STATISTICA, anno LXXXII, n. 2, 2022

MCDONALD-G POISSON FAMILY OF DISTRIBUTIONS

Keshav Pokhrel
Department of Mathematics and Statistics, University of Michigan-Dearborn, USA
Gokarna Raj Aryal[1]
Department of Mathematics and Statistics, Purdue University Northwest, USA
Ram Chandra Kafle
Department of Mathematics and Statistics, Sam Houston State University, USA
Bhikhari Tharu
Department of Mathematics, Spelman College, USA
Netra Khanal
Department of Mathematics, University of Tampa, USA

SUMMARY

In this article, we utilize the method proposed by Tahir and Cordeiro (2016) to study a new family of distributions called the McDonald Generalized Poisson (McGP) family. This family is defined by using the genesis of the McDonald distribution and the zero truncated Poisson (ZTP) distribution. We provide some mathematical properties and parameter estimation procedures of the McGP family. Three real-life data are analyzed to illustrate the potential applications of the McGP family. Our examples illustrate that the development of new probability distributions is of great interest to capture the nature of the data under study. However, one can't guarantee a better fit just because a probability distribution possesses a larger number of parameters than its sub-model.

Keywords: McDonald distribution; McDonald-G family; Truncated Poisson distribution; Parameter estimation.

1. INTRODUCTION

The probability distribution plays a major role in statistical modeling and analysis. There are numerous probability distributions in the literature that are widely used to describe data under study from many different areas including engineering, environmental

[1] Corresponding Author. E-mail: aryalg@pnw.edu

sciences, medical sciences, sports, finance and economics, among others. Various studies have shown that a particular data set following a standard probability distribution is more often the exception rather than the reality. Therefore, developing a more flexible distribution that can capture larger variability of data is of prime interest in the field applied statistics. A more flexible distribution can be developed by generalizing a classical distribution. Generalizations have been achieved in many different ways including compounding distributions or adding a new parameter(s) including a frailty parameter or a tilt parameter as discussed in Marshall and Olkin (2007). For the survey of compounding of distributions, readers are referred to Tahir and Cordeiro (2016). The generalization of classical distributions by adding different parameters have been discussed in many articles in the statistics literature. A brief survey of such techniques can be found in Ahmad *et al.* (2019). Although the addition of a parameter makes a given distribution more flexible, it may not produce significantly different results from that of the base distribution. Hence, having a higher number of parameters in a given distribution does not assure a better fit for given data. That is, a particular generalization might not end up with a better fit just because it possesses a larger number of parameters. In this paper, we have addressed the effectiveness of the generalization of McDonald-G Poisson family, which is generalized using the method introduced by Tahir and Cordeiro (2016). We also provide three different types of examples to address the usefulness of this generalized family. The McDonald generalized Poisson family is developed using the genesis of the zero truncated Poisson (ZTP) distribution by David and Johnson (1952) and the McDonald distribution by McDonald (1984). The McDonald distribution is also known as the generalized beta of first kind (GB1). The probability density function (PDF) and cumulative distribution function (CDF) of the McDonald ("Mc" for short) distribution are, respectively, given by

$$f(x) = \frac{c}{B(ac^{-1}, b)} x^{a-1}(1-x^c)^{b-1}, \quad 0 < x < 1 \tag{1}$$

and

$$F(x) = I_{x^c}\left(ac^{-1}, b\right) = \frac{1}{B(ac^{-1}, b)} \int_0^{x^c} \omega^{\frac{a}{c}-1}(1-\omega)^{b-1} \mathrm{d}\omega, \tag{2}$$

where $a > 0$, $b > 0$ and $c > 0$ are the shape parameters.

The McDonald distribution is very flexible, as it approaches different distributions when its parameters are appropriately chosen as below:

- for $c = 1$ Mc distribution reduces to the *beta* distribution.
- for $a = c$ Mc distribution reduces to the *Kumaraswamy* distribution.
- for $a = b = 1/2$ and $c = 1$ Mc distribution reduces to the *arcsine* distribution.

