Asian Journal of Statistical Sciences
Vol. 4, No-2, (2024), pp. 103-120

ARF INDIA
Academic Open Access Publishing
Journal homepage link:
http://arfjournals.com/ajss

Transformative Distribution Techniques for Data Analysis: A Fusion of Marshall-Olkin and Topp-Leone Nadarajah-Haghighi G Families

Sini K P[a], Davis Antony Mundassery [b] and V M Chacko[a]

[a]*Department of Statistics, St. Thomas College (Autonomous), Thrissur, affiliated to the University of Calicut, 680 001, Kerala, India.* [b]*Department of Statistics, Christ College (Autonomous), Irinjalakkuda, affiliated to the University of Calicut, 680 125, Kerala, India.*

ARTICLE HISTORY
Compiled April 3, 2025

Received 21 June 2024; Accepted 26 December 2024

ABSTRACT
This work presents a novel group of univariate distributions formed by integrating the Marshall-Olkin and the Topp-Leone Nadarajah-Haghighi G families. This new family, designed to enhance flexibility and applicability in data analysis, exhibits unique structural properties that make it suitable for various statistical applications. We discuss three notable members of the proposed family, each demonstrating distinct characteristics and potential use cases with a special case identified within the family. When considering submodels, the densities, as well as hazard rate plots, reveal a range of shapes, showcasing the versatility of the proposed family. A thorough analysis of certain structural characteristics is carried out. Additionally, a characterization derived from truncated moments is presented. Estimating parameters is conducted through maximum likelihood estimation and the efficacy of the same is evidenced through extensive simulation studies. Subsequently, with a specific family member already identified (the baseline distribution being the exponential distribution), a comprehensive analysis has been conducted to apply this novel model using real-life data. Compared to other leading competing models, the new model excels in all evaluated statistical criteria and tests.

KEYWORDS
Data analysis; Marshall-Olkin transformation; Marshall-Olkin Nadarajah-Haghighi Topp-Leone-G family; Maximum likelihood method; Nadarajah-Haghighi distribution; Nadarajah-Haghighi Topp-Leone-G family

1. Introduction

Choosing proper statistical distribution is crucial for making informed data analysis decisions. By selecting a suitable distribution, we can ensure that our models accurately reflect the underlying data and provide more accurate insights. In finance, healthcare, and marketing fields, choosing proper distribution for real-life data is crucial, where even minor errors in decision-making can have significant consequences. So, whenever analyzing sales data, patient outcomes, or customer behavior, carefully consider which statistical distribution is most appropriate for the present needs. Selecting a well-suited distribution in data modeling and analysis is vital to achieving precise decisions.

CONTACT Sini K P[a]. Email:sinikpstat@gmail.com

Throughout the years, a variety of distribution classes introduced in the literature to accommodate different data forms, including symmetric, skewed, multimodal, and heavy-tailed. Kotz and Vicari [26] highlighted key milestones in this development, including significant methods like the system of differential equations, the Method of Transformation (Translation), and the Method of Quantiles.

In recent decades, there has been a revived focus on creating more adaptable and versatile distributions. Methodologies for generating flexible distributions have shifted towards integrating additional parameters into existing distributions [18], generating skew distributions [3], the beta-generated method [6], the Transformed-transformer method (T-X family) [2], and the Composite method [17]. It is fascinating to observe how various distributions are applied to model real-world phenomena, from stock prices to weather patterns. However, relying solely on classical distributions for fitting these data sets may lead to unreliable outcomes. Therefore, for the most dependable results, it is essential to explore alternative methods and consider modifications to keep pace with the evolving world.

Researchers have increasingly recognized the effectiveness of adding new parameters to pre-existing distributions. Among these, the Marshall-Olkin (MO) family [18], and the T-X family [2] are particularly notable. These approaches have led to the generalization of several distribution families. Refer[1, 5, 12, 13, 20, 21, 23].

The survival function (*sf*) and the probability distribution function (*pdf*) for the MO family are defined in (1) and (2), respectively.

$$\overline{H}(y) = \frac{\alpha F(y)}{F(y) + \alpha \overline{F}(y)}, \quad \alpha > 0, \quad y \in \mathbf{R}. \tag{1}$$

$$h(y) = \frac{\alpha f(y)}{(1 - \overline{\alpha}\overline{F}(y))^2}. \tag{2}$$

The Nadarajah Haghighi Topp Leone-G (NHTL-G) family [22], serves as an instance of the T-X family. It is derived from the Nadarajah Haghighi (NH) distribution [19], and utilizes the parameters β and λ as its generators. For any distribution characterized by a *cdf* $G(y;\zeta)$, the *cdf* of the NHTL-G family, along with its corresponding *pdf* can be represented as follows

$$F(y;\beta,\lambda,\theta,\zeta) = 1 - e^{1-\{1+\lambda[1-\overline{G}(y;\zeta)^2]^\theta\}^\beta}, \quad y \in \mathbf{R}. \tag{3}$$

$$\begin{aligned} f(y;\beta,\lambda,\theta,\zeta) = {} & 2\beta\lambda\theta g(y;\zeta)(1-\overline{G}(y;\zeta)^2)^{\theta-1}\overline{G}(y;\zeta) \\ & \{1+\lambda[1-\overline{G}(y;\zeta)^2]^\theta\}^{\beta-1} \\ & e^{1-\{1+\lambda[1-\overline{G}(y;\zeta)^2]^\theta\}^\beta}, \quad y \in \mathbf{R}. \end{aligned} \tag{4}$$

The present paper merges the MO-G family with the NHTL-G family to offer a broader choice of distributions. This development results in a more versatile set of options that can better serve different types of users. Combining two generator families of distributions has become a popular and practical trend in distribution theory.

Researchers such as [15] and [11] are actively exploring combinations and their properties, uncovering new patterns and relationships. This approach has provided new insights and expanded the range of applications, leading to significant progress in the field.

This paper is organized into several sections to explore the Marshall-Olkin Nadarajah-Haghighi Topp-Leone-G (MONHTL-G) family. Section 2 provides an overview of this family, while Section 3 examines some of its sub-models. In Section 4, we discuss various statistical properties of the family. Section 5 focuses on characterization derived from truncated moments. Section 6 looks into the use of the maximum likelihood estimation (MLE) approach for estimating parameters. followed by a comprehensive simulation study in Section 7. A practical application is presented in Section 8, and ultimately, the paper concludes in Section 9.

