STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **x**, Month 202x, pp 1–18.
Published online in International Academic Press (www.IAPress.org)

A Flexible Pareto-Type Distribution: Properties, Estimation, and Real-Data Application

Eman Kh. Khalaf[1,*], Akram Al-Sabbagh[1], Omar AL-Tameemi[1], Ruaa Wana[2], Jamal N. Al-Abbasi[3]

[1]*Department of Mathematics and Computer Applications, College of Sciences, Al-Nahrain University, Jadiriya, Baghdad, Iraq*
[2]*Studies Planning and Follow-Up Directorate, Ministry of Higher Education and Scientific Research, Baghdad, Iraq*
[3]*Department of Biomedical Sciences, College of Applied Sciences, Cihan, University of Erbil, Kurdistan Region, Iraq*

Abstract In this paper, we define and analyze a three-parameter Pareto-type distribution (PTD) with two shape degrees of freedom and an unknown lower bound. It provides analytical results for density, distribution, survival and hazard functions, quantile function via Lambert W function, moments, residual-life measures, entropy measures, and order statistics. In addition, we explore a two-stage likelihood procedure with independent endpoint estimation, as well as joint weighted least squares (WLS) and Anderson-Darling (AD) estimation, which avoid the complications experienced when it comes to likelihood estimation with a support that depends on its parameters. Parameter bias and root mean squared error are compared at three parameter settings in a Monte Carlo study with 500 replications per configuration, based on samples of size 50, 100 and 200. Out of the configurations contrasted, the two-stage likelihood procedure has the lowest per-parameter RMSE (at n=200), while WLS and AD provide alternative joint estimation methods. Across all three procedures, similar fitted PTD densities are produced when applied to 569 Wisconsin Breast Cancer (Diagnostic) worst-texture measurements. Compared with five benchmark distributions, PTD achieves the minimal AIC, while Gamma beats BIC and shifted Weibull achieves Kolmogorov-Smirnov distance. The results demonstrate that multiple modelling approaches of positive biomedical measurements can yield consistent results without indicating overall superiority for any one model and without empirical confirmation of bathtub-shaped lifetime hazards.

Keywords Pareto-type distribution, entropy measures, lower-endpoint estimation, weighted least squares, Anderson–Darling estimation

AMS 2010 subject classifications 62E15, 62F10, 62B10, 62P10

DOI: 10.19139/soic-2310-5070-4654

1. Introduction

Positive continuous distributions find many applications in reliability and survival analysis, actuarial science, finance and biomedical research. Skewness, concentration near a threshold, and varying hazard-rate shape are common features of two settings that are unlikely to be captured using classical models. These features motivate us to construct and perform a thorough empirical assessment of distributions with extra shape parameters.

The Pareto distribution and its generalizations are a fundamental benchmark for modelling positive data. Arnold[5] provides a survey of the Pareto family; Idemudia and Ekhosuehi[12] suggested three-parameter paralogistic distribution while Alsolmi et al. [2] performed the estimation for an alpha-power-transformed Pareto model in several estimation procedures.

Even more generally, the Lambert-F family of mixed areas as proved by Iriarte et al. [13]. Moreover, few recent models providing bounded and unit (or lifetime) distributional features[15, 11, 4] have been proposed as examples

*Correspondence to: Eman Kh. Khalaf (Email: eman.khalid@nahrainuniv.edu.iq). Department of Mathematics and Computer Applications, College of Sciences, Al-Nahrain University, Baghdad, Iraq.

ISSN 2310-5070 (online) ISSN 2311-004X (print)
Copyright © 202x International Academic Press

of how to introduce different levels of flexibility in the injected distributions. The achievements in Statistics, Optimization and Information Computing involve Pareto-type record inference (Saadati Nik et al. Khalfallah[16]), alpha-power Weibull model: Simulation and Applications (Uma[23]), odd generalized exponential-Topp-Leone Weibull Distribution (Egala et al. [10]).

Evaluating a generic parametric proposal only using its density plots is not sufficient. The main purpose of a characterisation of the model is done through its survival and hazard functions, quantiles, moments, entropy measures and order statistics; furthermore estimation studies and comparisons with already known distributions explain how it works in practice. Al-Tameemi et al. [1] proposed an approach associated with a large zipped flexible distribution framework . The cumulative-hazard function has a particular form, which is what the present study attempts to clarify, giving rise to a three-parameter Pareto-type distribution (PTD). We do not pretend that the construction of logarithm and exponent itself is new. Its structure facilitates a closed-form Lambert-W quantile and many different hazard and density shapes. It is, however, *not* Pareto-regularly-varying at its upper tail (the term "Pareto-type" refers to the positive lower threshold and the logarithmic transformation.

Apart from the underlying mathematical properties of a probability distribution, reliable parameter estimation and empirical verification are crucial for evaluating its practical applicability. While the MLE methods are widely used, exploratory estimates might provide us with further information about the statistical properties of our proposed model. We contrast this approach with weighted least squares (WLS) estimation, as was followed by Swain et al. [26] and Anderson-Darling (AD) estimation[3] that puts more weight to differences in distribution tails. These procedures are assessed in finite sample with Monte Carlo simulations based on bias and root mean squared error. In addition, the applicability of the developed distribution with real-world data is also explored on the Wisconsin Breast Cancer (Diagnostic) dataset [28] in comparison to these commonly used probability distributions based on information criteria and goodness-of-fit statistics.

The remaining sections give various definitions. Section 2 describes the PTD, Section 3 survival and hazard functions, Section 4 the quantile function, Sections 5-6 respectively raw and incomplete moments, Section 7 is dedicated to entropy and Section 8 for order statistics. The estimation methods and the Monte Carlo study and real-data application are provided in Sections9-11, respectively. Finally, Section 12 concludes.

2. The Proposed Three-Parameter Pareto-Type Distribution

A three-parameter family of Pareto-type distributions is defined by changing logarithm, with a positive generator being the exponential transformation. This results in a family with one lower endpoint that is adjustable and two shape parameters. For the entire paper, we refer to this family as a Pareto-type distribution (PTD).

Suppose that X has support (c, ∞), for scale parameter $c > 0$. Set

$$g(x) = \ln\left(\frac{x}{c}\right)$$

For $a, b > 0$ define the nonnegative cumulative-hazard generator by

$$z(x) = g(x)^a \left(\frac{x}{c}\right)^b,$$

As $z(x)$ monotonically increases from zero to infinity for $x > c$, the expression $1 - e^{-z(x)}$ is a valid cdf. Define

$$F(x) = 1 - \exp\left[-z(x)\right]$$

Therefore, for $x > c$,

$$F(x) = 1 - \exp\left[-\left(\ln\frac{x}{c}\right)^a \left(\frac{x}{c}\right)^b\right], \qquad x > c. \tag{1}$$

For the sake of completeness, it holds that $F(x) = 0$ for all $x \leq c$; there is no point mass at c.
Differentiating the CDF with respect to x produces the probability density function,

$$f(x;a,b,c)=\frac{1}{x}\left(\frac{x}{c}\right)^{b}\left(\ln\frac{x}{c}\right)^{a-1}\left[a+b\ln\left(\frac{x}{c}\right)\right]\exp\left\{-\left(\ln\frac{x}{c}\right)^{a}\left(\frac{x}{c}\right)^{b}\right\},\qquad x>c. \tag{2}$$

The parameter c is the lower end point and a scale parameter. The density and hazard functions are jointly influenced by a and b, where the power of the distance between the logarithm of c is dictated by parameter a, parameter b modifies the growth rate of cumulative hazard. The analytical properties and a numerical example below highlight these effects.

This distribution is identifiable because the lower endpoint, c, is uniquely determined by the support. In addition, by setting $y=\ln(x/c)>0$, the survival function is given by $\ln[-\ln S(x)]=a\ln y+by$.

Since $\ln y$ and y are linearly independent on $((0,\infty)$. the shape parameters a and b are uniquely determined.

