Earthline Journal of Mathematical Sciences
ISSN (Online): 2581-8147
Volume 3, Number 2, 2020, Pages 167-190
https://doi.org/10.34198/ejms.3220.167190

The $T - R\{Y\{U\}\}$ Family of Distributions of Type I: Some Properties and Applications

Clement Boateng Ampadu

31 Carrolton Road, Boston, MA 02132-6303, USA
e-mail: drampadu@hotmail.com

Abstract

CDF-quantile distributions appeared in [1]. In the present paper, we show it can be used to generalize the $T - R\{Y\}$ class of distributions [2] to a new family which we call $T - R\{Y\{U\}\}$ family of distributions. Some properties and applications associated with the $T - R\{Y\{U\}\}$ family of distributions are obtained.

1. The $T - R\{Y\}$ Family of Distributions

This family of distributions was proposed in [2]. In particular, let T, R, Y be random variables with CDF's $F_T(x) = P(T \leq x)$, $F_R(x) = P(R \leq x)$, and $F_Y(x) = P(Y \leq x)$, respectively. Let the corresponding quantile functions be denoted by $Q_T(p)$, $Q_R(p)$, and $Q_Y(p)$, respectively. Also if the densities exist, let the corresponding PDF's be denoted by $f_T(x)$, $f_R(x)$, and $f_Y(x)$, respectively. Following this notation, the CDF of the $T - R\{Y\}$ is given by

$$F_X(x) = \int_a^{Q_Y(F_R(x))} f_T(t)dt = F_T\{Q_Y(F_R(x))\}$$

Received: October 17, 2019; Accepted: January 27, 2020

2010 Mathematics Subject Classification: 62Exx.

Keywords and phrases: CDF-quantile distributions, $T - R\{Y\}$ distributions.

and the PDF of the $T - R\{Y\}$ family is given by

$$f_X(x) = \frac{f_R(x)}{f_Y\{Q_Y(F_R(x))\}} f_T\{Q_Y(F_R(x))\}.$$

2. The CDF-quantile Family of Distributions

Let $G(x, \mu, \sigma)$ denote a CDF for random variable X with support $(0, 1)$, where $\mu \in \mathbb{R}$ is a location parameter, and $\sigma > 0$ is a scale parameter. Then,

$$G(x, \mu, \sigma) = F[U(H^{-1}(x), \mu, \sigma)],$$

where F is a CDF with support which is denoted D_1, H is an invertible CDF with support which is denoted by D_2, and $U : D_2 \mapsto D_1$ is an appropriate transform for imposing the location and scale parameters, μ and σ.

Case I: When F has support $(-\infty, \infty)$, and H is an invertible CDF with support $(-\infty, \infty)$, we take $U : (-\infty, \infty) \mapsto (-\infty, \infty)$ as

$$U(y, \mu, \sigma) = \frac{y - \mu}{\sigma}.$$

Case II: When F has support $(-\infty, \infty)$, and H is an invertible CDF with support $(0, \infty)$, we take $U : (0, \infty) \mapsto (-\infty, \infty)$ as

$$U(y, \mu, \sigma) = \frac{\log(y) - \mu}{\sigma}.$$

Case III: When F has support $(0, \infty)$, and H is an invertible CDF with support $(-\infty, \infty)$, we take $U : (-\infty, \infty) \mapsto (0, \infty)$ as

$$U(y, \mu, \sigma) = \exp\left(\frac{-\mu}{\sigma}\right)\exp\left(\frac{y}{\sigma}\right).$$

Case IV: When F has support $(0, \infty)$, and H is an invertible CDF with support $(0, \infty)$, we take $U : (0, \infty) \mapsto (0, \infty)$ as

$$U(y, \mu, \sigma) = \exp\left(\frac{-\mu}{\sigma}\right) y^{\frac{1}{\sigma}}.$$

3. The New Class of Distributions

To motivate the new class of distributions, we first make the following observation for Case I of the previous section. Note that similar observations hold for the remaining cases. For Case I, the CDF can be written as

$$G_1(x, \mu, \sigma) = F\left[\frac{H^{-1}(x) - \mu}{\sigma}\right].$$

If $F' = f$, then G_1 has the following integral representation

$$G_1(x, \mu, \sigma) = \int_{-\infty}^{\frac{H^{-1}(x)-\mu}{\sigma}} f(t)dt.$$

Further observe that

$$G_1(x, 0, 1) = \int_{-\infty}^{H^{-1}(x)} f(t)dt$$

and

$$G_1(S(x), 0, 1) = \int_{-\infty}^{H^{-1}(S(x))} f(t)dt.$$

If the random variable T with support $(-\infty, \infty)$ has PDF $f(t)$ and CDF $F(t)$, the random variable R has CDF $S(x)$, and the random variable Y with support $(-\infty, \infty)$ has quantile function, $H^{-1}(x)$, then the relation to the $T - R\{Y\}$ family of distributions [2] is clear. If we define $U_1 : \text{supp}(T) \mapsto \text{supp}(Y)$ by $U_1(y, \mu, \sigma) = \frac{y - \mu}{\sigma}$, then it follows that

$$G_1(S(x), 0, 1) = \int_{-\infty}^{U_1(H^{-1}(S(x)), 0, 1)} f(t)dt,$$

that is,

$$G_1(S(x), 0, 1) = F[U_1(H^{-1}(S(x)), 0, 1)].$$

Now we present the $T - R\{Y\{U\}\}$ class of distributions of type I as follows

Definition 3.1. Let the random variable T with support $(-\infty, \infty)$ have PDF $f_T(t)$ and CDF $F_T(t)$, the random variable R have CDF $F_R(x)$, and the random variable Y with support $(-\infty, \infty)$ have quantile function Q_Y, and let $U_1 : \text{supp}(T) \mapsto \text{supp}(Y)$ be defined as $U_1(y, \mu, \sigma) = \frac{y-\mu}{\sigma}$, where $\mu \in \mathbb{R}$, and $\sigma > 0$, then we say a random variable X is $T - R\{Y\{U_1\}\}$ distributed or $T - R\{Y\{U\}\}$ distributed of TYPE I if the CDF is given by the following

$$G_1(x, \mu, \sigma) = \int_{-\infty}^{U_1(Q_Y(F_R(x)))} f_T(t)dt = F_T[U_1(Q_Y(F_R(x)))] = F_T\left[\frac{Q_Y(F_R(x)) - \mu}{\sigma}\right].$$

By differentiating the CDF in the previous definition, we have the following

Theorem 3.2. *The PDF of the* $T - R\{Y\{U_1\}\}$ *class of distributions is given by*

$$g_1(x, \mu, \sigma) = \frac{f_T\left[\frac{Q_Y(F_R(x)) - \mu}{\sigma}\right] f_R(x)}{\sigma f_Y(Q_Y(F_R(x)))},$$

where the random variable T with support $(-\infty, \infty)$ *has PDF* f_T, *the random variable R has CDF* $F_R(x)$ *and PDF* $f_R(x)$, *the random variable Y with support* $(-\infty, \infty)$ *has quantile function* Q_Y, $\sigma > 0$, *and* $\mu \in \mathbb{R}$.

