n} > t \right) = c_{s-1} \int_{0}^{1} \frac{F(t+u\theta(t))}{F(t)} (\sum_{i=1}^{s} a_i(s)(1 - \nu)^{\gamma_i-1} d\nu). \tag{18}
\]

Hence taking into account (15), (16) and (18) we have
\[ \frac{\bar{F}(t+u\theta(t))}{F(t)} = \bar{F}(u). \]

That is, \( F \) is GPD. The proof of the ‘only if’ part of the theorem is straightforward and hence is omitted. This completes the proof of the theorem.
Theorem 2.2 gives the following corollary which is proved recently by Asadi and Bayramoglu (2006).

**Corollary 2.2.** Let $X_1, X_2, \ldots, X_n$ be independent non-negative random variables from a continuous distribution function $F$. Let also $X_{1:n}, X_{2:n}, \ldots, X_{n:n}$ denote the corresponding order statistics. (That is, in the progressively type-II right censored order statistics model $m = n$ and $R_1 = R_2 = \ldots = R_m = 0$). Then $F$ is GPD if and only if for $t > 0$ and $s = 1, 2, \ldots, n$,

$$
\frac{X_{s:n} - t}{\theta(t)} | X_{1:n} > t \overset{d}{=} X_{s:n}
$$

where $d$ stands for distribution.

**Remark.** To prove the ‘only if’ part of the Theorem 2.1, and the ‘if’ part of Theorem 2.2 one does not actually need to assume initially that $m(x)$ is the mean residual life function of $F$. In fact, if we assume, in these cases, that the results are true for some non-negative function $\theta(.)$ then we see at the end that $m(x)$ must be the mean residual life function of $F$ and that it must be a linear function.

In the following theorem we prove a result which is stronger than the result obtained in Theorem 2.2.

**Theorem 2.3.** Let $X_{1:m:n}, X_{2:m:n}, \ldots, X_{m:m:n}$ be non-negative progressively type-II right censored order statistics from a population with absolutely continuous distribution function $F$. Furthermore assume that $\theta(x) = \frac{m(x)}{m(0)}$, where $m$ denotes the mean residual life function of $F$. Then $F$ is GPD if and only if for $t > 0$ and $s = 1, 2, \ldots, m$,

$$
E\left(\frac{X_{s:m:n} - t}{\theta(t)} | X_{1:m:n} > t \right) = E(X_{s:m:n}).
$$

(19)

**Proof.** First we prove the ‘if part’ of the theorem. If $s = 1$, then we have

$$
E(X_{1:m:n} - t | X_{1:m:n} > t) = E(X_{1:m:n})\theta(t) = a(m, n)m(t),
$$

(20)

where $a(m, n) = \frac{E(X_{1:m:n})}{m(0)}$. Since

$$
m(t) = \int_t^\infty \frac{\bar{F}(x)}{F(t)} dx,
$$

(21)
the equation (20) can be written as
\[ \int_0^\infty P(X_{1:m:n} - t > u|X_{1:m:n} > t)du = a(m, n) \int_t^\infty \frac{\tilde{F}(x)}{F(t)}dx, \]
or
\[ \int_t^\infty \left( \frac{\tilde{F}(x)}{F(t)} \right)^n dx = a(m, n) \int_t^\infty \frac{\tilde{F}(x)}{F(t)}dx. \]
On differentiating the last equality with respect to \( t \), we obtain
\[ (\tilde{F}(t))^n = a(m, n)(n - 1)f(t)(\tilde{F}(t))^{n-2} \int_t^\infty \tilde{F}(x)dx + a(m, n)(\tilde{F}(t))^n, \]
which is equivalent to
\[ a(m, n)(n - 1)r(t)m(t) = (1 - a(m, n)). \]
This implies, using the fact that \( r(t) = \frac{m'(t)+1}{m(t)} \), the mean residual \( m(t) \) is linear. That is, \( F \) is GPD.

Now suppose that \( 2 \leq s \leq m \). Then, from (19), we have
\[ E(X_{s:m:n} - t|X_{1:m:n} > t) = E(X_{s:m:n})\theta(t) = a(s, m, n)m(t), \tag{22} \]
where \( a(s, m, n) = \frac{E(X_{s:m:n})}{m(0)} \). On the other hand
\[ P(X_{s:m:n} > t + u|X_{1:m:n} > t) = c_{s-1} \sum_{i=1}^s \frac{a_i(s)}{\gamma_i} \left( \frac{\tilde{F}(t + u)}{F(t)} \right)^{\gamma_i}, \]
which, in turn, implies that
\[ E(X_{s:m:n} - t|X_{1:m:n} > t) = \int_0^\infty c_{s-1} \sum_{i=1}^s \frac{a_i(s)}{\gamma_i} \left( \frac{\tilde{F}(t + u)}{F(t)} \right)^{\gamma_i} du \\
= \int_t^\infty c_{s-1} \sum_{i=1}^s \frac{a_i(s)}{\gamma_i} \left( \frac{\tilde{F}(x)}{F(t)} \right)^{\gamma_i} dx \\
= M_{s:m:n}(t). \tag{23} \]
On differentiating the last equality with respect to \( t \), and noting that (see the form of the distribution function of \( X_{r:m:n} \) in Introduction section),
\[ c_{s-1} \sum_{i=1}^s \frac{a_i(s)}{\gamma_i} = 1, \]
we obtain
\[ M_{s:m:n}'(t) = -1 + r(t)cs^{-1} \int_t^\infty \sum_{i=1}^s a_i(s) \left( \frac{\bar{F}(x)}{F(t)} \right)^{\gamma_i} dx. \]  

(24)

One can easily verify that
\[ M_{s:m:n}(t) - M_{s-1:m:n}(t) = cs^{-2} \int_t^\infty \sum_{i=1}^s a_i(s) \left( \frac{\bar{F}(x)}{F(t)} \right)^{\gamma_i} dx. \]

Hence, equation (24) can be written as
\[ M_{s:m:n}'(t) = -1 + r(t) \frac{cs^{-1}}{cs^{-2}} (M_{s:m:n}(t) - M_{s-1:m:n}(t)), \]

which is (by taking into account (22)) equivalent to
\[ r(t) = \frac{a(s, m, n)m'(t)}{b(s) (a(s, m, n) - a(s - 1, m, n)) m(t)}, \]

where
\[ b(s) = \frac{cs^{-1}}{cs^{-2}}. \]

From this it can be easily concluded that \( m \) is linear. That is, \( F \) is GPD. The ‘if’ part of the theorem is straightforward and hence is omitted.

**Remark.** Similar to Theorem 2.2 one can see that the result of Theorem 2.3 is also true for ordinary order statistics.

**Acknowledgment**

The authors would like to thank the referees and the guest editor for valuable comments which resulted in the improvement of the presentation of this paper.

**References**

Hashemi and Asadi