In order to show the versatility of the proposed PTD, Figure 1 includes some plots of PDF for selected values of parameters *a, b* and *c* illustrating the range of densities obtainable from the proposed model, including right-skewed shapes and densities strongly concentrated near the lower endpoint.

The density's peak location and skewness are largely influenced by the shape parameter *a*. For small *a*, the density is closely clustered around $x=c$, producing an extremely right-skewed form. The density will be smoother as *a* increases and can take on a more quasi-unimodal shape. Therefore, a plays a role in the mass of probability closer to the lower endpoint itself and the skewness.

The second shape parameter *b*, primarily controls the tail behavior and how fast the density decays in relation to x. Higher values of *b* generate a more rapid falloff of the density, while lower values permit heavier tails. Therefore the value of b plays an significant role in upper-tail decay. And even though $0<b<1$ shows how the PTD tail is lighter than any power-law tail, it still decays much slower than an exponential tail when $b\geq 1$ gives at least exponentially-order decay.

In contrast with the classical Pareto distribution, the survival function of a PTD is not regularly varying: $x^{k}S(x)\to 0$ as $x\to\infty$, for every $k>0$. For $0<b<1$, the survival function decreases at a slower rate than $e^{-\lambda x}$ for any fixed $\lambda>0$.

Therefore scaling c scales the distribution and rightwards translates its support endpoint, but has no effect on the distribution of the dimensionless ratio, X/c, for fixed a,b. This provides both scale and threshold interpretations to c.

Therefore scaling c scales the distribution and rightwards translates its support endpoint, but has no effect on the distribution of the dimensionless ratio, X/c, for fixed a,b. This provides both scale and threshold interpretations to c.

To summarize, Figure 1 shows that the influence of the two shape parameters on skewness, concentration near c and upper-tail decay. Evaluation of the practical adequacy of the distribution against observed data rather than inferred from density flexibility is needed.

3. Survival and Hazard Rate Functions

The survival and hazard rate functions incorporate the basic principles from reliability theory and survival analysis that describe the lifetime characteristics of a probability distribution [17, 18]. The survival function is the probability that a unit survives time (x), and the hazard rate function is the instantaneous potential per unit time of failure at time (x), given that it has survived up to that point in time. These functions are useful for the proposed PTD, as they may give insights about the reliability behavior and aging features of the model.

Using the cumulative distribution function given in Equation (1), the survival function of the PTD is obtained as

$$S(x)=1-F(x)=\exp\left\{-\left(\ln\frac{x}{c}\right)^{a}\left(\frac{x}{c}\right)^{b}\right\},\qquad x>c. \tag{3}$$

The survival function meets the boundary conditions $S(c)=1$ and $\lim_{x\to\infty}S(x)=0$, proving that it is indeed a sensible survival function for the PTD under consideration. The shape parameters *a* and *b* together manage how rapidly the survival probability achieves zero over the support of the distribution. The survival probability decreases

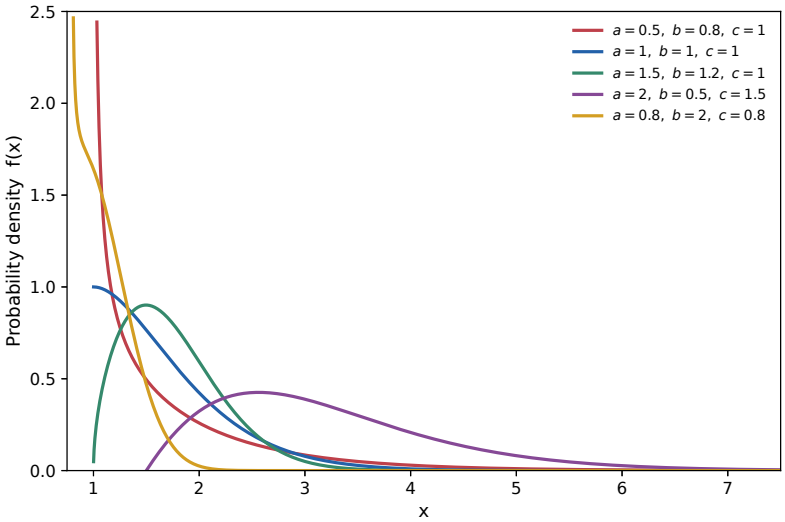


Figure 1. Probability density functions of the proposed PTD for selected parameter combinations.

as b increases and the effects of increasing a are determined by whether $\ln(x/c)$ is less or greater than one for a fixed $x > c$. This implies that the shape parameters do not need to order survival curves uniformly over support. As a result, the PTD can accommodate fairly diverse patterns of lifetime data reliability.

The hazard rate function, which reflects the time-dependent instantaneous risk of an object failing at time (x), given that it has survived up to and including time (x) [17, 19]. It is defined as

$$h(x) = \frac{f(x)}{S(x)}.$$

The hazard rate function of PTD stems from substituting Equations (2) and (3) into the above expression as follows:

$$h(x) = \frac{1}{c}\left(\ln\frac{x}{c}\right)^{a-1}\left(\frac{x}{c}\right)^{b-1}\left[b\,\ln\frac{x}{c} + a\right] \tag{4}$$

In other words, by writing $y = \ln(x/c) > 0$, one can see that the sign of $h\prime(x)$ is equal to the sign of $a(a-1) + a(2b-1)y + b(b-1)y^2$. Specifically, $a \geq 1$ and $b \geq 1$ gives a nondecreasing hazard while having $0 < a < 1$ and $b > 1$ begets a bathtub-shaped hazard. These articulations describe potential shapes and do not, a proof of their existence in nature.

The PTD, therefore, can exhibit a broad range of failure-rate behavior because the hazard rate function is linked to only three model parameters. Among the possible uses of hazard patterns, these analytically identifiable patterns suggest a potential application for lifetime modelling, but its empirical adequacy requires a matching failure-time distribution.

Figure 2 shows the hazard rate of the proposed PTD for some parameter combinations of *a, b,* and *c*, where it can be seen that the PTD is able to produce many different shapes such that it can fit a wide variety of survival (and reliability) data.

Many combinations of parameters lead to hazard rates that are non-decreasing, which means the instantaneous risk of failure is increasing in time. This type of behavior is typical for aged systems that show gradual degradation of their components due to wear, fatigue or material degradation. As an example, increasing hazard functions are quite common in engineering reliability, mechanical systems and biomedical survival studies.

In general, combinations of other parameters lead to a bathtub-shaped hazard function which is defined by the existence of an initial high failure rate that decreases over early operation before finally increasing again at

longer times. The model follows the similar early-failure-useful-life-wear-out pattern commonly used in reliability engineering [17, 19]. The generality so achieved essentially increases the relevance of the PTD to real life lifetime modeling, in which both increasing and bathtub-shaped hazards need to be considered.

In general, the PTD can have increasing and bathtub hazards for some parameter values. Such theoretical attributes ought to be of an entirely diverse nature than excellent fit evidenced by leaving out to specific survival or reliability data sets.

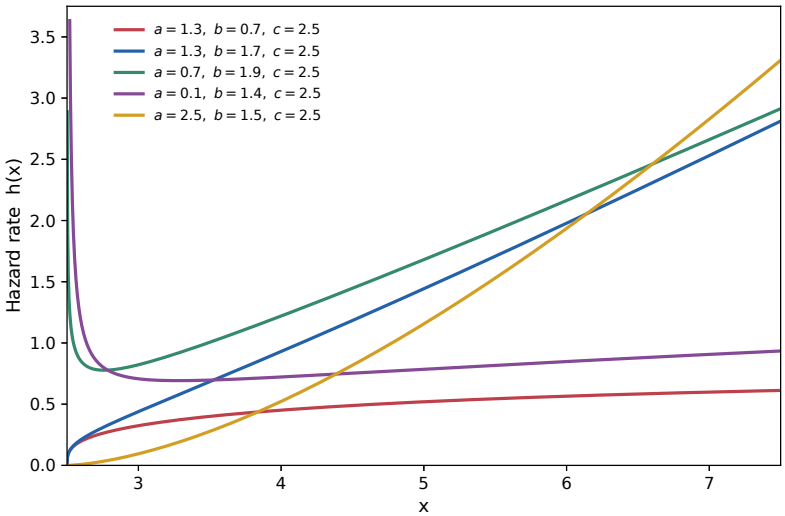


Figure 2. Hazard-rate functions of the proposed PTD for selected parameter combinations.