Note that the CDF of the McDonald distribution can be written in terms of the hypergeometric function as

$$F(x) = \frac{cx^a}{aB\left(ac^{-1}, b\right)} {}_2F_1\left(ac^{-1}, 1-b; ac^{-1}+1; x^a\right), \tag{3}$$

where

$${}_2F_1(a,b;c;x) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)} \sum_{j=0}^{\infty} \frac{\Gamma(a+j)\Gamma(b+j)}{\Gamma(c+j)} \frac{x^j}{j!}. \tag{4}$$

First, we provide a physical interpretation for the motivation behind the study as outlined in Tahir and Cordeiro (2016). This will illustrate the situation where the newly generalized family will be useful in practice. Suppose that a system has N subsystems functioning independently at a given time where N has a ZTP distribution with probability mass function (PMF)

$$P(N=n) = \frac{1}{[1-\exp(-\lambda)]} \frac{\exp(-\lambda)\lambda^n}{n!} \quad \text{for } n = 1, 2, \ldots. \tag{5}$$

Also, suppose that the failure time of each subsystem has the McDonald Generalized distribution with its CDF and PDF given by

$$H(x; a, b, c, \underline{\varphi}) = I_{G^c(x;\underline{\varphi})}(a, b) = \frac{1}{B(ac^{-1}, b)} \int_0^{G^c(x;\underline{\varphi})} \omega^{\frac{a}{c}-1} (1-\omega)^{b-1} \, d\omega \tag{6}$$

and

$$h(x; a, b, c, \underline{\varphi}) = \frac{c}{B(ac^{-1}, b)} \left[G(x;\underline{\varphi})\right]^{a-1} \left[1 - G^c(x;\underline{\varphi})\right]^{b-1} g(x;\underline{\varphi}), \tag{7}$$

respectively, where $a > 0, b > 0$ and $c > 0$ are the shape parameters and $\underline{\varphi}$ is the set of parameters for the random variable X with CDF $G(x;\underline{\varphi})$. Let Y_i denote the failure time of the ith subsystem and let $X = \min\{Y_1, Y_2, \ldots, Y_N\}$. Then, the conditional CDF of X given N is

$$\begin{aligned} F(x \mid N) &= 1 - P(X > x \mid N) \\ &= 1 - \left[1 - H\left(x; a, b, c, \underline{\varphi}\right)\right]^N. \end{aligned} \tag{8}$$

Therefore, using Eq. (5) and Eq. (8) the unconditional CDF of X can be expressed as

$$\begin{aligned} F(x) &= P(X \leq x) \\ &= \sum_{n=1}^{\infty} F(x|N) P(N=n) \\ &= \frac{1 - \exp[-\lambda H(x; a, b, c, \underline{\varphi})]}{1 - \exp(-\lambda)}. \end{aligned} \tag{9}$$

The CDF in Eq. (9) is called the McDonald-G Poisson (McGP) family of distributions. The corresponding PDF is given by

$$\begin{aligned} f\left(x;a,b,c,\lambda,\underline{\varphi}\right) &= \frac{\lambda h(x;a,b,c,\underline{\varphi})\exp[-\lambda H(x;a,b,c,\underline{\varphi})]}{1-\exp(-\lambda)} \\ &= \frac{c\lambda g(x)}{B(ac^{-1},b)}\frac{[G(x)]^{a-1}[1-\{G(x)\}^c]^{b-1}\exp[-\lambda I_{G^c}(a,b)]}{1-\exp(-\lambda)}. \end{aligned} \tag{10}$$

In the rest of the Section 1 we establish the relationship between the McGP family and the McDonald-G family as well as the relationship between McGP family and exponentiated-G family via Lemma 1 and Lemma , respectively. Section 2 explores six different cases of McGP family. Some mathematical and statistical characteristics of the McGP family are provided in Section 3. In Section 4 we describe different parameter estimation procedures. Real life application of the McGP family to model three different types of data is presented in Section 5. Section 6 provides some concluding remarks.