2. MONHTL-G family

This section delves into the MONHTL-G family of distributions, which possesses distinctive and advantageous properties, making it highly suitable for various applications. For any continuous baseline distribution, we construct the new family of distributions by integrating the well-established MO family and NHTL-G family. The corresponding *cdf* is expressed in (5) and the associated *pdf* is in (6).

$$F^{MONHTLG}(y;\zeta) = \frac{1-e^{1-\{1+\lambda[1-\overline{G}(y;\Phi)^2]^\theta\}^\beta}}{\alpha+\overline{\alpha}[1-e^{1-\{1+\lambda[1-\overline{G}(y;\Phi)^2]^\theta\}^\beta}]}. \tag{5}$$

$$f^{MONHTLG}(y;\zeta) = \frac{2\alpha\beta\lambda\theta g(y;\Phi)\overline{G}(y;\Phi)(1-\overline{G}(y;\Phi)^2)^{\theta-1}}{(\alpha+\overline{\alpha}[1-e^{1-\{1+\lambda[1-\overline{G}(y;\Phi)^2]^\theta\}^\beta}])^2} \{1+\lambda[1-\overline{G}(y;\Phi)^2]^\theta\}^{\beta-1}e^{1-\{1+\lambda[1-\overline{G}(y;\Phi)^2]^\theta\}^\beta}, \tag{6}$$

where $\alpha, \beta, \lambda, \theta$ are all positive, and $G(y;\Phi)$, where Φ represents the parameter set, is the *cdf* of the baseline distribution.

2.1. Quantile function

Let us define

$$u = e^{1-(1+\lambda(1-\overline{G}(y;\Phi)^2)^\theta)^\beta}.$$

Then (5) can be reformulated as

$$F^{MONHTLG}(y;\zeta) = \frac{1-u}{\alpha+\overline{\alpha}(1-u)}. \tag{7}$$

The quantile function (qf) of the MONHTL-G family, denoted as $Q(w) = F^{-1}(w)$, for $w \in (0,1)$, $\alpha \neq 0$, $\beta \neq 0$, $\lambda \neq 0$, $\theta \neq 0$, is the solution to the non-linear equation (8).

$$Q(w) = G^{-1}\left(1 - \left[1 - \left(\frac{1}{\lambda}\left\{\left[1 - \log\left(\frac{1-w}{1-w\overline{\alpha}}\right)\right]^{1/\beta} - 1\right\}\right)^{1/\theta}\right]^{1/2}\right). \tag{8}$$

3. The Sub-models within the MONHTL-G family

This section delves into three notable sub-models within the MONHTL-G family, each distinguished by unique characteristics and advantages. We will explore these sub-models in depth to highlight their individual contributions and strengths. By the end of this section, readers will have a comprehensive understanding of the MONHTL-G family and the versatility and power of its sub-models.

3.0.1. The MONHTL-Lomax (MONHTLLx) Model

Here, we select a random variable (r.v.) X following the Lomax (Lx) distribution with the *sf* in (9) as the baseline distribution to characterize MONHTLLx. The resulting *cdf* is obtained in (10) and the corresponding *pdf* is in (11).

$$\overline{G}(x;\Phi) = (1+bx)^{-a}, \quad x \geq 0, \quad a, b > 0. \tag{9}$$

$$F_{MONHTLLx}(x;\zeta) = \frac{1 - e^{1-\{1+\lambda[1-(1+bx)^{-2a}]^{\theta}\}^{\beta}}}{\alpha + \overline{\alpha}[1 - e^{1-\{1+\lambda[1-(1+bx)^{-2a}]^{\theta}\}^{\beta}}]}. \tag{10}$$

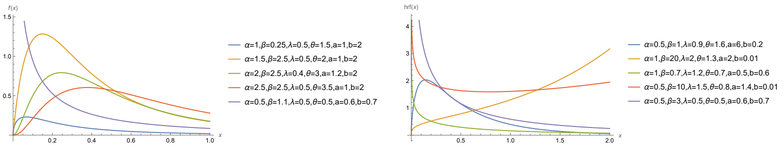
$$f^{MONHTLLx}(x;\zeta) = \frac{2\alpha\beta\lambda\theta ab(1+bx)^{-2a}(1-(1+bx)^{-2a})^{\theta-1}}{(\alpha + \overline{\alpha}[1 - e^{1-\{1+\lambda[1-(1+bx)^{-2a}]^{\theta}\}^{\beta}}])^2} \{1+\lambda[1-(1+bx)^{-2a}]^{\theta}\}^{\beta-1} e^{1-\{1+\lambda[1-(1+bx)^{-2a}]^{\theta}\}^{\beta}}. \tag{11}$$

Figure 1 displays the *pdf* and hazard function (*hrf*) graphs for some random parameter values. The *pdf* plot illustrates the impact of parameter adjustments on the distribution's form. Certain curves display unimodal properties, while others demonstrate a monotonically decreasing tendency. Moreover, specific cases demonstrate heavier tails, signifying an increased likelihood of extreme values. The *hrf* figure illustrates various hazard rate patterns, including increasing decreasing, and bathtub-shaped trends. These differences underscore the adaptability of the MONHTLLx distribution in modeling various real-world circumstances, including reliability analysis and survival studies.

3.0.2. The MONHTL-Kumaraswamy (MONHTLKw) Model

Consider the Kumaraswamy distribution with *sf* in (12).

$$\overline{G}(x;\Phi) = (1-x^a)^b, \quad 0 \leq x \leq 1, \quad a, b > 0. \tag{12}$$



(a) The *pdf* plot (b) The *hrf* plot.

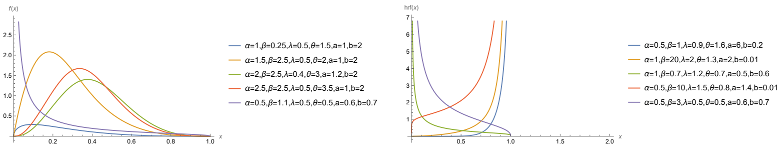
Figure 1. The MONHTLLx *pdf* and *hrf* plot for random values of parameters.