4. Some Statistical Measures

Theorem 4.1. (Transformation) *If W is uniform on* $(0, 1)$, *then the random variable*

$$X = Q_R\{e^{-e^{\sigma \ln\left(\frac{1-W}{W}\right) - \mu}}\}$$

follows the standard Logistic-R$\{$*standard Gumbel*$\{U_1\}\}$ *class of distributions, where* Q_R *is the quantile of the random variable* R, $\sigma > 0$, *and* $\mu \in \mathbb{R}$.

Proof. Assume Y is standard Gumbel with quantile function, $Q_Y(p) = -\ln(-\ln(p))$ for $0 < p < 1$, and T is standard Logistic with CDF $F_T(t) = (1 + e^{-x})^{-1}$, $x \in \mathbb{R}$, and $F_R(x)$ is the CDF of the random variable R. We know the CDF of W is

$F_W(w) = P(W \le w) = w.$ Now we show the CDF of X is given by the standard Logistic-$R\{\text{standard Gumbel}\{U_1\}\}$ class of distributions as follows

$$P(X \le x) = P(Q_R\{e^{-e^{\sigma \ln\left(\frac{1-W}{W}\right)-\mu}}\} \le x).$$

However, $Q_R = F_R^{-1}$, where F_R is the CDF of the random variable R. Thus, the above implies the following

$$\begin{aligned}
P(X \le x) &= P(e^{-e^{\sigma \ln\left(\frac{1-W}{W}\right)-\mu}} \le F_R(x)) \\
&= P(-e^{\sigma \ln\left(\frac{1-W}{W}\right)-\mu} \le \ln(F_R(x))) \\
&= P(e^{\sigma \ln\left(\frac{1-W}{W}\right)-\mu} \ge -\ln(F_R(x))) \\
&= P\left(\sigma \ln\left(\frac{1-W}{W}\right) - \mu \ge \ln(-\ln(F_R(x)))\right) \\
&= P\left(\ln\left(\frac{1-W}{W}\right) \ge \frac{\ln(-\ln(F_R(x))) + \mu}{\sigma}\right) \\
&= P\left(\frac{1-W}{W} \ge e^{\frac{\ln(-\ln(F_R(x)))+\mu}{\sigma}}\right) \\
&= P(1 \ge W + We^{\frac{\ln(-\ln(F_R(x)))+\mu}{\sigma}}) \\
&= P(1 \ge W(1 + e^{\frac{\ln(-\ln(F_R(x)))+\mu}{\sigma}})) \\
&= P\left(W \le \frac{1}{(1 + e^{\frac{\ln(-\ln(F_R(x)))+\mu}{\sigma}})}\right) \\
&= (1 + e^{\frac{\ln(-\ln(F_R(x)))+\mu}{\sigma}})^{-1}.
\end{aligned}$$

It now follows that the CDF of X is the standard Logistic-$R\{$standard Gumbel$\{U_1\}\}$ class of distributions, that is,

$$F_X(x) = P(X \le x) = (1 + e^{\frac{\ln(-\ln(F_R(x)))+\mu}{\sigma}})^{-1}.$$

Theorem 4.2. (Quantile) *The quantile of the $T - R\{Y\{U_1\}\}$ class of distributions or the $T - R\{Y\{U\}\}$ class of distributions of TYPE I is given by*

$$Q(p) = Q_R\{F_Y[\sigma Q_T(p) + \mu]\},$$

where $0 < p < 1$, $Q_R = F_R^{-1}$ *is the quantile of the random variable R with CDF* F_R, F_Y *is the CDF of the random variable Y,* $Q_T = F_T^{-1}$ *is the quantile of the random variable T with CDF* F_T, $\sigma > 0$, *and* $\mu \in \mathbb{R}$.

Proof. Using the fact that $F_T^{-1} = Q_T$, $Q_Y = F_Y^{-1}$, and $Q_R = F_R^{-1}$. The result follows from Definition 3.1 by solving the following equation for $Q(p)$

$$p = F_T\left[\frac{Q_Y(F_R(Q(p))) - \mu}{\sigma}\right].$$

Theorem 4.3. (CDF Power Series) *The standard Logistic-standard Gumbel$\{$standard Gumbel$\{U_1\}\}$ class of distributions has the following representation as a power series for its CDF*

$$F_Y(y) = \sum_{k,q=0}^{\infty} \sum_{m=0}^{q} \frac{(-1)^{k+m+q}(1+k)^q \mu^{q-m}}{\sigma^q q!} \binom{q}{m} y^m,$$

where $\sigma > 0$, μ, $y \in \mathbb{R}$.

Proof. From Theorem 4.1, we know that the CDF of the standard Logistic-$R\{$standard Gumbel$\{U_1\}\}$ class of distributions, is given by

$$F_X(x) = P(X \le x) = (1 + e^{\frac{\ln(-\ln(F_R(x)))+\mu}{\sigma}})^{-1}.$$

Now if R is standard Gumbel, then it follows that

$$F_R(x) = e^{-e^{-x}}$$

thus, the CDF of the standard Logistic-standard Gumbel{standard Gumbel$\{U_1\}$} class of distributions, call it, $F_Y(y)$, is given by

$$F_Y(y) = (1 + e^{\frac{-y+\mu}{\sigma}})^{-1}.$$

By the negative binomial series, we can write

$$(1 + e^{\frac{-y+\mu}{\sigma}})^{-1} = \sum_{k=0}^{\infty} (-1)^k e^{\frac{(-y+\mu)(-1-k)}{\sigma}}.$$

By the power series representation for the exponential function, we can write

$$e^{\frac{(-y+\mu)(-1-k)}{\sigma}} = \sum_{q=0}^{\infty} \frac{(-y+\mu)^q (-1-k)^q}{\sigma^q q!}.$$

By the binomial theorem, we can write

$$(-y+\mu)^q = \sum_{m=0}^{q} \binom{q}{m} (-1)^m y^m \mu^{q-m}.$$

It now follows that

$$F_Y(y) = \sum_{k,q=0}^{\infty} \sum_{m=0}^{q} \frac{(-1)^{k+m+q} (1+k)^q \mu^{q-m}}{\sigma^q q!} \binom{q}{m} y^m.$$

Theorem 4.4. (PDF Power Series) *The standard Logistic-standard Gumbel{standard Gumbel$\{U_1\}$} class of distributions has the following representation as a power series for its PDF*

$$f_Y(y) = \sum_{k,q=0}^{\infty} \sum_{m=0}^{q} \frac{(-1)^{k+q+m} (1+k)^q \mu^{q-m}}{\sigma^{q+1} q!} \binom{1+k}{k} \binom{q}{m} y^m,$$

where $\sigma > 0$, μ, $y \in \mathbb{R}$.