LEMMA 1. *The McDonald-G Poisson family of distribution can be expressed as a linear combination of McDonald-G (McG) distributions.*

PROOF. We know that the power series expansion of $\exp(-y)$ is given by

$$\exp(-y)=\sum_{k=0}^{\infty}\frac{(-y)^k}{k!}=\sum_{k=0}^{\infty}\frac{(-1)^k y^k}{k!}.$$

Therefore, using power series expansion to $\exp[-\lambda H(x;a,b,c,\underline{\varphi})]$ the PDF in Eq. (10) becomes

$$\begin{aligned} f(x;a,b,c,\lambda,\underline{\varphi}) &= \frac{\lambda h(x;a,b,c,\underline{\varphi})}{[1-\exp(-\lambda)]}\sum_{k=0}^{\infty}\frac{(-1)^k\lambda^k H^k(x;a,b,c,\underline{\varphi})}{k!} \\ &= h(x;a,b,c,\underline{\varphi})\sum_{k=0}^{\infty}\frac{(-1)^k\lambda^{k+1}}{k![1-\exp(-\lambda)]}H^k(x;a,b,c,\underline{\varphi}) \\ &= h(x;a,b,c,\underline{\varphi})\sum_{k=0}^{\infty}u_k H^k(x;a,b,c,\underline{\varphi}), \end{aligned} \tag{11}$$

where $u_k=\frac{(-1)^k\lambda^{k+1}}{k![1-\exp(-\lambda)]}$ and $k=0,1,2,\ldots$.

By using the power series expansion of $(1-\omega)^{b-1}$, we have

$$\begin{aligned} H(x;a,b,c,\underline{\varphi}) &= \frac{1}{B(ac^{-1},b)}\int_0^{G^c(x;\underline{\varphi})}\omega^{\frac{a}{c}-1}(1-\omega)^{b-1}\mathrm{d}\omega \\ &= \frac{1}{B(ac^{-1},b)}\sum_{i=0}^{\infty}(-1)^i\binom{b-1}{i}\frac{\left[G^c(x;\underline{\varphi})\right]^{\frac{a}{c}+i}}{\frac{a}{c}+i} \\ &= \frac{G^a(x;\underline{\varphi})}{B(ac^{-1},b)}\sum_{i=0}^{\infty}\frac{c(-1)^i}{a+ci}\binom{b-1}{i}\left[G(x;\underline{\varphi})\right]^{ci}. \end{aligned} \tag{12}$$

On letting $v_i = \frac{c(-1)^i}{a+ci}\binom{b-1}{i}$, the PDF in Eq. (10) reduces to

$$
\begin{aligned}
f(x) &= h(x;a,b,c,\underline{\varphi})\sum_{k=0}^{\infty} u_k \left(\frac{G^a(x;\underline{\varphi})}{B(ac^{-1},b)} \sum_{i=0}^{\infty} \frac{c(-1)^i}{a+ci}\binom{b-1}{i}\left[G(x;\underline{\varphi})\right]^{ci}\right)^k \\
&= h(x;a,b,c,\underline{\varphi})\sum_{k=0}^{\infty} u_k \frac{G^{ak}(x;\underline{\varphi})}{[B(ac^{-1},b)]^k}\left(\sum_{i=0}^{\infty} v_i G^{ci}(x;\underline{\varphi})\right)^k .
\end{aligned}
\tag{13}
$$

From Section 0.314 of Gradshteyn and Ryzhik (2000), we know that for any positive integer k

$$
\left(\sum_{i=0}^{\infty} m_i n^i\right)^k = \sum_{i=0}^{\infty} d_{k,i} n^i,
\tag{14}
$$

where the coefficients $d_{k,i}$ for $i = 1,2,3,\ldots$ can be determined from the recurrence equation $d_{k,0} = m_0^k$ and $d_{k,i} = (i m_0)^{-1}\sum_{q=1}^{i}[q(k+1)-i]m_q d_{k,i-q}$.