Then the corresponding MONHTLKw distribution is defined in (13) and (14).

$$F^{MONHTLKw}(x;\zeta) = \frac{1-e^{1-\{1+\lambda(1-x^a)^{b\theta}\}^\beta}}{\alpha+\overline{\alpha}[1-e^{1-\{1+\lambda(1-x^a)^{b\theta}\}^\beta}]}, \tag{13}$$

$$f^{MONHTLKw}(x;\zeta) = \frac{2\alpha\beta\lambda\theta abx^{a-1}(1-x^a)^{2b-1}[(1-x^a)^{2b}]^{\theta-1}}{\{\alpha+\overline{\alpha}[1-e^{1-\{1+\lambda(1-x^a)^{b\theta}\}^\beta}]\}^2} \{1+\lambda[1-(1-x^a)^{2b}]^\theta\}^{\beta-1}e^{1-\{1+\lambda[1-(1-x^a)^{2b}]^\theta\}^\beta}. \tag{14}$$

The density and *hrf* plots for MONHTL-Kw distribution in Fig. 2 effectively capture unimodal, skewed, and long-tailed distributions, as well as ascending, descending, and bathtub-like hazard rates. This makes it suitable for applications in reliability analysis, survival studies, and risk assessment.



(a) The *pdf* plot (b) The *hrf* plot.

Figure 2. The MONHTLKw *pdf* and *hrf* plot for random parameter values.

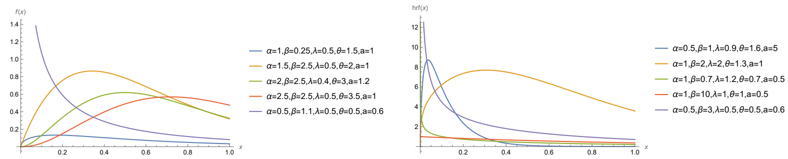
3.0.3. The MONHTL-Exponential (MONHTLEx) Model

Finally, we consider the exponential (Ex) distribution with *cdf* $\overline{G}(x;\lambda) = e^{-\lambda x}; \quad x \geq 0, \quad \lambda > 0$. Then the *cdf* MONHTLEx distribution is obtained in (15) and the associated *pdf* is in (16).

$$F^{MONHTLEx}(x;\zeta) = \frac{1-e^{1-[1+\lambda(1-e^{-2ax})^\theta]^\beta}}{\alpha+\overline{\alpha}[1-e^{1-[1+\lambda(1-e^{-2ax})^\theta]^\beta}]}, \tag{15}$$

$$f^{MONHTLEx}(x;\zeta) = \frac{2a\alpha\beta\lambda\theta e^{-2ax}\{1+\lambda[1-e^{-2ax}]^\theta\}^{\beta-1}(1-e^{-2ax})^{\theta-1}}{\{\alpha+\overline{\alpha}[1-e^{1-[1+\lambda(1-e^{-2ax})^\theta]^\beta}]\}^2} e^{1-\{1+\lambda[1-e^{-2ax}]^\theta\}^\beta}. \tag{16}$$

Similar to other sub-modals, the MONHTLEx *pdf* captures unimodal, skewed, and heavy-tailed distributions, while *hrf* exhibits increasing, decreasing, and bathtub-shaped patterns, as illustrated in Fig. 3.



(a) The *pdf* plot (b) The *hrf* plot.

Figure 3. The MONHTLEx *pdf* and *hrf* plot for random parameter values.

3.1. Special Cases

Choosing $\alpha = \frac{1}{\gamma}$ in (1), it reduces to

$$F(x) = \frac{\gamma G(y;\theta)}{1-(1-\gamma)G(y;\theta)},$$

which is the Geometric Generated family. Substituting baseline distribution as NHTLG, we get another novel generalized family. That is, the Geometric generalized NHTLG family.

4. Some Statistical and mathematical properties of MONHTL-G family

4.1. Order Statistics

Here, we explore the core concepts of order statistics within the framework of the MONHTL-G family.

Let $Y_1, Y_2, ..., Y_n$ be a simple random sample (r.s.) that follows the proposed distribution with *pdf* as stated in (5), and the associated order statistics, denoted as $Y_{1:n}, Y_{2:n}, ...Y_{n:n}$, are obtained accordingly. The *pdf* for $Y_{s:n}$, the s^{th} order statistics, is readily available as

$$f^{MONHTLG}_{Y_{s:n}}(y) = \frac{(-1)^s}{\beta(s, n-s+1)} \sum_{w=0}^{n-s} \binom{n-s}{w} f(y)F(y)^{s+w-1}, \tag{17}$$

where $\beta(.,.)$ is the beta function.

Now,

$$f(y)F(y)^{s+w-1} = \frac{2\alpha\beta\lambda\theta g(y;\Phi)\overline{G}(y;\Phi)\left[1-\overline{G}(y;\Phi)^2\right]^{\theta-1}\left\{1+\lambda[1-\overline{G}(y;\Phi)^2]^{\theta}\right\}^{\beta-1}}{\{\alpha+\overline{\alpha}[1-e^{1-\left\{1+\lambda[1-\overline{G}(y;\Phi)^2]^{\theta}\right\}^{\beta}}]\}^{s+w+1}}$$
$$e^{[1-\{1+\lambda[1-\overline{G}(y;\Phi)^2]^{\theta}\}^{\beta}]}\left[1-e^{1-\{1+\lambda[1-\overline{G}(y;\Phi)^2]^{\theta}\}^{\beta}}\right]^{s+w-1}. \tag{18}$$

Rewriting the denominator as (19) and making use of binomial expansion in (20),

(18) can be expressed by (21).

$$\left\{\alpha+\overline{\alpha}\left[1-e^{1-\left\{1+\lambda\left[1-\overline{G}(y;\Phi)^2\right]^{\theta}\right\}^{\beta}}\right]\right\}^{s+w+1}=\left\{1-\overline{\alpha}\left\{1-\left[1-e^{1-\left\{1+\lambda\left[1-\overline{G}(y;\Phi)^2\right]^{\theta}\right\}^{\beta}}\right]\right\}\right\}^{s+w+1} \tag{19}$$

$$(1-y)^n=\sum_{s=0}^{\infty}\binom{n}{s}(-y)^s, \quad \text{for } |y|<1. \tag{20}$$

$$f(y)F(y)^{s+w-1}=\sum_{q=0}^{\infty}\mu_q\Pi_{q+1}(y), \tag{21}$$

where

$$\mu_q=2e\alpha\beta\lambda\theta\sum_{j,i,l=0}^{\infty}\sum_{m,n,p=0}^{\infty}\binom{s+w+j}{j}\binom{s+w+i-1}{l}\binom{\beta(m+1)-1}{n}$$
$$\binom{\theta(n+1)-1}{p}\binom{j}{i}\binom{2p+1}{q}\frac{(-1)^{i+l+m+p+q}(1-\alpha)^j\lambda^n(l+1)^m e^l}{m!},$$

and

$$\Pi_{q+1}(y)=g(y;\Phi)G(y;\Phi)^q. \tag{22}$$

Substituting (21) into (18), we get

$$f_{Y_{s:n}}^{MONHTLG}(y)=\sum_{q=0}^{\infty}\mu_q^*\Pi_{q+1}(y), \tag{23}$$

where

$$\mu_q^*=\frac{(-1)^s}{\beta(s,n-s+1)}\sum_{w=0}^{n-s}\binom{n-s}{w}\mu_q. \tag{24}$$

In addition, the r^{th} moment of the s^{th} order statistic for this family can be expressed as

$$E(y_{s:n}^r)=\sum_{q=0}^{\infty}\mu_q^{**}\Pi_{q+1}(y), \tag{25}$$

where, $\mu_q^{**}=(q+1)\mu_q^*$.