Proof. By differentiating $F_Y(y)$ from the proof of the previous theorem, we know the PDF of the standard Logistic-standard Gumbel{standard Gumbel$\{U_1\}$} class of distributions is given by

$$f_Y(y) = \frac{e^{\frac{-y+\mu}{\sigma}}}{\sigma\left(1+e^{\frac{-y+\mu}{\sigma}}\right)^2}.$$

By the negative binomial series, we can write

$$\left(1+e^{\frac{-y+\mu}{\sigma}}\right)^{-2} = \sum_{k=0}^{\infty}(-1)^k\binom{1+k}{k}e^{\frac{(-y+\mu)(-2-k)}{\sigma}}.$$

It now follows that

$$e^{\frac{-y+\mu}{\sigma}}\left(1+e^{\frac{-y+\mu}{\sigma}}\right)^{-2} = \sum_{k=0}^{\infty}(-1)^k\binom{1+k}{k}e^{\frac{(-y+\mu)(-1-k)}{\sigma}}.$$

By the power series representation for the exponential function, we can write

$$e^{\frac{(-y+\mu)(-1-k)}{\sigma}} = \sum_{q=0}^{\infty}\frac{(-y+\mu)^q(-1)^q(1+k)^q}{\sigma^q q!}.$$

By the binomial theorem, we can write

$$(-y+\mu)^q = \sum_{m=0}^{q}\binom{q}{m}(-1)^m y^m \mu^{q-m}.$$

It now follows that

$$e^{\frac{-y+\mu}{\sigma}}\left(1+e^{\frac{-y+\mu}{\sigma}}\right)^{-2} = \sum_{k,q=0}^{\infty}\sum_{m=0}^{q}\frac{(-1)^{k+q+m}(1+k)^q\mu^{q-m}}{\sigma^q q!}\binom{1+k}{k}\binom{q}{m}y^m.$$

So the result follows from

$$\frac{1}{\sigma}e^{\frac{-y+\mu}{\sigma}}\left(1+e^{\frac{-y+\mu}{\sigma}}\right)^{-2} = \sum_{k,q=0}^{\infty}\sum_{m=0}^{q}\frac{(-1)^{k+q+m}(1+k)^q\mu^{q-m}}{\sigma^{q+1} q!}\binom{1+k}{k}\binom{q}{m}y^m.$$

Theorem 4.5. (Non-Central Moments) *The rth noncentral moments of the standard Logistic-R$\{$standard Gumbel$\{U_1\}\}$ class of distributions are given by*

$$\mu_r' = \sum_{i,q=0}^{\infty}\sum_{m=0}^{q}\sum_{v=0}^{m\sigma}\frac{\delta_{r,i}(-1)^{q+m\sigma-v}i^q\mu^{q-m}}{q!(1-v)}\binom{q}{m}\binom{m\sigma}{v},$$

where $\mu \in \mathbb{R}$, $\sigma > 0$, *and* $\delta_{r,i} = (ih_0)^{-1}\sum_{s=1}^{i}[s(r+1)-i]h_s\delta_{r,i-s}$ *with* $\delta_{r,0} = h_0^r$ *for* $i = 1, 2, \ldots$ [3].

Proof. From Theorem 4.1, the following random variable below follows the standard Logistic-$R\{$standard Gumbel$\{U_1\}\}$ class of distributions

$$Q_R\{e^{-e^{\sigma\ln\left(\frac{1-W}{W}\right)-\mu}}\},$$

where $Q_R(\cdot) = F_R^{-1}(\cdot)$ is a quantile function. According to [4], we can write

$$Q_R(u) = \sum_{i=0}^{\infty}h_iu^i,$$

where the coefficients are suitably chosen real numbers that depend on the parameters of the $F_R(x)$ distribution. For a power series raised to a positive integer $r \geq 1$, we have

$$(Q_R(u))^r = \left(\sum_{i=0}^{\infty}h_iu_i\right)^r = \sum_{i=0}^{\infty}\delta_{r,i}u^i,$$

where $\delta_{r,i}$ are obtained from the recurrence equation as stated in the theorem. Thus we have the following

$$\mu_r' = \sum_{i=0}^{\infty}\delta_{r,i}E[(e^{-e^{\sigma\ln\left(\frac{1-W}{W}\right)-\mu}})^i],$$

where $E(\cdot)$ is an expectation. Now observe we can write $(e^{-e^{\sigma\ln\left(\frac{1-W}{W}\right)-\mu}})^i$ as follows

$$(e^{-e^{\sigma\ln\left(\frac{1-W}{W}\right)-\mu}})^i = (e^{-\left(\frac{1-W}{W}\right)^{\sigma}-\mu})^i$$

$$= \left(e^{-\left\{\left(\frac{1-W}{W}\right)^{\sigma}+\mu\right\}}\right)^{i}$$

$$= e^{-i\left\{\left(\frac{1-W}{W}\right)^{\sigma}+\mu\right\}}.$$

By the power series representation for the exponential function, we can write

$$e^{-i\left\{\left(\frac{1-W}{W}\right)^{\sigma}+\mu\right\}} = \sum_{q=0}^{\infty} \frac{(-i)^{q}\left\{\left(\frac{1-W}{W}\right)^{\sigma}+\mu\right\}^{q}}{q!}.$$

By the Binomial theorem we can write

$$\left\{\left(\frac{1-W}{W}\right)^{\sigma}+\mu\right\}^{q} = \sum_{m=0}^{q}\binom{q}{m}\left(\frac{1-W}{W}\right)^{m\sigma}\mu^{q-m}.$$