4. Quantile Function

The quantile function is a fundamental statistical building block, used as part of many random variate generation algorithms. It also gives you direct access to get random samples from the proposed PTD using the inverse transform approach and support Monte Carlo simulation studies [9]. Also, the quantile function is very valuable for calculating descriptive things like median and other sample quantiles.

Let $0 < u < 1$ and put $z = -\ln(1-u)$. From Equation (1), the quantile equation is

$$z = [\ln(x/c)]^a \, (x/c)^b .$$

Consider the transformation

$$\vartheta = \ln(x/c), \tag{5}$$

Hence, the ath root and rearranging gives

$$\frac{b}{a}\vartheta \exp\left(\frac{b}{a}\vartheta\right) = \frac{b}{a} z^{1/a}.$$

Inverting this relation using the principal Lambert-W function [7] yields $\vartheta = (a/b)W_0((b/a)z^{1/a})$, and hence

$$Q(u) = c \exp\left(\frac{a}{b}\, W\, \left(\frac{b}{a}\ \ (-\ln(1-u))^{\frac{1}{a}}\right)\right) \tag{6}$$

In Equation (6), W denotes the principal real branch W_0 of the Lambert-W function, with $0 < u < 1$.

5. Raw Moments

Raw moments are basic identifying features of a probability distribution that quantify its location, variability, skewness and tail behavior [14]. This section derives analytical expressions for the raw moments of the proposed PTD, and from them, various descriptive measures: mean, variance, skewness and kurtosis.

Consider a random variable X which follows the proposed PTD with pdf (2). The $r-th$ raw moment about the origin is defined as,

$$\mu_r^{'} = E\left(X^r\right) = \int_c^{\infty} x^r f(x) dx. \tag{7}$$

Plugging the probability density (2) into Equation (7), we have this integral representation of $r-th$ raw moment of PTD:

$$\mu_r^{'} = c^r \int_0^{\infty} t^{a-1}(a+bt) \exp\left((r+b)t - t^a e^{bt}\right) dt, \tag{8}$$

which provides the base representation to compute any order moment.

The numerical values calculated for the mean, variance, skewness and kurtosis of the proposed PTD for a selection of combinations of *a, b* and *c* using numerical evaluation of the corresponding moment expressions are reported in Table 1.

Table 1 shows that, with a and c fixed, increasing b decreases the mean and variance; with b and c fixed, increasing a increases the mean and decreases the variance. In the same spirit, the level of asymmetry and tail heaviness are less pronounced for higher values of a and b (that is, since a and b increase, we observe lower coefficients of skewness and kurtosis).

The PTD demonstrates significant right-skewness and leptokurticity with these relatively small values of the shape parameters, which is often used to characterize right-skewed data. With increasing values of shape parameters, the distribution moves progressively from pronounced right-skewness to less skewed and more concentrated shapes. These results indicate that the PTD can be fitted over a wide range of distributional shapes, from strongly right-skewed distributions to more moderate and compact patterns.

Table 1. Descriptive measures of the PTD for different parameter combinations.

a	b	c	**Mean**	**Variance**	**Skewness**	**Kurtosis**
1.5	0.5	1.0	2.099790	0.751527	1.596724	6.943349
1.5	1.0	1.0	1.818652	0.282402	0.979019	4.112392
1.5	1.5	1.0	1.671570	0.151345	0.693116	3.309036
2.0	0.5	1.0	2.150088	0.526654	1.136161	4.836697
2.0	1.0	1.0	1.908971	0.239032	0.717414	3.468415
2.0	1.5	1.0	1.767479	0.138537	0.491076	3.005190
2.5	0.5	1.0	2.196537	0.401812	0.835503	3.866513
2.5	1.0	1.0	1.982151	0.205953	0.526341	3.126533
2.5	1.5	1.0	1.846224	0.126439	0.340799	2.852346

6. Incomplete Moments

Incomplete moments are important in reliability theory, actuarial science and survival analysis because they give information about the distribution of observations below or above a certain threshold [14]. They also form the basis of many descriptive statistics measures like median and mean residual life (MRL) function. This section makes a

mathematical derivation of the lower and upper incomplete moments of the proposed PTD, and then determines the median and MRL function correspondingly.

The lower incomplete moment of order r at a threshold x_0 is given by:

$$\mu_r^- (x_0) = \int_c^{x_0} x^r f(x) dx, \tag{9}$$

Likewise, the upper incomplete moment of order *r,*

$$\mu_r^+ (x_0) = \int_{x_0}^{\infty} x^r f(x) dx. \tag{10}$$

Inserting the probability density function of the PTD (2) into Equations (9) and (10), we have, respectively, the following integral representations of the lower and upper incomplete moments:

$$\mu_r^- (x_0) = c^r \int_0^{\ln(x_0/c)} e^{(r+b)t} t^{a-1} (a+bt) e^{-t^a e^{bt}} dt, \tag{11}$$

$$\mu_r^+ (x_0) = c^r \int_{\ln(x_0/c)}^{\infty} e^{(r+b)t} t^{a-1} (a+bt) e^{-t^a e^{bt}} dt. \tag{12}$$

These quantities are useful, in particular, actuarial science, income analysis and reliability studies where the interest is on observations below or above a fixed threshold.

The median is obtained from the quantile function by setting *u=0.5*. Hence,

$$M = c \exp\left[\frac{a}{b} W\left(\frac{b}{a} (\ln 2)^{1/a}\right)\right]. \tag{13}$$

Mean residual life (MRL) function is a reliability characteristic which indicates the expected lifetime of a unit that has already outlived some amount of time (x_0) . It is defined as

$$MRL(x_0) = E(X - x_0 \mid X > x_0) = \frac{1}{S(x_0)} \int_{x_0}^{\infty} (x - x_0) f(x) dx, \tag{14}$$

where $(S(x))$ denotes the survival function.

The MRL function can then be expressed as a transformation of (5),

$$MRL(x_0) = \frac{1}{S(x_0)} \int_{\ln(x_0/c)}^{\infty} \left(ce^{\vartheta} - x_0\right) \vartheta^{a-1} (a + b\vartheta) e^{b\vartheta} e^{-\vartheta^a e^{b\vartheta}} d\vartheta. \tag{15}$$

We note that the mean residual life function carries some information related to the aging properties of the distribution. MRL estimates the expected further life given survival to age x_0; its trend with respect to x_0 may reveal useful information about aging but, by itself, the MRL does not provide evidence for increasing or decreasing reliability in old age.

Selected combinations of the parameters *a* and *b* are presented in Table 2 together with corresponding numerical values of median, incomplete moments, and the mean residual life.

The results are partially consistent in that the median tends to increase with larger values of *a*, suggesting a movement of probability mass towards higher observations. Moreover, for fixed a, both the upper incomplete first moment and mean residual life decrease as b increases.

Larger mean residual life values are observed for smaller parameter values as the distribution is heavier-tailed. This behaviour supports the ability of the PTD model to fit a wide variety of reliability and lifetime trends.

In summary, the numerical results show that PTD provides a flexible tool to model right-skewed and moderately concentrated data which can be subsequently used for applications in reliability engineering, actuarial science and survival analysis.