4.2. Expansion of the density function

To express both the *pdf* and the *cdf* of the new family as mixture representations of the exponentiated-G (Exp-G) distribution, we can express (6) in the following manner:

$$f^{MONHTLG}(y) = 2e\alpha\beta\lambda\theta \sum_{i,j,k=0}^{\infty} \sum_{l,m,n=0}^{\infty} \sum_{q=0}^{2n+1} (-1)^{i+j+l+n+q} \frac{e^i \lambda^m (i+1)^l (1-\alpha)^i}{l!} \binom{i}{j} \binom{j}{k} \binom{\beta(l+1)-1}{m} \binom{\theta(m+1)-1}{n} \binom{2n+1}{q} g(y;\Phi) G(y;\Phi)^q.$$

Or equivalently,

$$f^{MONHTLG}(y) = \sum_{q=0}^{2n+1} \delta_q \Pi_{q+1}(y), \quad (26)$$

where,

$$\delta_q = 2e\alpha\beta\lambda\theta \sum_{i,j,k=0}^{\infty} \sum_{l,m,n=0}^{\infty} \sum_{q=0}^{2n+1} (-1)^{i+j+l+n+q} \frac{e^i \lambda^m (i+1)^l (1-\alpha)^i}{(q+1)l!} \binom{i}{j} \binom{j}{k} \binom{\beta(l+1)-1}{m} \binom{\theta(m+1)-1}{n} \binom{2n+1}{q}$$

and

$$\Pi_{q+1}(y) = (q+1) g(y;\Phi) G(y;\Phi),$$

which corresponds to the Exp-G distribution with power parameter (q+1). Similarly, *cdf* in (5) can be represented as

$$F(y) = \sum_{q=0}^{2n+1} \delta_q \Phi_{q+1}(y), \quad (27)$$

where $\Phi_{q+1}(y) = G(y;\Phi)^{q+1}$.

4.3. Moments

The moments function is employed to examine various essential distributional characteristics, such as dispersion, kurtosis, asymmetry, and central tendency. Then the s^{th}

moment about the origin for the MONHTL-G family is:

$$\begin{aligned}\mu_s &= \int_0^\infty y^s f^{MONHTLG}(y;\zeta)\,dy \\ &= \Delta \int_0^\infty y^s g(y;\Phi) G(y;\Phi)^q\,dy \\ &= \sum_{q=0}^{2n+1} \Delta\Phi_{s,q}, \end{aligned} \tag{28}$$

where $\Delta = (q+1)\delta_q$.

4.4. Moment Generating Function (MGF)

For the MONHTL-G family, the MGF can be expressed as given in (29).

$$\begin{aligned} M_y(t) &= \sum_{s=0}^{\infty} \frac{t^s}{s!} E[y^s] \\ &= \sum_{s=0}^{\infty} \frac{t^s}{s!} \sum_{n=0}^{\infty} \Delta\Phi_{s,2n+1}, \end{aligned} \tag{29}$$

where

$$\Delta = 2e\alpha\beta\lambda\theta \sum_{i,j,k=0}^{\infty} \sum_{l,m=0}^{\infty} (-1)^{i+j+l+n} \frac{e^i}{l!}(k+1)(1-\alpha)^k \lambda^m$$
$$(i+1)^l \binom{i}{j}\binom{j}{k}\binom{\beta(l+1)-1}{m}\binom{\theta(m+1)-1}{n},$$

and

$$\Phi_{s,2n+1} = \int_0^\infty y^s g(y;\Phi)\overline{G}(y;\Phi)^{2n+1}\,dy.$$

4.5. Probability Weighted Moments (PWM)

The PWM of order $(r+s)$ for the r.v. Y under the new model, denoted as $\nu_{r,s}$, is given by (30).

$$\nu_{r,s} = E(Y^r F(y)^s) = \int_{-\infty}^{\infty} Y^r F(y)^s f(y) dy. \tag{30}$$

Considering,

$$f(y)F(y)^s = 2e\alpha\beta\lambda\theta \sum_{i,j,k=0}^{\infty} \sum_{l,m,n=0}^{\infty} (-1)^{i+k+l+n} \frac{e^k(k+1)^l \lambda^m (1-\alpha)^j}{l!} \binom{j}{i} \binom{i+s}{k} \binom{2+s+j-1}{j} \binom{\beta(l+1)-1}{m} \binom{\theta(m+1)-1}{n} g(y;\Phi)\overline{G}(y;\Phi)^{2n+1}. \tag{31}$$

Or equivalently

$$f(y)F(y)^s = \sum_{q=0}^{2n+1} b_q \Pi_{q+1}(y), \tag{32}$$

where

$$b_q = 2e\alpha\beta\lambda\theta \sum_{i,l,k=0}^{\infty} \sum_{l,m,n=0}^{\infty} \frac{(-1)^{i+k+n} e^k (k+1)^l \lambda^m (1-\alpha)^j}{l!} \binom{j}{i} \binom{i+s}{k} \binom{2+s+j-1}{j} \binom{\beta(l+1)-1}{m} \binom{\theta(m+1)-1}{n},$$

and

$$\Pi_{q+1}(y) = g(y;\Phi)\overline{G}(y;\Phi)^{2n+1}. \tag{33}$$

Substituting the above result to (30), we can achieve (34).

$$\nu_{r,s} = \sum_{q=0}^{2n+1} b_q^* \Psi_{r,q}, \tag{34}$$

where $b_q^* = (q+1)b_q$ and $\Phi_{r,q} = \int_{-\infty}^{\infty} y^r g(y)G(y)^q dx$ is the PWM of the parent distribution.

5. Characterizations

Probability distributions can be characterized in multiple manners, one of which includes truncated moments. This methodology was first presented by [7] and subsequently examined by many scholars including [8, 10, 14, 16, 25], among others. It is important to recognize that obtaining characterization results in this context frequently necessitates extensive mathematical methods. This section analyzes particular characterization of the MONHTL-G family through the truncated moment (conditional expectation) of specific functions of a r.v..