Again by the Binomial theorem we have

$$\left(\frac{1-W}{W}\right)^{m\sigma} = W^{-m\sigma}(1-W)^{m\sigma}$$

$$= W^{-m\sigma} = \sum_{v=0}^{m\sigma}\binom{m\sigma}{v}(1)^{v}(-W)^{m\sigma-v}$$

$$= W^{-m\sigma}\sum_{v=0}^{m\sigma}\binom{m\sigma}{v}(-1)^{m\sigma-v}(W)^{m\sigma-v}$$

$$= \sum_{v=0}^{m\sigma}\binom{m\sigma}{v}(-1)^{m\sigma-v}W^{-v}.$$

It now follows that we have the following

$$\left(e^{-e^{\sigma\ln\left(\frac{1-W}{W}\right)-\mu}}\right)^{i} = \sum_{q=0}^{\infty}\sum_{m=0}^{q}\sum_{v=0}^{m\sigma}\frac{(-1)^{q+m\sigma-v}\, i^{q}\mu^{q-m}}{q!}\binom{q}{m}\binom{m\sigma}{v}W^{-v}.$$

Now using the expression immediately above in

$$\mu_r' = \sum_{i=0}^{\infty}\delta_{r,i}E\left[\left(e^{-e^{\sigma\ln\left(\frac{1-W}{W}\right)-\mu}}\right)^{i}\right]$$

we deduce the following

$$\mu_r' = \sum_{i, q=0}^{\infty} \sum_{m=0}^{q} \sum_{v=0}^{m\sigma} \frac{\delta_{r,i}(-1)^{q+m\sigma-v} i^q \mu^{q-m}}{q!} \binom{q}{m} \binom{m\sigma}{v} E[W^{-v}].$$

From Theorem 4.1, we know W is uniform on (0, 1). Let $Y = W^{-v}$. By the transformation technique, the CDF of Y for $0 \le y \le 1$ is given by

$$F_Y(y) = y^{\frac{-1}{v}}.$$

Consequently, the PDF is given by

$$f_Y(y) = \frac{-1}{v} y^{-\frac{1+v}{v}},$$

where $0 \le y \le 1$. Thus,

$$E[Y] = \int_0^1 y f_Y(y) dy = \frac{1}{1-v}.$$

It now follows that

$$\mu_r' = \sum_{i, q=0}^{\infty} \sum_{m=0}^{q} \sum_{v=0}^{m\sigma} \frac{\delta_{r,i}(-1)^{q+m\sigma-v} i^q \mu^{q-m}}{q!(1-v)} \binom{q}{m} \binom{m\sigma}{v}.$$

Given a random variable X, one defines the moment generating function as

$$M_X(z) = E[e^{zX}],$$

where $E[\cdot]$ is an expectation. Now using the series expansion for e^{zX}, one can write

$$M_X(z) = \sum_{r=0}^{\infty} \frac{z^r \mu_r'}{r!},$$

where μ_r' is the rth non-central moment of the random variable X. Thus from the previous theorem, the following is immediate

Theorem 4.6. (Moment Generating Function) *The moment generating function of the standard Logistic-$R\{$standard Gumbel$\{U_1\}\}$ class of distributions are given by*

$$\sum_{r,i,q=0}^{\infty}\sum_{m=0}^{q}\sum_{v=0}^{m\sigma}\frac{z^{r}\delta_{r,i}(-1)^{q+m\sigma-v}i^{q}\mu^{q-m}}{r!\,q!\,(1-v)}\binom{q}{m}\binom{m\sigma}{v},$$

where $\mu \in \mathbb{R}$, $\sigma > 0$, *and* $\delta_{r,i} = (ih_0)^{-1}\sum_{s=1}^{i}[s(r+1)-i]h_s\delta_{r,i-s}$ *with* $\delta_{r,0} = h_0^r$ *for* $i = 1, 2, \ldots$ [3].

Theorem 4.7. (Shannon Entropy) *If a random variable V follows the* $T - R\{Y\{U_1\}\}$ *class of distributions, then the Shannon entropy of V, call it* S_V, *is given by*

$$S_V = \eta_T - E[\log f_R(Q_R\{F_Y[\sigma T + \mu]\})] + \log\sigma + E[\log f_Y(\sigma T + \mu)],$$

where the random variable T has Shannon entropy η_T, *the random variable R has PDF* f_R *and quantile function* Q_R, *the random variable Y has CDF* F_Y *and PDF* f_Y, $\sigma > 0$, *and* $\mu \in \mathbb{R}$.

Proof. From Theorem 3.2, $T = \frac{Q_Y(F_R(X)) - \mu}{\sigma}$ has PDF $r(t)$, thus the result follows by noting that we have the following

$$-E\left\{\log f_T\left[\frac{Q_Y(F_R(X)) - \mu}{\sigma}\right]\right\} = E[-\log f_T(t)] = \eta_T$$

$$E[\log f_R(X)] = E[\log f_R(Q_R\{F_Y[\sigma T + \mu]\})]$$

$$E[\log f_Y(Q_Y(F_R(X)))] = E[\log f_Y(\sigma T + \mu)]$$

$$E[\log\sigma] = \log\sigma.$$

5. Practical Illustration and Numerical Comparison

In this section, we show a member of the $T - R\{Y\{U\}\}$ family of distributions of type I is a good fit to the coupons data, Table 5 [5]. We also compare the new member arising from the $T - R\{Y\{U\}\}$ framework with a member of the $T - R\{Y\}$ framework. We assume the random variable T with support $(-\infty, \infty)$ is Normally distributed with CDF given by

$$F_T(x; c, d) = \frac{1}{2} erfc\left(\frac{c - x}{\sqrt{2}\, d}\right),$$

where $x, c \in \mathbb{R}$, with $d > 0$, and

$$erfc(z) = 1 - erf(z) = 1 - \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt.$$

We also assume that the random variable Y with support $(-\infty, \infty)$ is (standard) Cauchy distributed, so that the quantile function is given by

$$Q_Y(x) = \tan\left(\pi\left(x - \frac{1}{2}\right)\right),$$

where $0 < x < 1$. Finally, we have the added assumption that the random variable R is Pareto distributed with CDF

$$F_R(x) = 1 - \left(\frac{a}{x}\right)^b,$$

where $x \geq a$, and $a, b > 0$.