Table 2. Median, incomplete moments and mean residual life.

a	b	x_0	**Median**	$\mu_1^-(x_0)$	$\mu_1^+(x_0)$	$\mathrm{MRL}(x_0)$
1.5	0.5	2.00	1.885167	0.845778	1.254012	0.836205
1.5	1.0	2.00	1.724122	1.045002	0.773650	0.453536
1.5	1.5	2.00	1.621139	1.226915	0.444655	0.274543
2.0	0.5	2.00	2.011857	0.783679	1.366409	0.695677
2.0	1.0	2.00	1.845644	0.986248	0.922723	0.412057
2.0	1.5	2.00	1.734658	1.186361	0.581118	0.261719
2.5	0.5	2.00	2.104757	0.710369	1.486168	0.616640
2.5	1.0	2.00	1.939767	0.909249	1.072903	0.387805
2.5	1.5	2.00	1.825519	1.118614	0.727609	0.255540

7. Entropy Measures

Entropy is a measure of uncertainty and randomness associated with a probability distribution, widely used in information theory, reliability analysis statistical inference, engineering etc. Shannon entropy and Rényi entropy are two of the most commonly adopted measures of entropy, providing descriptions of the uncertainty in a random variable [24, 21]. This section derives the expressions for these entropy measures for the proposed PTD.

Formally, the Shannon entropy of a continuous random variable X with probability density function $f(x)$ is defined as

$$H(X) = -\int_c^\infty f(x)\ln\left[f(x)\right]dx. \tag{16}$$

Thus, PTD Shannon entropy can be calculated as

$$H(X) = -\int_0^\infty t^{a-1}(a+bt)e^{bt}e^{-t^a e^{bt}}\ln\left[f\left(ce^t\right)\right]dt. \tag{17}$$

Expression (17) can be computed numerically for any admissible parameter values.

The Rényi entropy is a generalization of the concept of uncertainty, defined as

$$H_\alpha(X) = \frac{1}{1-\alpha}\ln\left[\int_c^\infty f(x)^\alpha dx\right], \alpha > 0, \quad \alpha \neq 1. \tag{18}$$

Substituting (2) into (18), with t = ln(x/c), gives:

$$H_\alpha(X) = \frac{1}{1-\alpha}\ln\left[c^{1-\alpha}\int_0^\infty \left(t^{a-1}(a+bt)\right)^\alpha e^{(\alpha b+1-\alpha)t}e^{-\alpha t^a e^{bt}}dt\right]. \tag{19}$$

The Rényi integral is finite for all orders $\alpha > 0$ when a $\geq$ 1; if $0 < a < 1$, finiteness additionally requires $\alpha <$ 1/(1 − a).

This yields an expression which is the starting point for numerical evaluation of the Rényi entropy at different orders of α.

The obtained entropy forms yield quantitative measures of the uncertainty involved in the PTD. As both entropy measures explicitly depend on the model parameters, they also provide some additional knowledge on how learning shape and scale parameters indeed might cause a typical distribution to have more variability and dispersion.

Figures 3 and Table 3 show the numerical values of Rényi entropy and its graphical behaviour for various values ranges of the order parameter, where $a = 1.5,\ b = 2.0$, and $c = 1.0$. respectively.

We see that for larger values of α the Rényi entropy decreases monotonically. For small values of α, the entropy achieves moderately large values, suggesting heavier weighting to low-probability and extreme events. The weights

of rare observations decays as the number of α increases and consequently the entropy will be dominated by regions with higher probability densities. This reflects the increasing emphasis on high-density regions as the Rényi order increases.

From the numerical results, one can see that with increasing order parameter, the entropy decreases from $(H_{0.1} = 0.7590)$ to $(H_{3.0} = 0.0274)$, which indicates a significant reduction of uncertainty. In addition, the entropy values near (α =1) approach the Shannon entropy, confirming a well-known limiting relation.

$$\lim_{\alpha \to 1} H_\alpha = H. \tag{20}$$

As illustrated in Figure 3, larger values of α emphasize high-density regions of the proposed distribution.

Overall, the entropy analysis further highlights some of the features of uncertainty behind the proposed distribution and demonstrates its flexibility in accommodating various levels of concentration and tail sensitivity.

Table 3. Rényi entropy for selected orders.

α	H_α
0.10	0.7590444630
0.25	0.4965501908
0.50	0.3321534577
0.75	0.2491536650
0.90	0.2147359857
0.99	0.1974209456
1.01	0.1938449972
1.10	0.1788044245
1.25	0.1569384713
1.50	0.1270844781
2.00	0.0830759557
3.00	0.0273702252

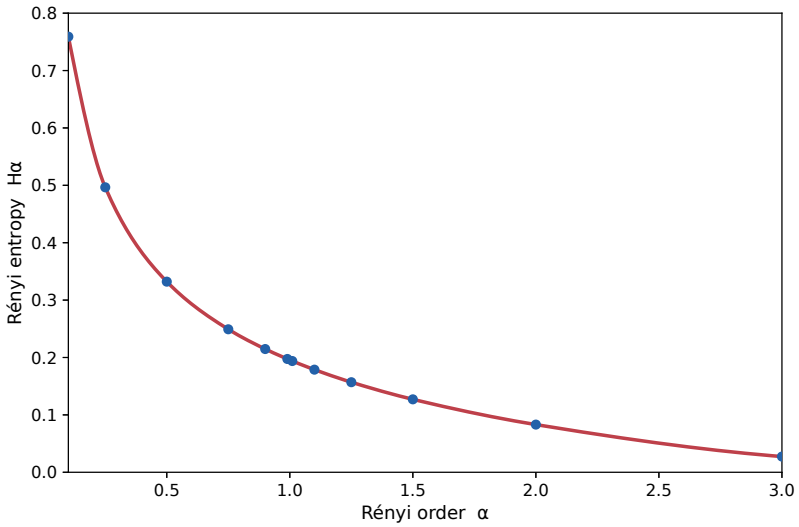


Figure 3. Rényi entropy of the PTD versus order parameter.

8. Order Statistics

Order statistics are a key concept in probability theory, statistical inference and reliability analysis and can be used to extend the applications of classical estimation methods, life testing approaches, quality control techniques by performing so called extreme-value analysis. They describe the sorted values of a random sample and underpin many inferential procedures David and Nagaraja [8]. In this section, we derive the probability density function (pdf) of order statistics associated with the proposed PTD.

Let $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ be the ordered sample from n random sampled data form the proposed distribution. The probability density function of the $r-th$ order statistic $X_{r:n}$ out of n elements from a continuous distribution with cumulative distribution function $F(x)$ and probability density function $f(x)$, is

$$f_{r:n}(x) = \frac{n!}{(r-1)!(n-r)!} \left[F(x)\right]^{r-1} \left[1-F(x)\right]^{n-r} f(x), x > c, r = 1, 2, \ldots, n. \tag{21}$$

Replacing the $F(x)$ and $f(x)$ in Equation(21) with Equations (1) and (2), respectively, leads to the probability density function of the r-th order statistic for PTD. Here $z(x) = [\ln(x/c)]^a (x/c)^b$.

$$f_{r:n}(x) = \frac{n!}{(r-1)!(n-r)!} \left(1-e^{-z(x)}\right)^{r-1} e^{-(n-r+1)z(x)} \cdot \frac{1}{c} \left(\ln\left(\frac{x}{c}\right)\right)^{a-1} \left(\frac{x}{c}\right)^{b-1} \left[b \ln\left(\frac{x}{c}\right) + a\right]. \tag{22}$$

The form derived serves as the foundation for studying the statistics of minima, maxima and other order statistics from the PTD. These findings can be applied in, for example, reliability study, life testing and extreme data inference contexts.

The first order statistic $X_{1:n}$ corresponds to the smallest sample observation. By setting $r = 1$ in Eq (22), we obtain the following

$$f_{X_{1:n}}(x) = \frac{n}{c} \left(\ln\left(\frac{x}{c}\right)\right)^{a-1} \left(\frac{x}{c}\right)^{b-1} \left[b \ln\left(\frac{x}{c}\right) + a\right] e^{-nz(x)}. \tag{23}$$

The minimum order statistic is particularly important in reliability studies because it characterizes the lifespan of a series system that fails as soon as any one component fails.