We revisit Theorem 10.1 by [8]. As demonstrated by [9], this characterization remains stable under weak convergence.

Proposition 5.1. *Let $Y:\Omega\rightarrow$ (0,∞) be a continuous r.v. and let*

$$q_1(y)=[\alpha+(1-\alpha)(1-e^{\nabla(y)})]^2$$

where $\nabla(y)\ =1-\{1+\lambda[1-\overline{G}(y;\Phi)^2]^\theta\}^\beta$ *and*

$$q_2(y)=q_1(y)e^{\nabla(y)},\quad y>0.$$

Thus, Y has pdf in (6) if and only if the function ζ defined in Theorem 10.1 is of the form.

Proof. Suppose Y is a r.v. with *pdf* in (6), then

$$(1-F(y))E[q_1(Y)|Y\geq y]=e^{\nabla(y)},\quad y>0,$$

and

$$(1-F(y))E[q_2(Y)|Y\geq y]=\frac{1}{2}e^{2\nabla(y)}.$$

Hence

$$zeta(y)=\frac{1}{2}e^{\nabla(y)},\quad y>0.$$

We also have

$$\zeta(y)q_1(y)-q_2(y)=\frac{-1}{2}q_1(y)e^{\nabla(y)}<0,\quad y>0.$$

Conversely, if ζ is of the above form, then

$$\begin{aligned}s'(y)&=\frac{\zeta^{'}(y)q_1(y)}{\zeta(y)q_1(y)-q_2(y)}\\&=-2\beta\lambda\theta\{1+\lambda[1-\overline{G}(y;\Phi)^2]^\theta\}^{\beta-1}[1-\overline{G}(y;\Phi)^2]^{\theta-1}\overline{G}(y;\Phi)\end{aligned}$$

As stated in Theorem 10.1, Y follows a density in (6). □

Corollary 5.2. *Suppose Y is a continuous r.v.. Let $q_1(y)$ be as in Proposition 5.1. Then Y has a density in (6) if and only if there exist functions $q_2(y)$ and $\zeta(y)$ defined in Theorem 10.1 for which the following first-order differential equation holds*

$$\begin{aligned}s'(x)&=\frac{\zeta^{'}(y)q_1(y)}{\zeta(y)q_1(y)-q_2(y)}.\\&=-2\beta\lambda\theta\{1+\lambda[1-\overline{G}(y;\Phi)^2]^\theta\}^{\beta-1}[1-\overline{G}(y;\Phi)^2]^{\theta-1}\overline{G}(y;\Phi).\end{aligned}$$

The general solution for the differential equation in the above corollary is

$$\zeta(y)=e^{-\nabla(y)}\left[\int 2\beta\lambda\theta\{1+\lambda[1-\overline{G}(y;\Phi)^2]^\theta\}^{\beta-1}[1-\overline{G}(y;\Phi)^2]^{\theta-1}\overline{G}(y;\Phi)e^{\nabla(y)}(q_1(y))^{-1}q_2(y)+D\right]$$

where D is a constant. Proposition 5.1 provides a set of functions satisfying the above differential equation with D = 0.

6. Parameter Estimation

This section examines the MONHTL-G family, utilizing the maximum likelihood estimation (MLE) procedure to estimate the parameters because of its significant theoretical and practical benefits. Given a n r.s. extracted from the distribution of Y, the log-likelihood function, employing the *pdf* specified in (6), can be expressed as

$$L(y;\zeta)=\prod_{i=1}^{n}\left\{\frac{2\alpha\beta\lambda\theta g(y_i;\Phi)\overline{G}(y_i;\Phi)(1-\overline{G}(y_i;\Phi)^2)^{\theta-1}}{(\alpha+\overline{\alpha}[1-e^{1-\{1+\lambda[1-\overline{G}(y_i,\Phi)^2]^\theta\}^\beta}])^2}\right\}$$
$$\left\{\{1+\lambda[1-\overline{G}(y_i;\Phi)^2]^\theta\}^{\beta-1}e^{1-\{1+\lambda[1-\overline{G}(y_i,\Phi)^2]^\theta\}^\beta}\right\}.$$

While the log-likelihood ($\log L$) function is

$$\begin{aligned}\log L &= nlog2+nlog\alpha+nlog\beta+nlog\lambda+nlog\theta+\sum_{i=1}^{n}logg(y_i;\Phi)+\sum_{i=1}^{n}log\overline{G}(y_i;\Phi)\\ &+(\theta-1)\sum_{i=1}^{n}log[1-\overline{G}(y_i;\Phi)^2]+(\beta-1)\sum_{i=1}^{n}log\{1+\lambda[1-\overline{G}(y_i;\Phi)^2]^\theta\}\\ &-\sum_{i=1}^{n}\{1+\lambda[1-\overline{G}(y_i;\Phi)^2]^\theta\}^\beta-2\sum_{i=1}^{n}log(\alpha+\overline{\alpha}[1-e^{1-\{1+\lambda[1-\overline{G}(y_i;\Phi)^2]^\theta\}^\beta}]).\end{aligned} \tag{35}$$

Following standard practice, partial derivatives of $\log L$ are set to zero, as demonstrated below.

$$\frac{\partial\log L}{\partial\alpha}=\frac{n}{\alpha}-2\sum_{i=1}^{n}\frac{1}{\alpha+\overline{\alpha}\left[1-e^{1-\left\{1+\lambda[1-\overline{G}(x_i;\Phi)^2]^\theta\right\}^\beta}\right]}\left(1-e^{1-\left\{1+\lambda[1-\overline{G}(y_i;\Phi)^2]^\theta\right\}^\beta}\right) \tag{36}$$

$$\begin{aligned}\frac{\partial\log L}{\partial\lambda}&=\frac{n}{\lambda}+(\beta-1)\sum_{i=1}^{n}\frac{[1-\overline{G}(y_i;\Phi)^2]^\theta}{1+\lambda[1-\overline{G}(y_i;\Phi)^2]^\theta}-2\sum_{i=1}^{n}\frac{\overline{\alpha}\left[1-e^{1-\left\{1+\lambda[1-\overline{G}(y_i;\Phi)^2]^\theta\right\}^\beta}\right]}{\alpha+\overline{\alpha}\left[1-e^{1-\left\{1+\lambda[1-\overline{G}(y_i;\Phi)^2]^\theta\right\}^\beta}\right]}\\ &\left(-e^{1-\left\{1+\lambda[1-\overline{G}(y_i;\Phi)^2]^\theta\right\}^\beta}\left\{1+\lambda[1-\overline{G}(y_i;\Phi)^2]^\theta\right\}^{\beta-1}[1-\overline{G}(y_i;\Phi)^2]^\theta\right)\\ &+\sum_{i=1}^{n}\left(-\left\{1+\lambda[1-\overline{G}(y_i;\Phi)^2]^\theta\right\}^{\beta-1}[1-\overline{G}(y_i;\Phi)^2]^\theta\right)\end{aligned} \tag{37}$$