From the $T - R\{Y\{U\}\}$ framework we deduce the following

Proposition 5.1. *The CDF of the Normal-Standard Cauchy*$\{Pareto\{U\}\}$ *distribution of type I is given by*

$$G(x; a, b, c, d, \mu, \sigma) = \frac{1}{2} erfc\left(\frac{c - \frac{\tan\left(\pi\left(\frac{1}{2} - \left(\frac{a}{x}\right)^b\right)\right) - \mu}{\sigma}}{\sqrt{2}\, d}\right),$$

where $x \geq a$, $a, b, d, \sigma > 0$, $c, \mu \in \mathbb{R}$, $erfc(z) = 1 - erf(z) = 1 - \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt.$

Remark 5.2. We write $J \sim NSCPU(a, b, c, d, \mu, \sigma)$, if J is a random variable with the CDF given by the previous Proposition. When the parameters a, b, c, d are fixed we write $J \sim NSCPU_{(a_f ix, b_f ix, c_f ix, d_f ix)}(\mu, \sigma)$.

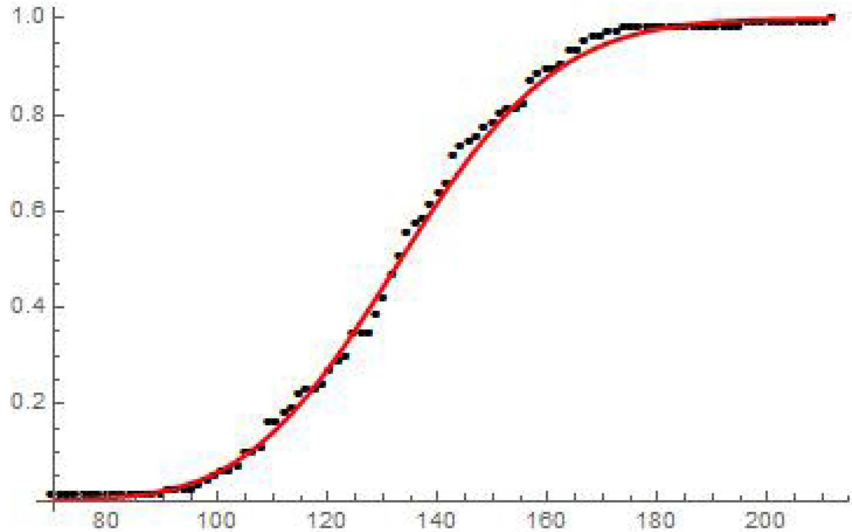

Figure 1. The CDF of $NSCPU_{(0.0820562, 0.703981, 57.9051, 6.79818)}(-0.00716473, 1.00037)$ fitted to the empirical distribution of the coupons data [5].

Remark 5.3. The PDF of the $NSCPU(a, b, c, d, \mu, \sigma)$ distribution can be obtained by differentiating the CDF.

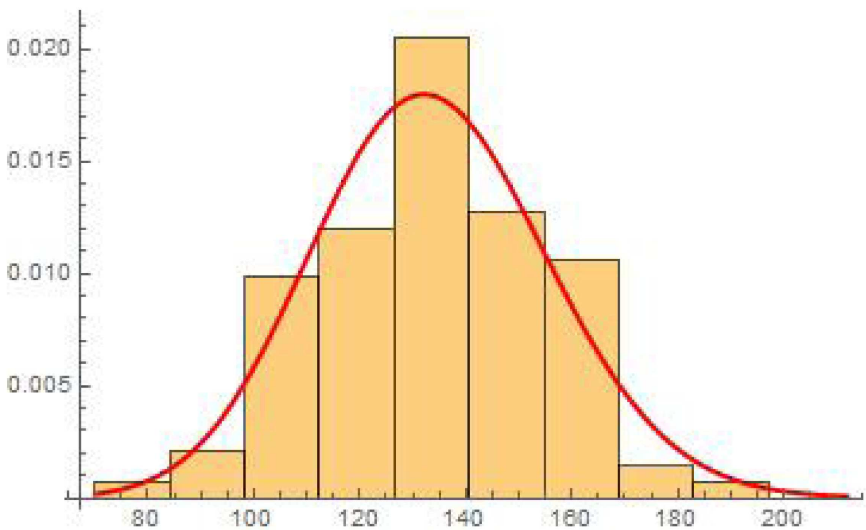

Figure 2. The PDF of $NSCPU_{(0.0820562, 0.703981, 57.9051, 6.79818)}(-0.00716473, 1.00037)$ fitted to the histogram of the coupons data [5].

From the $T - R\{Y\}$ framework we deduce the following

Proposition 5.4. *The CDF of the Normal-Standard Cauchy*$\{Pareto\}$ *distribution is given by*

$$F(x;\, a,\, b,\, c,\, d) = \frac{1}{2}\, erfc\left(\frac{c - \tan\left(\pi\left(\frac{1}{2} - \left(\frac{a}{x}\right)^{b}\right)\right)}{\sqrt{2}\, d}\right),$$

where $x \geq a,\ a,\ b,\ d > 0,\ c \in \mathbb{R},$ *and* $erfc(z) = 1 - erf(z) = 1 - \frac{2}{\sqrt{\pi}}\int_0^z e^{-t^2} dt.$

Remark 5.5. We write $Q \sim NSCP(a,\, b,\, c,\, d),$ if Q is a random variable with the CDF given by the previous proposition.

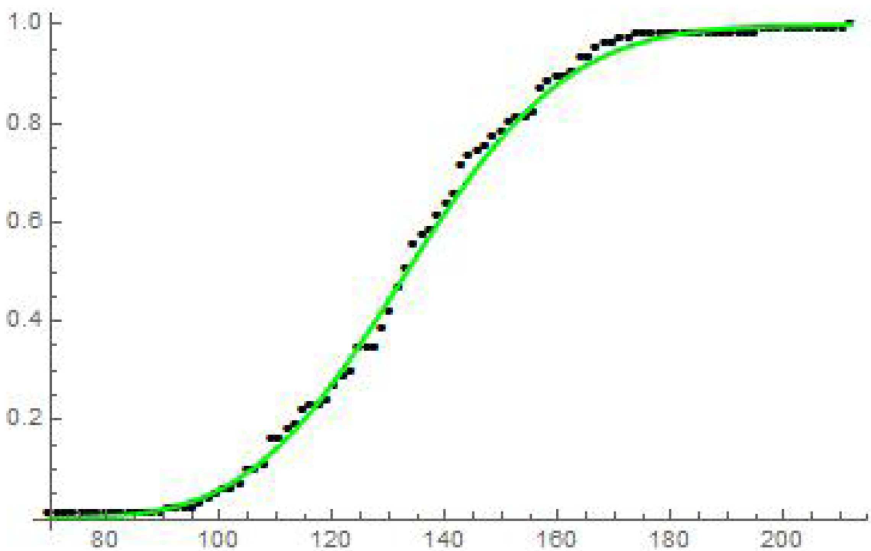

Figure 3. The CDF of $NSCP(0.0820562, 0.703981, 57.9051, 6.79818)$ fitted to the empirical distribution of the coupons data [5].