The largest order statistic $X_{n:n}$ represents the maximum observation in the sample. Setting *r=n* in (22) gives

$$f_{X_{n:n}}(x) = \frac{n}{c} \left(1-e^{-z(x)}\right)^{n-1} \left(\ln\left(\frac{x}{c}\right)\right)^{a-1} \left(\frac{x}{c}\right)^{b-1} \left[b \ln\left(\frac{x}{c}\right) + a\right] e^{-z(x)}. \tag{24}$$

The maximum order statistic is applied in extreme-value analysis, and corresponds to the lifetime of a parallel system (which fails only when all its components have failed).

The expected value of the smallest order statistic and the largest order statistic are given as

$$E\left[X_{1:n}\right] = \int_c^\infty x \cdot f_{X_{1:n}}(x)\, dx, \tag{25}$$

$$E\left[X_{n:n}\right] = \int_c^\infty x \cdot f_{X_{n:n}}(x)\, dx. \tag{26}$$

These quantities give information about the way sample extremes behave as one increases the size of the sample.

The expected values of the lowest and highest order statistics are shown in Figures 4 and 5 as functions of the sample size n.

The results show that $E\left(X_{1:n}\right)$ monotonically decreases as the sample size increases. One can expect this as larger samples are more likely to have observations nearer the lower end of a distribution.

On the other hand, $E(X_{n:n})$ grows in sample size (since larger samples provide greater chance of observing large sample values).

As n increases, $E(X_{1:n})$ will decrease towards the finite lower endpoint at c, while $E(X_{n:n})$ diverges to ∞, since the support of a PTD is unbounded above. Over the plotted range, their changes may be decreasing more and more, but the expected maximum will not tend to a finite constant.

Sample extremes under the PTD model are characterized by standard order-statistic theory, and these monotonicity properties follow immediately.

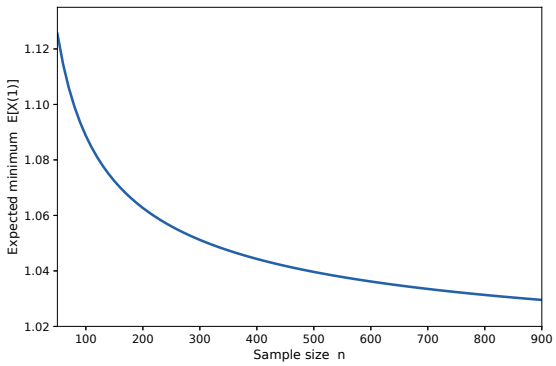


Figure 4. Expected value of the smallest order statistic versus sample size.

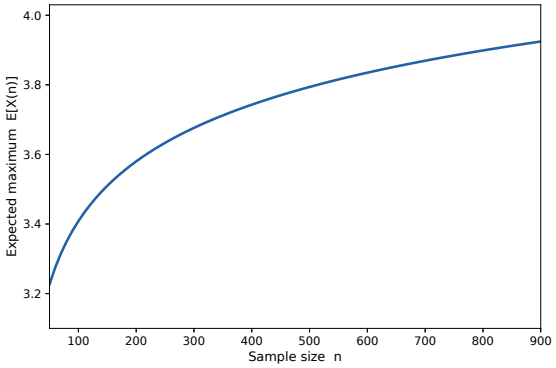


Figure 5. Expected value of the largest order statistic versus sample size.

9. Parameter Estimation

9.1. Maximum Likelihood Estimation

Maximum likelihood (ML) is a popular method among the commonly used methods for estimating unknown parameters since it has many attractive statistical properties, including consistency, asymptotic efficiency and asymptotic normality [6, 20]. This section considers inference for a random sample of independent and identically distributed observations and provides the maximum likelihood estimators of the parameters of PTD.

Suppose that $X_1, X_2, \ldots, X_n$ is a sample of size n, drawn from the PTD with parameter vector $\theta = (a, b, c)^\top$. Consequently, the likelihood function is

$$L(a, b, c) = \prod_{i=1}^{n} f\left(x_i; a, b, c\right), \tag{27}$$

where

$$f\left(x_i; a, b, c\right) = \frac{1}{x_i}\left(\frac{x_i}{c}\right)^b \left(\ln \frac{x_i}{c}\right)^{a-1} \left[a + b \ln\left(\frac{x_i}{c}\right)\right] \exp\left\{-\left(\ln \frac{x_i}{c}\right)^a \left(\frac{x_i}{c}\right)^b\right\}$$

Substituting the density (2) yields

$$L(a, b, c) = \prod_{i=1}^{n} \frac{1}{x_i}\left(\frac{x_i}{c}\right)^b \left(\ln \frac{x_i}{c}\right)^{a-1} \left[a + b \ln\left(\frac{x_i}{c}\right)\right] \exp\left\{-\left(\ln \frac{x_i}{c}\right)^a \left(\frac{x_i}{c}\right)^b\right\}. \tag{28}$$

To ease the computation, we consider taking the log of the likelihood function (28), which gives us the log-likelihood function

$$\ell(a, b, c) = \ln L(a, b, c).$$

By applying some algebraic manipulation, the log-likelihood function of PTD becomes

$$\ell(a,b,c) = -\sum_{i=1}^{n} \ln x_i + b\sum_{i=1}^{n} t_i + (a-1)\sum_{i=1}^{n} \ln t_i + \sum_{i=1}^{n} \ln(a+bt_i) - \sum_{i=1}^{n} z_i. \tag{29}$$

For $0 < c < \min_i x_i$, put $t_i = \ln(x_i/c)$ and $z_i = t_i^a e^{bt_i}$. Then the score functions corresponding to the parameters a, b, and c be respectively

$$\frac{\partial \ell}{\partial a} = \sum_{i=1}^{n} \left[\ln t_i + \frac{1}{a+bt_i} - z_i \ln t_i \right]. \tag{30}$$

$$\frac{\partial \ell}{\partial b} = \sum_{i=1}^{n} \left[t_i + \frac{t_i}{a+bt_i} - z_i t_i \right]. \tag{31}$$

$$\frac{\partial \ell}{\partial c} = \frac{1}{c}\sum_{i=1}^{n} \left[\frac{z_i(a+bt_i) - (a-1)}{t_i} - b - \frac{b}{a+bt_i} \right]. \tag{32}$$

For an interior stationary point of the likelihood, it must also fulfill the following nonlinear score equations:

$$\frac{\partial \ell}{\partial a} = 0, \quad \frac{\partial \ell}{\partial b} = 0, \quad \frac{\partial \ell}{\partial c} = 0. \tag{33}$$

There is no general closed-form solution to these equations; numerical optimization is necessary and stationary solutions should be checked against boundary conditions on parameters and competing local optima.

As c is unknown and represents the lower endpoint of the support, this likelihood needs to be optimized with respect to $0 < c < \min_i x_i$, so we cannot invoke standard regularity results for the full three-parameter MLE automatically. Specifically, for $0 < a < 1$ the likelihood is unbounded as c approaches from below the sample minimum; explicit checks are also required on boundary behavior and convergence diagnostics. Standard inference for c is a nontrivial assertion without further endpoint-specific analysis. At the lower endpoint, $f(x) \sim ac^{-a}(x-c)^{a-1}$ as $x \downarrow c$; this is a nonregular endpoint model considered by Smith [25].

In the simulation study, we estimate c separately with a formula from Wang et al. [27]. Let $x_{(1)}, x_{(2)}$ be the two smallest observations such that, $x_{(1)} < x_{(2)}$. Then the estimator is calculated by:

$$\widehat{c}_{\mathrm{RW}} = \left[\frac{1}{x_{(1)}} + \frac{1}{n}\left(\frac{1}{x_{(1)}} - \frac{1}{x_{(2)}} \right) \right]^{-1}. \tag{34}$$

the proposed estimator meets the constraints $nx_{(1)}/(n+1) < \widehat{c}_{\mathrm{RW}} < x_{(1)}$ as $x_{(2)} > x_{(1)}$. More importantly, $x_{(1)} \to c$ almost surely as $n \to \infty$, and by the squeeze theorem we also have $\widehat{c}_{\mathrm{RW}} \to c$ almost surely. This creates validity and consistency in the legitimacy of support, but not neutrality or optimal efficiency. We then can calculate a,b¿0 by maximizing $\ell(a, b \mid \widehat{c}_{\mathrm{RW}})$. This yields the two-stage plug-in ML estimates, (not joint three-parameter MLEs), $\widehat{c}_{\mathrm{RW}}$ uncertainty is not propagated by two-dimensional inverse-Hessian calculation. For the case of fitted shape above 1 in real data application, where constrained joint maximum-likelihood fit with both $a > 1$ and $0 < c < x_{(1)}$ was found as for original analysis, we report this latter option. These special implementations are indicated in the result sections following.