$$
\begin{aligned}
\frac{\partial \log L}{\partial \beta} &= \frac{n}{\beta} + \sum_{i=1}^{n} \frac{(\beta-1)\lambda[1-\overline{G}(y_i;\Phi)^2]^\theta}{1+\lambda[1-\overline{G}(y_i;\Phi)^2]^\theta} - 2\sum_{i=1}^{n} \frac{\overline{\alpha}\left[1-e^{1-\left\{1+\lambda[1-\overline{G}(y_i;\Phi)^2]^\theta\right\}^\beta}\right]}{\alpha+\overline{\alpha}\left[1-e^{1-\left\{1+\lambda[1-\overline{G}(y_i;\Phi)^2]^\theta\right\}^\beta}\right]} \\
&\left(-e^{1-\left\{1+\lambda[1-\overline{G}(y_i;\Phi)^2]^\theta\right\}^\beta}\left\{1+\lambda[1-\overline{G}(y_i;\Phi)^2]^\theta\right\}^\beta \log\left\{1+\lambda[1-\overline{G}(y_i;\Phi)^2]^\theta\right\}\right) \\
&+ \sum_{i=1}^{n}\left(-\left\{1+\lambda[1-\overline{G}(y_i;\Phi)^2]^\theta\right\}^\beta \log\left\{1+\lambda[1-\overline{G}(y_i;\Phi)^2]^\theta\right\}\right) \qquad (38)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \log L}{\partial \theta} &= \frac{n}{\theta} + \sum_{i=1}^{n} \log(1-\overline{G}(y_i;\Phi)^2) + (\beta-1)\sum_{i=1}^{n} \frac{\lambda[1-\overline{G}(y_i;\Phi)^2]^\theta \log[1-\overline{G}(y_i;\Phi)^2]}{1+\lambda[1-\overline{G}(y_i;\Phi)^2]^\theta} \\
&- 2\sum_{i=1}^{n} \frac{\overline{\alpha}\left[1-e^{1-\left\{1+\lambda[1-\overline{G}(y_i;\Phi)^2]^\theta\right\}^\beta}\right]}{\alpha+\overline{\alpha}\left[1-e^{1-\left\{1+\lambda[1-\overline{G}(y_i;\Phi)^2]^\theta\right\}^\beta}\right]} \\
&\left(-e^{1-\left\{1+\lambda[1-\overline{G}(y_i;\Phi)^2]^\theta\right\}^\beta}\left\{1+\lambda[1-\overline{G}(y_i;\Phi)^2]^\theta\right\}^{\beta-1} \lambda[1-\overline{G}(y_i;\Phi)^2]^\theta \log[1-\overline{G}(y_i;\Phi)^2]\right) \\
&+ \sum_{i=1}^{n}\left(-\left\{1+\lambda[1-\overline{G}(y_i;\Phi)^2]^\theta\right\}^{\beta-1} \lambda[1-\overline{G}(y_i;\Phi)^2]^\theta \log[1-\overline{G}(y_i;\Phi)^2]\right). \qquad (39)
\end{aligned}
$$

Due to the numerous parameters involved, finding a solution to the equation system where each derivative equals zero is extremely laborious. Software like Mathematica, R, and Python can be used in such cases to obtain maximum likelihood estimates. In Mathematica, the $Maximize$ function optimizes parameter estimates while enforcing positivity constraints. In R, optimization methods such as *nlm* (nonlinear minimization), *optim*, and *constrOptim* provide flexible and efficient solutions for both constrained and unconstrained optimization problems. In Python, the *scipy.optimize* module, specifically *minimize*, $fmin$, and $differential_evolution$, is used to efficiently solve the system of equations.

7. Simulation

The Monte Carlo (MC) simulation approach analyzes the performance of estimators for the parameters associated with the newly proposed MONHTL-Ex distribution. For each experiment, 1000 pseudo-random samples were generated from the proposed model using selected population parameter values and sample sizes. The selected parameter combination is $(\alpha, \beta, \lambda, \theta, a) = (0.5, 1.5, 1.5, 2, 1.2)$. Table 1 presents the simulation outcomes, whereas Figure 4 illustrates the decreasing trend of bias as the sample size increases.

Table 1. MONHTL-Ex: Estimates, Bias, and RMSE for different parameters and sample sizes.

n	Parameter	Estimates	Bias	RMSE
50	α	0.2883	-0.2117	0.2669
	β	1.5843	0.0843	0.092
	λ	1.4106	-0.0894	0.1298
	θ	3.3281	1.3281	1.3725
	a	1.9776	0.7776	0.7983
100	α	0.2898	-0.2102	0.2594
	β	1.5822	0.0822	0.0884
	λ	1.4469	-0.0531	0.0732
	θ	3.3383	1.3383	1.3754
	a	2.0031	0.8031	0.8176
500	α	0.2930	-0.2070	0.2630
	β	1.5800	0.0800	0.0910
	λ	1.4510	-0.0490	0.0910
	θ	3.3050	1.3050	1.3600
	a	1.9710	0.7710	0.7910
1000	α	0.294	-0.206	0.26
	β	1.578	0.078	0.119
	λ	1.451	-0.049	0.091
	θ	3.295	1.295	1.35
	a	1.961	0.761	0.78

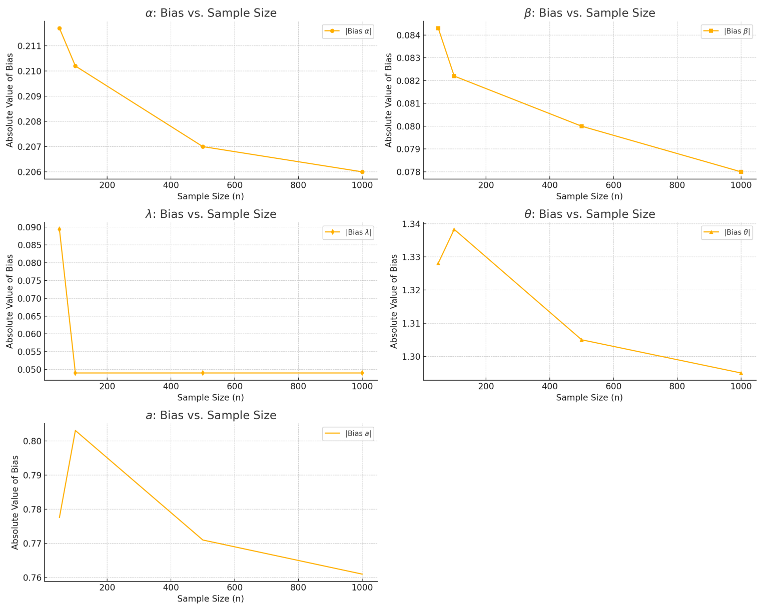


Figure 4. Bias v/s Sample size for each parameter.