Remark 5.6. The PDF of the $NSCP(a,\, b,\, c,\, d)$ distribution can be obtained by differentiating the CDF.

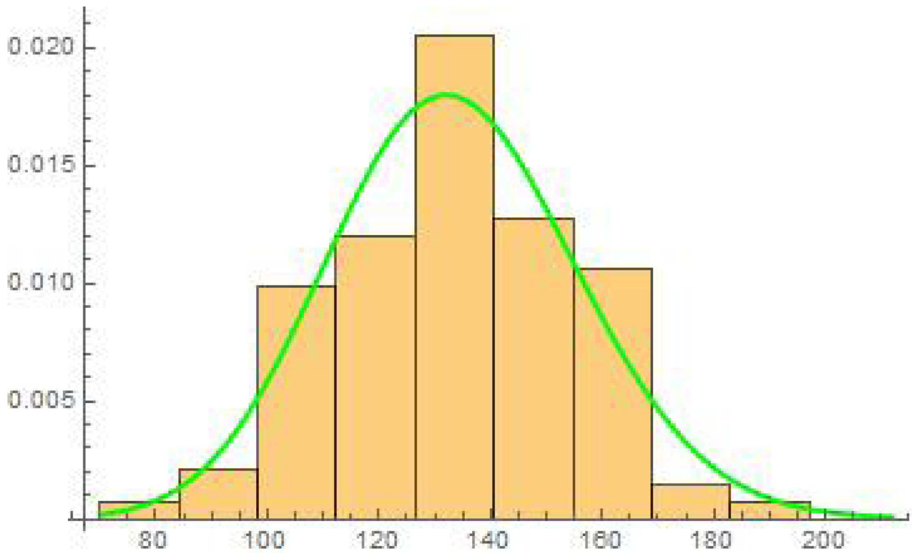

Figure 4. The PDF of $NSCP(0.0820562, 0.703981, 57.9051, 6.79818)$ fitted to the histogram of the coupons data [5].

In the rest of this section we compare the Normal-Standard Cauchy{Pareto} distribution, and the Normal-Standard Cauchy{Pareto{U}} distribution of type I in fitting the coupons data[5].

Table 1. Estimated parameters for the coupons data.

Model	Parameter Estimate	Standard Error
$NSCP(a,b,c,d)$	(0.0820562,0.703981,57.9051,6.79818)	(0.0816029, 0.142213, 40.9521, 5.91504)
$NSCPU_{(0.0820562,0.703981,57.9051,6.79818)}(\mu,\sigma)$	(-0.00716473,1.00037)	(4.13149, 0.0703856)

Table 2. Criteria for comparison.

Model	-2(Log-likelihood)	AIC	AICC	BIC
$NSCP(a,b,c,d)$	912.332	920.332	920.748	930.792
$NSCPU_{(0.0820562,0.703981,57.9051,6.79818)}(\mu,\sigma)$	912.331	916.331	916.454	921.561

In order to compare the two distribution models, we used the following criteria: -2(Loglikelihood), AIC (Akaike information criterion), AICC (corrected Akaike information criterion), and BIC (Bayesian information criterion) for the data set. The better distribution corresponds to the smaller -2(Log-likelihood) AIC, AICC, and BIC values:

$$AIC = 2k - 2l,$$

$$AICC = AIC + \frac{2k(k+1)}{n-k-1},$$

$$BIC = k\log(n) - 2l,$$

where k is the number of parameters in the statistical model, n is the sample size, and l is the maximized value of the log-likelihood function under the considered model. From Table 2, it is clear that the $NSCPU_{(0.0820562, 0.703981, 57.9051, 6.79818)}(\mu, \sigma)$ distribution has the smallest values across $\frac{3}{4}$ criteria considered, hence we see the

$$NSCPU_{(0.0820562, 0.703981, 57.9051, 6.79818)}(\mu, \sigma)$$

distribution is a better fit than the $NSCP(a, b, c, d)$ distribution to the coupons data.

6. Simulation Study

In this section a Monte Carlo simulation study is carried out to assess the performance of the maximum likelihood estimation method in the distribution

$$NSCPU_{(0.0820562, 0.703981, 57.9051, 6.79818)}(\mu, \sigma)$$

which is a member of the newly introduced $T - R\{Y\{U\}\}$ framework. Samples of sizes 200, 350, 500, and 700, are drawn from the

$$NSCPU_{(0.0820562, 0.703981, 57.9051, 6.79818)}(\mu, \sigma)$$

distribution, and the samples have been drawn for the following set of parameters

(a) Set I: $(\mu, \sigma) = (0.1, 0.9)$,

(b) Set II: $(\mu, \sigma) = (0.9, 0.9)$.

The maximum likelihood estimators for the parameters μ, and σ are obtained. The procedure has been repeated 200 times, and the standard deviation and variance for the estimates are computed, and the results are summarized in Table 3 and Table 5 for each of sets I and II, respectively, considered above.

Table 3. Result of simulation study for set I.

	Parameter μ	
Sample Size	Standard Deviation	Variance
200	1.089441	1.186881
350	0.888419	0.7892883
500	0.7158677	0.5124666
700	0.5997037	0.3596446
	Parameter σ	
Sample Size	Standard Deviation	Variance
200	0.01643016	0.0002699502
350	0.01326639	0.0001759971
500	0.01072819	0.0001150941
700	0.00908572	0.00008255031

From Table 3, we observe that the estimated standard deviation and variance consistently decrease with increasing sample size as seen in Table 4, hence the estimation method is adequate.