9.2. Weighted Least Squares Estimation

Denote the ordered observations by $x_{(1)} \le \cdots \le x_{(n)}$, and define $F_i(\theta) = F(x_{(i}; a, b, c), \theta = (a, b, c)^\top$. This order statistic $F(X_{(i)}; \theta)$ has expected value $\frac{i}{n+1}$, and variance $\dfrac{i(n-i+1)}{(n+1)^2(n+2)}$, given a continuous correctly specified model[8]. We proceed to define the least-squares principle on weighted loss[26] as follows:

$$Q_{\mathrm{WLS}}(\theta) = \sum_{i=1}^{n} \frac{(n+1)^2(n+2)}{i(n-i+1)} \left[F_i(\theta) - \frac{i}{n+1} \right]^2. \tag{35}$$

The WLS estimator $\widehat{\theta}_{\mathrm{WLS}}$ is the minimizer of Q_{WLS} over $a, b > 0$ and $0 < c < x_{(1)}$.. The weights are largest at the ends of the sample, corresponding to points where the variance of the theoretical uniform order statistics is smallest. This approach estimates all three parameters of the PTD simultaneously.

9.3. Anderson-Darling Estimation

The Anderson-Darling (AD) method is a minimum-distance procedure that gives more weight to the tails of the distributions [3]. Let $F_i = F(x_{(i)}; \theta)$. The objective function is

$$Q_{\mathrm{AD}}(\theta) = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\left[\ln F_i + \ln(1 - F_{n+1-i})\right]. \tag{36}$$

$\widehat{\theta}_{\mathrm{AD}}$ is the AD estimator, which minimizes Q_{AD} with respect to $a, b > 0$ and $0 < c < x_{(1)}$. Approach Both WLS and AD use multi-start numerical constrained optimization. During the run-time, strict endpoint inequality is enforced. Once fitted, these estimation goals differ from the reported goodness-of-fit statistics.

10. Monte Carlo Simulation

In order to assess the finite-sample properties of the three estimation methods, The results are reported from a Monte Carlo simulation with 500 independent replications (executed in R by the authors) in each configuration. The random observations are obtained by inverse PTD quantile function ()(6)). The sample sizes are varied as $n = 50, 100, 200$ and for each of the three parameters $\theta = (2, 0.5, 1), (0.7, 1.5, 1)$ and $(2, 2, 1)$. In particular, the second case involves only the nonregular likelihood region $a < 1$. Thus, all procedures were performed on the same generated samples for each configuration.

The ML method used the reciprocal-Wang endpoint estimate (eq.(34)), and then maximized the likelihood over $a, b > 0$ given that estimated endpoint. In contrast, WLS and AD jointly optimize a, b, c via (35) and (36). Hence the comparison is between a two-stage likelihood method and two joint minimum-distance methods, not three unrestricted joint MLEs. Values for the numeric convergence of the experiment, and boundary flags are recorded , which should be treated with caution as optimizer diagnostics rather than an assertion of consistency [22].

The empirical bias and root mean squared error of an estimator $\widehat{\theta}$ are

$$\mathrm{Bias}(\widehat{\theta}) = \frac{1}{N}\sum_{j=1}^{N}(\widehat{\theta}_j - \theta), \tag{37}$$

$$\mathrm{RMSE}(\widehat{\theta}) = \left[\frac{1}{N}\sum_{j=1}^{N}(\widehat{\theta}_j - \theta)^2\right]^{1/2}, \tag{38}$$

where $N = 500$ and the calculations apply separately to each parameter. Smaller RMSE indicates more accurate parameter recovery.

Table 4 provides the three parameter-specific RMSEs for $n = 200$; bias and RMSE for all sample sizes are given in the computational file provided as supplementary material. Last but not least, Figure 6 yields the RMSE evolution for the reference configuration $\theta = (2, 0.5, 1)$. In this setting, the two-stage ML estimator produces smaller RMSE for all three parameters at n=200 while at $n = 50$, WLS and AD have lower endpoint RMSE. For the $(0.7, 1.5, 1)$ setting, two-stage ML has the lowest RMSE for all three parameters when $n = 200$. Out of these methods, two-stage ML gives the lower RMSE for all three parameters $(2, 2, 1)$. Thus, WLS and AD can be viable alternatives to ML estimation methods, but not one-size-fits-all substitutes for ML which are always more efficient.

Table 4. RMSE of the three PTD parameters at $n = 200$ (500 Monte Carlo replications per setting).

True (a, b, c)	Estimator	RMSE(a)	RMSE(b)	RMSE(c)
$(2, 0.5, 1)$	ML	0.4340	0.1932	0.07276
$(2, 0.5, 1)$	WLS	0.4981	0.2402	0.07871
$(2, 0.5, 1)$	AD	0.4874	0.2357	0.07665
$(0.7, 1.5, 1)$	ML	0.0546	0.1041	0.00127
$(0.7, 1.5, 1)$	WLS	0.0652	0.1165	0.00268
$(0.7, 1.5, 1)$	AD	0.0729	0.1149	0.00588
$(2, 2, 1)$	ML	0.4617	0.2303	0.06560
$(2, 2, 1)$	WLS	1.1157	0.5993	0.11815
$(2, 2, 1)$	AD	1.1725	0.6303	0.12059

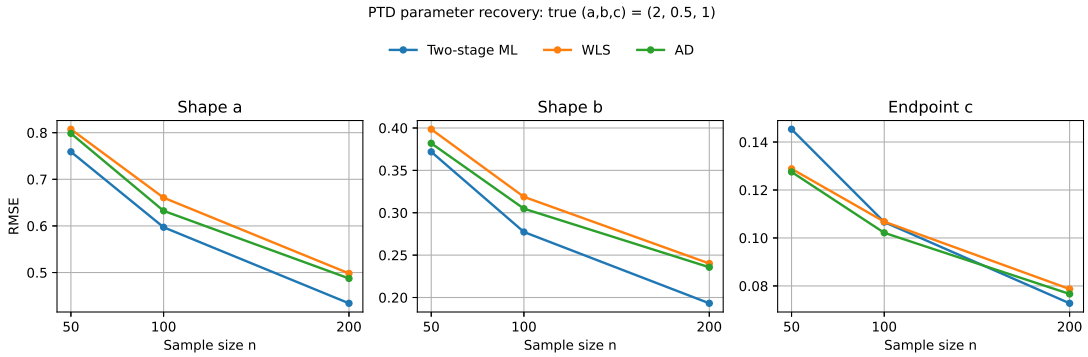


Figure 6. RMSE of the two-stage ML, joint WLS, and joint AD estimators for $(a, b, c) = (2, 0.5, 1)$ at $n = 50, 100, 200$ (500 replications per sample size).

11. Real-Data Application and Model Comparison

To show how well the PDT distribution works realistically, we use the $n = 569$ worst-texture measurements from the Breast Cancer Wisconsin (Diagnostic) dataset ([Wolberg et al. [28]), i.e., image-derived positive-valued biomedical measurements. These data are not survival times observed. In the first part we compare three PTD parameter estimation procedures (ML, WLS and AD) followed by comparison of maximized likelihood fit of PTD with five benchmark distribution families.