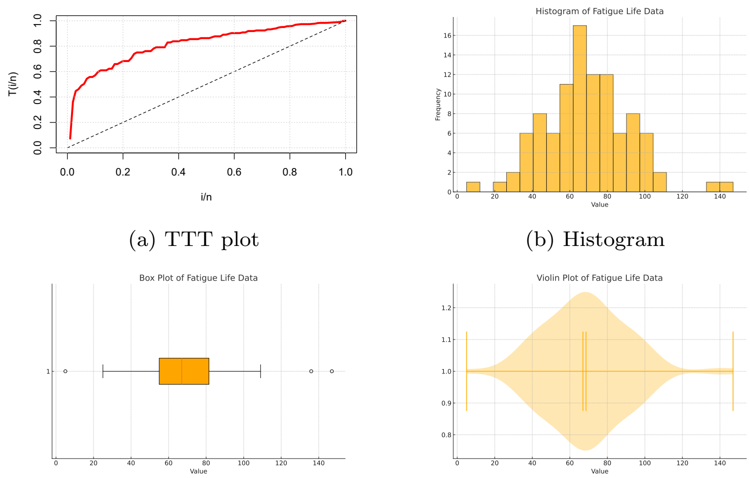
8. Data Analysis

This section demonstrates how the MONHTLEx distribution can be effectively utilized in real-world contexts. We examine the goodness-of-fit matrics and maximum likelihood estimates (MLEs) of model parameters between the MONHTLEx model and its competitors. The competing models selected are: MO Topp-Leone Weibull (MOTLWe), MO Topp-Leone Burr XII (MOTLBXII), and MO Topp-Leone Lomax (MOTLLo) distributions. To determine the most effective model, we utilized R software to compute several statistical metrics, including the Kolmogorov-Smirnov (KS) statistic, the Akaike Information Criterion (AICr) and its adjusted variant (CAICr), the Bayesian Information Criterion (BICr), the Hannan-Quinn Information Criterion (HQICr), and the negative log-likelihood.

To demonstrate this, we use the fatigue data from Birnbaum and Saunders [4], which yielded a p-value of 0.985. This dataset, presented in Table 2, has been referenced in recent studies [24], and its descriptive statistics, including skewness and kurtosis, are provided in Table 3. The TTT plot, histogram, box plot, and violin plot of the data are shown in Figure 5. The TTT plot supports models with decreasing hazard rates. Table 4 lists the estimated parameter values, while Table 5 presents statistical metrics, clearly demonstrating that our model achieves the lowest values across all conditions.

Table 2. The fatigue life of 6061 - T6 aluminum coupons cut parallel to the direction of rolling and oscillated at 18 cycles per second. The data set consists of 101 observations with a maximum stress per cycle of 31,000 psi. The data ($\times 10^{-3}$) are presented below (after subtracting 65).

5	25	31	32	34	35	38	39	39	40
42	43	43	43	44	44	47	47	48	49
49	49	51	54	55	55	55	56	56	56
58	59	59	59	59	63	63	64	64	64
65	65	65	66	66	66	66	67	67	67
67	68	69	69	69	71	71	72	73	73
73	73	74	74	76	76	76	77	77	77
77	79	79	80	81	83	84	84	86	86
87	90	91	92	92	92	92	93	94	97
98	98	98	99	101	103	105	109	136	147



(a) TTT plot (b) Histogram

(c) Boxplot (d) Violin plot

Figure 5. TTT plot, histogram, box plot, and violin plot of the fatigue data.

9. Conclusion

This research presents a novel class of distributions, referred to as the MONHTL-G family, which is formed by combining the Topp-Leone, Nadarajah-Haghighi, and Marshall-Olkin distributions. As a result, the proposed family includes distributions

Table 3. Descriptive Statistics of Fatigue Life Data.

Statistic	Value
Count	100
Mean	68.64
Standard Deviation	22.58
Minimum	5.0
25th Percentile	55.0
Median (50th Percentile)	67.0
75th Percentile	81.5
Maximum	147.0
Skewness	0.383
Kurtosis	1.139

Table 4. Estimates of the competitive models fitted to the fatigue data.

Model	Estimates				
MONHTLEx$(\theta, \beta, \alpha, \lambda, a)$	1.998735	2.004019	0.9977184	1.006349	1.674635
MOTLWe$(\theta, \beta, \alpha, \lambda)$	24.1879	20.1026	0.0155	1.5561	
MOTLBXII$(\theta, \beta, \alpha, \lambda)$	636.0990	634.5008	0.3459	3.8591	
MOTLLo$(\theta, \alpha, \lambda)$	5.8836	7.0752	0.0026	6.8090	

Table 5. Statistics of the competitive models fitted to the fatigue data.

Model	-LL	AIC	CAIC	BIC	HQIC
MONHTLEx	**447.8021**	**885.6043**	**894.966**	**872.5784**	**880.3325**
MOTLWe	452.2929	912.5858	913.0069	923.0065	916.8032
MOTLBXII	533.9673	1075.935	1076.356	1086.355	1080.152
MOTLLo	506.3483	1020.697	1021.118	1031.117	1024.914

with a minimum of five parameters, offering greater flexibility, even when adjusting parameter settings for different applications. We discussed three prominent members of this family and observed that the *pdf* and *hrf* exhibit diverse trends, making the new family suitable for various real-world applications. Some statistical and mathematical properties of this family were derived, along with a characterization derived from truncated moments. To estimate the parameters, we employed the MLE technique. A simulation study was conducted, considering the exponential distribution as a special case, where we also applied the MLE method. Furthermore, we analyzed a real-world dataset under the MONHTLEx model. Among the competing models considered, our model outperformed others based on various comparison criteria and statistical tests, demonstrating its effectiveness and applicability.