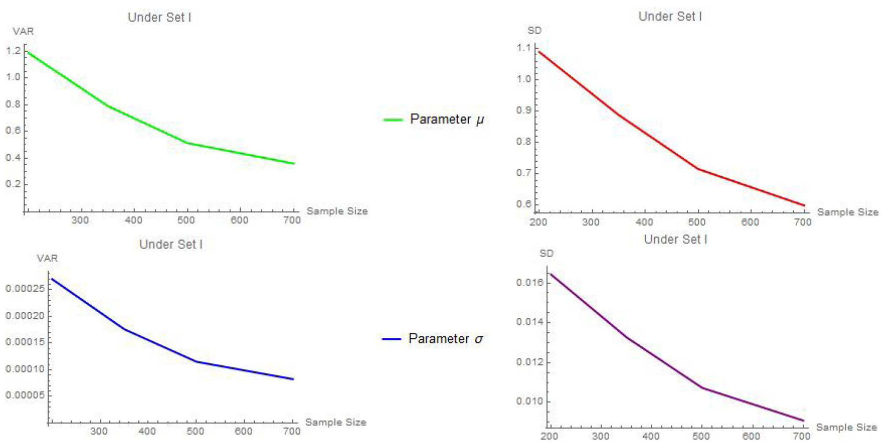

Table 4. Decreasing variance (VAR) and standard deviation (SD) for increasing sample size.

Table 5. Result of simulation study for set II.

	Parameter μ	
Sample Size	Standard Deviation	Variance
200	1.08944	1.186879
350	0.8884191	0.7892885
500	0.7158675	0.5124663
700	0.57897	0.3352062
	Parameter σ	
Sample Size	Standard Deviation	Variance
200	0.01643015	0.0002699498
350	0.0132664	0.0001759972
500	0.01072819	0.0001150941
700	0.008833336	0.00007802782

From Table 5, we observe that the standard deviation and variance consistently decrease with increasing sample size as seen in Table 6, hence the estimation method is adequate.

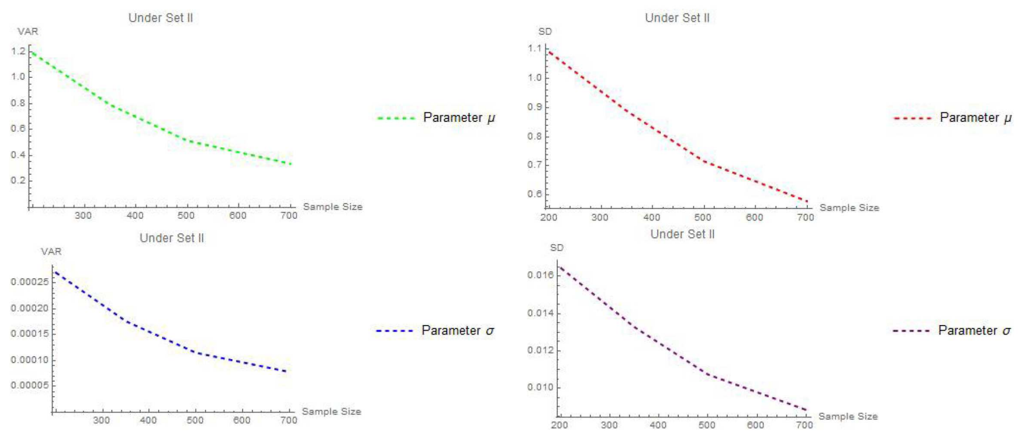

Table 6. Decreasing variance (VAR) and standard deviation (SD) for increasing sample size.

7. A Characterization Theorem

The characterization of statistical distributions plays a major role in stochastic modeling. In this section we present a characterization of the $T - R\{Y\{U\}\}$ distribution of type I. Our characterization theorem is based on a simple relationship between two truncated moments, and for related works in this direction, the reader is referred to [6]-[11].

At first, we recall the following which will be useful later

Theorem 7.1. [7] *Let* $(\Omega, \sum, \mathbb{P})$ *be a given probability space, and let* $I = [a, b]$ *be an interval for some* $a < b$ $(a = -\infty, b = \infty$ *might as well be allowed*). *Let* $X : \Omega \mapsto I$ *be a continuous random variable with probability distribution function F, and let* q_1 *and* q_2 *be two real functions on I such that*

$$\mathbb{E}[q_1(X) \mid X \geq x] = \mathbb{E}[q_2(X) \mid X \geq x]\eta(x), \quad x \in I,$$

is defined with some real function η. *Assume that* $q_1, q_2 \in C^1(I)$, *and* $\eta \in C^2(I)$, *and F is twice continuously differentiable and strictly monotone increasing on the set I. Finally, assume that the equation* $nq_2 = q_1$ *has no real solutions in the interior of I. Then F is uniquely determined by the functions* q_1, q_2, η. *In particular,*

$$F(x) = \int_a^x C \left| \frac{\eta'(u)}{\eta(u)q_2(u) - q_1(u)} \right| \exp(-s(u))du,$$

where the function s is a solution of the differential equation

$$s' = \frac{\eta' q_2}{\eta q_2 - q_1}$$

and C is a constant chosen to make $\int_I dF = 1$.

Remark 7.2. The characterization based on the ratio of two truncated moments is stable in the sense of weak convergence, and for more details see [12].

The main result of this section is as follows

Proposition 7.3. *Let* $X : \Omega \mapsto \mathbb{R}$ *be a continuous random variable, and let* $q_2(x) = 1$, *and*

$$q_1(x) = F_T\left[\frac{Q_Y(F_R(x)) - \mu}{\sigma}\right]$$

then the PDF of X is

$$\frac{f_T\left[\frac{Q_Y(F_R(x)) - \mu}{\sigma}\right] f_R(x)}{\sigma f_Y(Q_Y(F_R(x)))}$$

iff the function η *in Theorem* 7.1 *has the form*

$$\eta(x) = \frac{1}{2} F_T\left[\frac{Q_Y(F_R(x)) - \mu}{\sigma}\right],$$

where the random variable T with support $(-\infty, \infty)$ *has PDF* f_T *and CDF* F_T, *the random variable R has CDF* $F_R(x)$ *and PDF* $f_R(x)$, *the random variable Y with support* $(-\infty, \infty)$ *has quantile function* Q_Y *and PDF* f_Y, $\sigma > 0$, *and* $\mu \in \mathbb{R}$.