The ML fit for the PTD is based on constrained joint likelihood optimization with $a > 1, b > 0$ and $0 < c < x_{(1)}$, which excludes the boundary singularity for $0 < a < 1$. We estimate the full parameter vector under $a, b > 0$ and $0 < c < x_{(1)}$ using WLS and AD. Multiple numerical initializations were used. This dataset consists of observed ties due to measurement rounding; the WLS and AD order-statistic objectives are computed on all 569 observations without artificial jitter. Hence, their continuous-data Order-statistic weighting and calibration are indeed approximations for rounded data, with goodness-of-fit measures interpreted as descriptive. In the fitted solutions, the WLS and AD estimates are interior for all parameters; however, this makes small estimates of b sensitive in their effects on individual parameter interpretation.

Table 5 presents the estimated parameters, negative log-likelihood calculated at each fitted parameter vector, as well as the fit statistics (D, W^2 and A^2) for Kolmogorov-Smirnov, Cram'er-von Mises and Anderson-Darling respectively. The last three are evaluated from the fitted cumulative distribution, regardless of how the parameters were obtained. As expected, the log-likelihood from the fit with maximum-likelihood (ML) is maximized; and AD had slightly lower values of all three statistics based on the cumulative distribution function than ML and WLS. However, the absolute differences are small; not a single one of these results demonstrates that any estimator is

generally better. Because the parameters have been estimated from observations and the measurements are rounded, we report these statistics descriptively rather than via ordinary KS p-values.

Table 5. PTD parameter estimates and goodness-of-fit statistics for Wisconsin worst texture ($n = 569$).

Method	$\widehat{a}$	$\widehat{b}$	$\widehat{c}$	$-\ell$	D	W^2	A^2
ML	3.9408	0.1610	10.4390	1824.447	0.02443	0.04087	0.30854
WLS	3.6799	0.2767	10.7557	1824.508	0.02442	0.03902	0.30509
AD	4.0773	0.0803	10.2455	1824.484	0.02407	0.03812	0.30217

Figure 7 displays the three PTD fitted densities over top of the density-scaled histogram. The curves are nearly identical over much of the data range, with small differences emerging in the upper tail. The resemblance of these graphical results reveals numerical similarity of the three estimators, not exact agreement in estimated parameters.

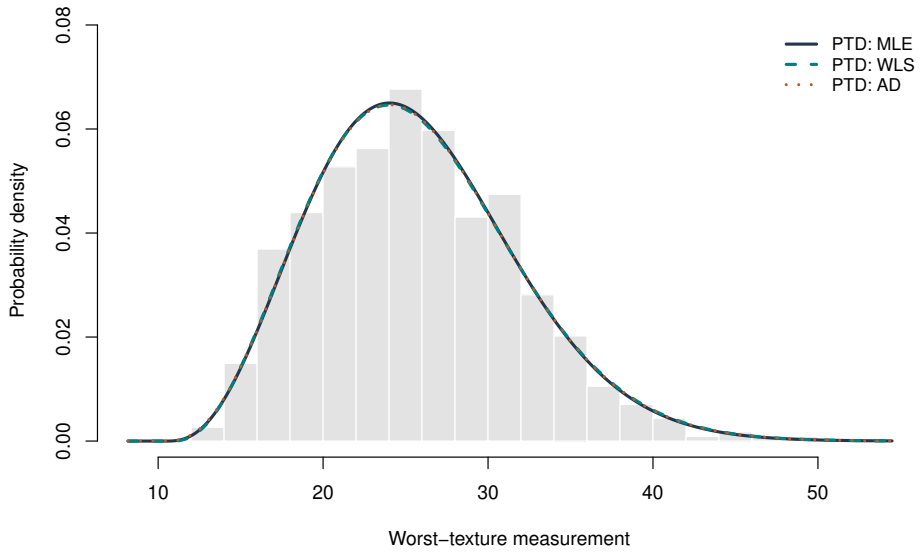


Figure 7. Wisconsin worst-texture histogram with fitted PTD densities obtained by ML, WLS, and AD estimation.

For model comparison, We keep the maximum-likelihood fits of Weibull, Gamma, lognormal, Burr XII and shifted Weibull for model benchmarking on a common set of data containing 569 measurements. The models whose support begins at zero were unshifted, so PTD, Burr XII and shifted Weibull include three estimated parameters (the other models: Weibull, Gamma and lognormal each include only two). Information criteria for the likelihood-fitted models are $\mathrm{AIC} = 2k - 2\ell$ and $\mathrm{BIC} = k \ln n - 2\ell$, with k number of fitted parameters. Under the fitted support-dependent models, these information criteria are descriptive; standard large-sample interpretations of likelihood inference necessitate separate regularity analysis. Table 6 summarizes these comparisons. AIC and BIC measure how similar two distribution families in their ML fits are from each other; they do not rank WLS or AD as estimation methods within a single family versus ML.

The model with the smallest AIC (3654.894) is PTD, the one with lowest BIC (3665.677) is Gamma, and finally shifted Weibull has the lowest KS distance (0.0195). As a result, it's competitive in many respects without crushing every criterion. Histogram with benchmark fitted density curves, shown in figure 8. This comparison assesses a slight distribution of the marginal biomarker difficulty, but do not demonstrate bathtub-shaped hazard accuracy on lifetime observations.

Table 6. Maximum-likelihood benchmark distribution comparison for Wisconsin worst texture ($n = 569$).

Model	k	$-\ell$	AIC	BIC	KS
PTD	3	1824.447	3654.894	3667.925	0.0244
Weibull	2	1855.353	3714.706	3723.394	0.0587
Gamma	2	1826.495	3656.989	3665.677	0.0231
Lognormal	2	1827.053	3658.106	3666.794	0.0367
Burr XII	3	1833.798	3673.596	3686.628	0.0314
Shifted Weibull	3	1825.690	3657.379	3670.411	0.0195

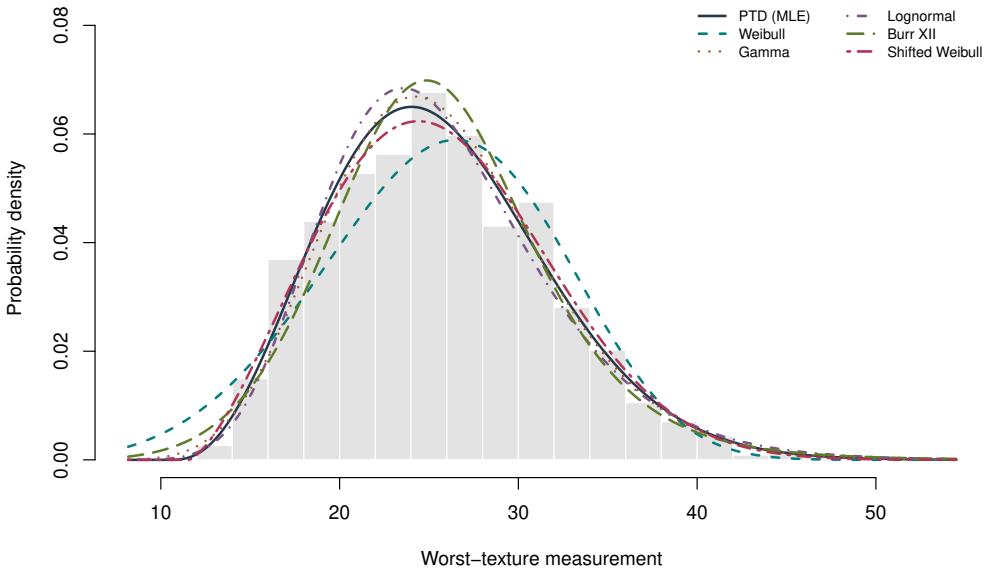


Figure 8. Wisconsin worst-texture histogram with maximum-likelihood fitted PTD and five competing distribution densities.

12. Conclusions

This study examined the distributional properties and estimation of a three-parameter Pareto-type model parameterized by two shape parameters a, b in addition to unknown lower endpoint c.