Acknowledgement(s)

We sincerely appreciate the valuable feedback and insightful suggestions from the referees and the editor, which have significantly enhanced the quality of this paper.

References

[1] Afify, A. Z., Cordeiro, G. M., Ibrahim, N. A., Jamal, F., Elgarhy, M., and Nasir, M. A. (2020). The marshall-olkin odd burr iii-g family: theory, estimation, and engineering applications. *IEEE Access*, 9:4376–4387.

[2] Alzaatreh, A., Lee, C., and Famoye, F. (2013). A new method for generating families of continuous distributions. *Metron*, 71(1):63–79.

[3] Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian journal of statistics*, pages 171–178.

[4] Birnbaum, Z. W. and Saunders, S. C. (1969). Estimation for a family of life distributions with applications to fatigue. *Journal of applied probability*, 6(2):328–347.

[5] Chakraborty, S. and Handique, L. (2017). The generalized marshall-olkin-kumaraswamy-g family of distributions. *Journal of data Science*, 15(3):391–422.

[6] Eugene, N., Lee, C., and Famoye, F. (2002). Beta-normal distribution and its applications. *Communications in Statistics-Theory and methods*, 31(4):497–512.

[7] Galambos, J. and Kotz, S. (2006). *Characterizations of Probability Distributions.: A Unified Approach with an Emphasis on Exponential and Related Models.*, volume 675. Springer.

[8] Glänzel, W. (1987). A characterization theorem based on truncated moments and its application to some distribution families. In *Mathematical statistics and probability theory: volume B statistical inference and methods proceedings of the 6th Pannonian symposium on mathematical statistics, bad Tatzmannsdorf, Austria, September 14–20, 1986*, pages 75–84. Springer.

[9] Glanzel, W. (1990). Some consequences of a characterization theorem based on truncated moments. *Statistics*, 21(4):613–618.

[10] Glänzel, W. and Hamedani, G. (2001). Characterizations of nivariate continuous distributions. *Studia Scientiarum Mathematicarum Hungarica*, 37(1-2):83–118.

[11] Handique, L., Chakraborty, S., and de Andrade, T. A. (2019). The exponentiated generalized marshall–olkin family of distribution: its properties and applications. *Annals of Data Science*, 6:391–411.

[12] Jamal, F., Reyad, H., Chesneau, C., Nasir, M. A., and Othman, S. (2020). The marshall-olkin odd lindley-g family of distributions: Theory and applications. *Punjab University Journal of Mathematics*, 51(7).

[13] Khaleel, M. A., Oguntunde, P. E., Al Abbasi, J. N., Ibrahim, N. A., and AbuJarad, M. H. (2020). The marshall-olkin topp leone-g family of distributions: A family for generalizing probability models. *Scientific African*, 8:e00470.

[14] Kim, J. H. and Jeon, Y. (2013). Credibility theory based on trimming. *Insurance: Mathematics and Economics*, 53(1):36–47.

[15] Klakattawi, H., Alsulami, D., Elaal, M. A., Dey, S., and Baharith, L. (2023). Correction: A new generalized family of distributions based on combining marshal-olkin transformation with tx family. *Plos one*, 18(10):e0293100.

[16] Kotz, S. and Shanbhag, D. (1980). Some new approaches to probability distributions. *Advances in Applied Probability*, 12(4):903–921.

[17] Liu, B. and Ananda, M. M. (2022). A generalized family of exponentiated composite distributions. *Mathematics*, 10(11):1895.

[18] Marshall, A. W. and Olkin, I. (1997). A new method for adding a parameter to a family of distributions with application to the exponential and weibull families. *Biometrika*, 84(3):641–652.

[19] Nadarajah, S. and Haghighi, F. (2011). An extension of the exponential distribution. *Statistics*, 45(6):543–558.

[20] Nasiru, S. and Abubakari, A. G. (2022). Marshall-olkin zubair-g family of distributions. *Pakistan Journal of Statistics and Operation Research*, pages 195–210.

[21] Nassar, M., Kumar, D., Dey, S., Cordeiro, G. M., and Afify, A. Z. (2019). The marshall–olkin alpha power family of distributions with applications. *Journal of Computational and Applied Mathematics*, 351:41–53.

[22] Reyad, H., Selim, M. A., and Othman, S. (2019). The nadarajah haghighi topp leone-g family of distributions with mathematical properties and applications. *Pakistan Journal of Statistics and Operation Research*, pages 849–866.

[23] Santos-Neto, M., Bourguignon, M., Zea, L. M., Nascimento, A. D., and Cordeiro, G. M. (2014). The marshall-olkin extended weibull family of distributions. *Journal of Statistical*

Distributions and Applications, 1(1):1–24.

[24] Shanker, R. and Shukla, K. (2016). On modeling of lifetime data using three-parameter generalized lindley and generalized gamma distributions. *Biometrics & biostatistics International journal*, 4(7):1–7.

[25] Telcs, A., Schubert, A., et al. (1984). Characterization by truncated moments and its application to pearson-type distributions. *Open Access publications from Katholieke Universiteit Leuven.*

[26] Vicari, S. K.-D. and Kotz, S. (2005). Survey of developments in the theory of continuous skewed distributions. *Metron*, 63(2):225–261.

10. Appendices

10.1. Appendix A

Theorem 10.1. *Let (Ω, F, P) be a given probability space and let $\mathbf{H} = [a, b]$ be an interval for $a < b$ ($d = -\infty, e = \infty$ might as well be allowed). Let $Y : \Omega -> \mathbf{H}$ be a continuous r.v. with the cdf F_p and let Q_1 and Q_2 be two real functions defined on H such that,*

$$E[Q_1(Y)|Y \geq y] = E[Q_2(Y)|Y \geq y]\zeta(y), \quad y \in \mathbf{H},$$

is defined with some real function ζ. Assume that $Q_1(y), Q_2(y) \in \mathbf{C}^1(\mathbf{H})$ and F is twice strictly monotone function and continuously differentiable on the set $\mathbf{H}$. Finally, assume that the equation $Q_2\zeta = g$ has no real solution in the interior of $\mathbf{H}$. Then F can be uniquely determined by the functions Q_1, Q_2, and ζ, particularly

$$F(y) = \int_a^y C|\frac{\zeta'(u)}{\zeta(u)Q_2(u) - Q_1(u)}|e^{-s(u)}du,$$

where the function s is a solution of the differential equation $s' = \frac{\zeta' Q_2}{\zeta Q_1 - Q_2}$ and C is the normalization constant, such that $\int_{\mathbf{H}} dF = 1$.