Proof. Let *X* have PDF

$$\frac{f_T\left[\frac{Q_Y(F_R(x)) - \mu}{\sigma}\right] f_R(x)}{\sigma f_Y(Q_Y(F_R(x)))},$$

then for all $x \in \mathbb{R}$ we deduce the following

$$(1 - F(x))\mathbb{E}[q_2(X) \mid X \geq x] = F_T\left[\frac{Q_Y(F_R(x)) - \mu}{\sigma}\right]$$

and

$$(1 - F(x))\mathbb{E}[q_1(X) \mid X \geq x] = \frac{1}{2}\left(F_T\left[\frac{Q_Y(F_R(x)) - \mu}{\sigma}\right]\right)^2$$

and finally

$$\begin{aligned} &\eta(x) q_2(x) - q_1(x) \\ &= -\frac{1}{2} F_T\left[\frac{Q_Y(F_R(x)) - \mu}{\sigma}\right] \\ &< 0. \end{aligned}$$

Conversely, if

$$\eta(x) = \frac{1}{2} F_T\left[\frac{Q_Y(F_R(x)) - \mu}{\sigma}\right],$$

then we can check that

$$s'(x) = -\frac{f_T\left[\frac{Q_Y(F_R(x)) - \mu}{\sigma}\right] f_R(x)}{\sigma f_Y(Q_Y(F_R(x))) F_T\left[\frac{Q_Y(F_R(x)) - \mu}{\sigma}\right]}$$

and hence

$$s(x) = -\log\left(F_T\left[\frac{Q_Y(F_R(x)) - \mu}{\sigma}\right]\right).$$

Now in view of Theorem 7.1, *X* has PDF

$$\frac{f_T\left[\frac{Q_Y(F_R(x)) - \mu}{\sigma}\right] f_R(x)}{\sigma f_Y(Q_Y(F_R(x)))}.$$

If q_2 is given by the previous proposition, then we have the following

Corollary 7.4. *Let* $X : \Omega \mapsto \mathbb{R}$ *be a continuous random variable, the random variable T with support* $(-\infty, \infty)$ *have PDF* f_T *and CDF* F_T, *the random variable R have CDF* $F_R(x)$ *and PDF* $f_R(x)$, *the random variable Y with support* $(-\infty, \infty)$ *have quantile function* Q_Y *and PDF* f_Y, $\sigma > 0$, *and* $\mu \in \mathbb{R}$. *The PDF of X is*

$$\frac{f_T\left[\frac{Q_Y(F_R(x)) - \mu}{\sigma}\right] f_R(x)}{\sigma f_Y(Q_Y(F_R(x)))}$$

$\Leftrightarrow$

there exists functions q_1 *and* η *defined in Theorem* 7.1 *satisfying the following differential equation*

$$\frac{\eta'(x) q_2(x)}{\eta(x) q_2(x) - q_1(x)} = -\frac{f_T\left[\frac{Q_Y(F_R(x)) - \mu}{\sigma}\right] f_R(x)}{\sigma f_Y(Q_Y(F_R(x))) F_T\left[\frac{Q_Y(F_R(x)) - \mu}{\sigma}\right]}.$$

Remark 7.5. The general solution of the differential equation in the above corollary is given by

$$\eta(x) = \frac{1}{F_T\left[\frac{Q_Y(F_R(x)) - \mu}{\sigma}\right]}\left[\int F_T\left[\frac{Q_Y(F_R(x)) - \mu}{\sigma}\right] \frac{f_T\left[\frac{Q_Y(F_R(x)) - \mu}{\sigma}\right] f_R(x)}{\sigma f_Y(Q_Y(F_R(x)))} dx + D\right]$$

for $x \in \mathbb{R}$, where *D* is a constant.

8. Concluding Remarks

The present paper has introduced a new family of distributions called $T - R\{Y\{U\}\}$ as a generalization of the $T - R\{Y\}$ family of distributions via the CDF-quantile distribution framework. Apart from applying the transformation technique to this new class of distributions, the quantile function, power series representation for the CDF and PDF, *r*th non-central moments, moment generating function, and the Shannon entropy are derived. A member of the $T - R\{Y\{U\}\}$ family of distributions of type I is shown to be practically superior to a member of the $T - R\{Y\}$ family of distributions in fitting the coupons data, showing the new family should be practical in fitting related data sets. A simulation study conducted shows the method of maximum likelihood is adequate in estimating parameters of members of this new class of distributions. Finally the new class of distributions is characterized in terms of a simple relationship between two truncated moments.

References

[1] Michael Smithson and Yiyun Shou, CDF-quantile distributions for modelling random variables on the unit interval, *British Journal of Mathematical and Statistical Psychology* 70(3) (2017), 412-438. https://doi.org/10.1111/bmsp.12091

[2] M. A. Aljarrah, C. Lee and F. Famoye, On generating T-X family of distributions using quantile functions, *J. Stat. Distrib. Appl.* 1(2) (2014). https://doi.org/10.1186/2195-5832-1-2

[3] I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products*, Academic Press, New York, 2007

[4] Suleman Nasiru, Peter N. Mwita and Oscar Ngesa, Exponentiated Generalized Transformed-Transformer Family of Distributions, *Journal of Statistical and Econometric Methods* 6(4) (2017), 1-17.

[5] Ayman Alzaatreh, Felix Famoye and Carl Lee, Gamma-Pareto Distribution and Its Applications, *Journal of Modern Applied Statistical Methods* 11(1) (2012). https://doi.org/10.22237/jmasm/1335845160

[6] W. Glänzel, A. Teles and A. Schubert, Characterization by truncated moments and its application to Pearson-type distributions, *Z. Wahrscheinlichkeitstheorie verw Gebiete* 66(2) (1984), 173-183. https://doi.org/10.1007/BF00531527

[7] W. Glänzel, A characterization theorem based on truncated moments and its application to some distribution families, in: *Mathematical Statistics and Probability Theory*, Springer, Netherlands, 1987, pp. 75-84. https://doi.org/10.1007/978-94-009-3965-3_8

[8] W. Glanzel, Some consequences of a characterization theorem based on truncated moments, *Statistics* 21(4) (1990), 613-618. https://doi.org/10.1080/02331889008802273

[9] W. Glänzel and G. G. Hamedani, Characterizations of univariate continuous distributions, *Studia Scientiarum Mathematicarum Hungarica* 37(1-2) (2001), 83-118. https://doi.org/10.1556/SScMath.37.2001.1-2.5

[10] G. G. Hamedani, Characterizations of univariate continuous distributions. II, *Studia Scientiarum Mathematicarum Hungarica* 39(3-4) (2002), 407-424. https://doi.org/10.1556/SScMath.39.2002.3-4.11

[11] G. G. Hamedani, Characterizations of univariate continuous distributions, III, *Studia Scientiarum Mathematicarum Hungarica* 43(3) (2006), 361-385. https://doi.org/10.1556/SScMath.43.2006.3.4

[12] Morad Alizadeh, Faton Merovci and G. G. Hamedani, Generalized transmuted family of distributions: properties and applications, *Hacettepe University Bulletin of Natural Sciences and Engineering Series B: Mathematics and Statistics* 46(4) (2016), 645-668. https://doi.org/10.15672/HJMS.201610915478