The probability density, cumulative distribution, survival and hazard functions are obtained as well as the quantile representation using the Lambert W function. In addition, moments, residual-life characteristics, Shannon and R'enyi entropies and order statistics were also analyzed. Depending on the values of the parameters, it may show different hazard-rate behaviour (bathtub-shaped) that is a property of the fitted family and not one established by this current real-data application.

The lower endpoint, which relies on support, introduces non-regularity into the estimation of unrestricted joint maximum likelihood when $0 < a < 1$. We therefore examined a two-stage likelihood method that estimates the endpoint marginally before maximizing the shape parameters with weighted least squares (WLS) and Anderson-Darling (AD) methods for joint estimation of all three parameters.The two-stage approach had the smallest parameter-wise RMSE among these three procedures in the Monte Carlo study (based on 500 replications at each of three combinations of parameter settings and sample sizes $n = 50, 100, 200$) for the biggest sample size setting

($n = 200$). That finite-sample result should not be interpreted as general understating nor assert standard likelihood asymptotics for the flip-side.

For all 569 Wisconsin Breast Cancer (Diagnostic) worst-texture measurements, constrained joint likelihood, WLS and AD estimation produced roughly identical PTD density fits - though individual parameter estimates diverged. Across five standard distribution families, PTD recorded the lowest AIC (3654.894), Gamma the lowest BIC (3665.677) and shifted Weibull with the best fit according to Kolmogorov-Smirnov distance (0.0195). Thus, the model we propose fits positively-biased biomedical data competitive with existing models across comparisons of fit though none of the models was best by all measures. This application does not provide empirical evidence of a bathtub-shaped hazard because these measurements are not failure times.

Further research could explore interval estimation allowing for uncertainty in the separately estimated endpoint, censored-data models, generalizations to regression and applications to true time-to-event data. Such developments would enable a more complete examination of the feasibility and non-regular estimation characteristics of such models.

REFERENCES

1. AL-Tameemi, Omar, Yamama N. Mahmood, Jamal N. Al-Abbasi, and Ruaa Wana. "The family of generalized flexible distributions." In AIP Conference Proceedings, vol. 3460, no. 1, p. 040004. AIP Publishing LLC, 2026. https://doi.org/10.1063/5.0328484
2. Alsolmi, M. M., Almulhim, F. A., Amine, M. M., Aljohani, H. M., Alrumayh, A., & Belouadah, F. (2024). Statistical Analysis and Several Estimation Methods of New Alpha Power-Transformed Pareto Model with Applications in Insurance. *Symmetry*, *16*(10), 1367. https://doi.org/10.3390/sym16101367
3. Anderson, T. W., & Darling, D. A. (1952). Asymptotic theory of certain goodness-of-fit criteria based on stochastic processes. *The Annals of Mathematical Statistics*, 23(2), 193-212. https://doi.org/10.1214/aoms/1177729437
4. AlTwijri, M. I., Alshahrani, N. D., Elgarhy, M., Elsehetry, M., & Elkalzah, B. (2026). A new three-parameter statistical distribution with applications to biomedical and radiation data. Journal of Radiation Research and Applied Sciences, 19(1), 102167. https://doi.org/10.1016/j.jrras.2026.102167
5. Arnold, B.C. (2015). Pareto Distributions (2nd ed.). Chapman and Hall/CRC. https://doi.org/10.1201/b18141
6. Casella, G. and Berger, R.L. (2002) Statistical Inference. 2nd Edition, Duxbury Press, Pacific Grove.
7. Corless, R. M., Gonnet, G. H., Hare, D. E., Jeffrey, D. J., & Knuth, D. E. (1996). On the Lambert W function. Advances in Computational mathematics, 5(1), 329-359. https://doi.org/10.1007/BF02124750
8. David, H. A., & Nagaraja, H. N. (2003). *Order Statistics* (3rd ed.). John Wiley & Sons. https://doi.org/10.1002/0471722162
9. Devroye, L. (1986). Non-Uniform Random Variate Generation. Springer. https://doi.org/10.1007/978-1-4613-8643-8
10. Egala, R. R. D., Otoo, H., Kwofie, C., & Appiah, S. T. (2026). Odd Generalized Exponential-Topp-Leone Weibull Distribution: Actuarial and Statistical Properties with Applications. Statistics, Optimization & Information Computing, 16(4), 3065-3088. https://doi.org/10.19139/soic-2310-5070-3992
11. Gemeay, A. M., Sapkota, L. P., Tashkandy, Y. A., Bakr, M. E., Balogun, O. S., & Hussam, E. (2024). New bounded probability model: Properties, estimation, and applications. Heliyon, 10(23), e38965. https://doi.org/10.1016/j.heliyon.2024.e38965
12. Idemudia, R., & Ekhosuehi, N. (2019). A new three-parameter paralogistic distribution: its properties and application. *Journal of Data Science*, 17(2), 239-258. https://doi.org/10.6339/JDS.201904_17(2).0001
13. Iriarte, Y. A., de Castro, M., & Gómez, H. W. (2020). The Lambert-F Distributions Class: An Alternative Family for Positive Data Analysis. Mathematics, 8(9), 1398. https://doi.org/10.3390/math8091398
14. Johnson, N. L., Kotz, S., & Balakrishnan, N. (1994). Continuous univariate distributions, volume 1 (Vol. 1). John wiley & sons.
15. Karakaya, K., Rajitha, C. S., Sağlam, Ş., Tashkandy, Y. A., Bakr, M. E., Muse, A. H., Kumar, A., Hussam, E., & Gemeay, A. M. (2024). A new unit distribution: properties, estimation, and regression analysis. Scientific reports, 14(1), 7214. https://doi.org/10.1038/s41598-024-57390-7
16. Khalfallah, M. E. (2025). Alpha Power One-Parameter Weibull Distribution: Its Properties, Simulations and Applications to Real-Life Data. Statistics, Optimization & Information Computing, 14(5), 2788-2812. https://doi.org/10.19139/soic-2310-5070-2892
17. Lawless, J. F. (2011). Statistical models and methods for lifetime data. John Wiley & Sons.
18. Meeker, W. Q., Escobar, L. A., & Pascual, F. G. (2021). Statistical methods for reliability data. John Wiley & Sons.
19. Nelson, W. B. (2005). Applied life data analysis. John Wiley & Sons.
20. Pawitan, Y. (2001) In All Likelihood: Statistical Modelling and Inference Using Likelihood. Oxford Science Publications, Clarendon Press, Oxford.
21. Rényi, A. (1961). On measures of entropy and information. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. I: Contributions to the Theory of Statistics (pp. 547-561). University of California Press.
22. Robert, C. P., & Casella, G. (2004). *Monte Carlo Statistical Methods* (2nd ed.). Springer. https://doi.org/10.1007/978-1-4757-4145-2

23. Saadati Nik, A., Asgharzadeh, A., & Baklizi, A. (2023). Inference Based on New Pareto-Type Records With Applications to Precipitation and Covid-19 Data. Statistics, Optimization & Information Computing, 11(2), 243-257. https://doi.org/10.19139/soic-2310-5070-1591
24. Shannon, C.E. (1948), A Mathematical Theory of Communication. Bell System Technical Journal, 27: 379-423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x
25. Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. Biometrika, 72(1), 67-90. https://doi.org/10.1093/biomet/72.1.67
26. Swain, J. J., Venkatraman, S., & Wilson, J. R. (1988). Least-squares estimation of distribution functions in Johnson's translation system. *Journal of Statistical Computation and Simulation*, 29(4), 271-297. https://doi.org/10.1080/00949658808811068
27. Wang, X., Yu, C., & Li, Y. (2015). A new finite interval lifetime distribution model for fitting bathtub-shaped failure rate curve. *Mathematical Problems in Engineering*, 2015, Article ID 954327. https://doi.org/10.1155/2015/954327
28. Wolberg, W., Mangasarian, O., Street, N., & Street, W. (1993). Breast Cancer Wisconsin (Diagnostic) [Dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C5DW2B