Thus, the PDF of McGP can be further simplified as

$$
\begin{aligned}
f(x) &= h(x;a,b,c,\underline{\varphi})\sum_{k=0}^{\infty} u_k \frac{G^{ak}(x;\underline{\varphi})}{[B(ac^{-1},b)]^k}\sum_{i=0}^{\infty} d_{k,i} G^{ci}(x;\underline{\varphi}) \\
&= \sum_{k=0}^{\infty}\sum_{i=0}^{\infty} \frac{u_k d_{k,i}}{[B(ac^{-1},b)]^k} G^{ak+ci}(x;\underline{\varphi}) h(x;a,b,c,\underline{\varphi}) \\
&= \sum_{k=0}^{\infty}\sum_{i=0}^{\infty} \frac{u_k d_{k,i}}{[B(ac^{-1},b)]^k} \times \\
&\times \left(\frac{c}{B(ac^{-1},b)}\left[G(x;\underline{\varphi})\right]^{a-1}\left[1-G^c(x;\underline{\varphi})\right]^{b-1} g(x;\underline{\varphi})\right) G^{ak+ci}(x;\underline{\varphi}) \\
&= \sum_{k=0}^{\infty}\sum_{i=0}^{\infty} \frac{u_k d_{k,i}}{[B(ac^{-1},b)]^k} \times \\
&\times \left(\frac{c}{B(ac^{-1},b)}\left[G(x;\underline{\varphi})\right]^{a(k+1)+ci-1}\left[1-G^c(x;\underline{\varphi})\right]^{b-1} g(x;\underline{\varphi})\right) \\
&= \sum_{k=0}^{\infty}\sum_{i=0}^{\infty} \frac{u_k d_{k,i} B\left((k+1)ac^{-1}+i,b\right)}{[B(ac^{-1},b)]^{k+1}} \\
&\times \left(\frac{c}{B\left((k+1)ac^{-1}+i,b\right)}\left[G(x;\underline{\varphi})\right]^{a(k+1)+ci-1}\left[1-G^c(x;\underline{\varphi})\right]^{b-1} g(x;\underline{\varphi})\right) \\
&= \sum_{k=0}^{\infty}\sum_{i=0}^{\infty} \pi_{ki} h(x;a(k+1)+ci,b,c,\underline{\varphi}),
\end{aligned}
\tag{15}
$$

where $\pi_{ki} = \frac{u_k d_{k,i} B\left((k+1)ac^{-1}+i,b\right)}{[B(ac^{-1},b)]^{k+1}}$.

The CDF corresponding to Eq. (15) can be expressed as

$$F(x;a,b,c,\lambda,\underline{\varphi})=\sum_{k=0}^{\infty}\sum_{i=0}^{\infty}\pi_{ki}H(x;a(k+1)+ci,b,c,\underline{\varphi}) \tag{16}$$

$$=\sum_{k=0}^{\infty}\sum_{i=0}^{\infty}\pi_{ki}I_{G^c(x;\underline{\varphi})}(a(k+1)+ci,b,c). \tag{17}$$

□

LEMMA 2. *The McDonald-G Poisson family of distributions can be expressed as a linear combination of exponentiated-G distributions.*

PROOF. Consider a random variable X with its CDF $G(x)$ and let $\alpha>0$ be a parameter. Then the exponentiated-G will have its PDF and CDF given by $\Psi_\alpha(x;\underline{\varphi})=G^\alpha(x;\underline{\varphi})$ and $\psi_\alpha(x;\underline{\varphi})=\alpha g(x;\underline{\varphi})G^{\alpha-1}(x;\underline{\varphi})$, respectively.

Substituting Eq. (7) into Eq. (15) and using a binomial expansion, we get

$$f(x)=\sum_{k=0}^{\infty}\sum_{i=0}^{\infty}\sum_{l=0}^{\infty}\pi_{ki}(-1)^l\binom{b-1}{l}\frac{c}{B\left((k+1)ac^{-1},b\right)}\left[G(x;\underline{\varphi})\right]^{a(k+1)+ci+l-1}g(x;\underline{\varphi})$$

$$=\sum_{k=0}^{\infty}\sum_{i=0}^{\infty}\sum_{l=0}^{\infty}\frac{(-1)^l\binom{b-1}{l}c\pi_{ki}}{B\left((k+1)ac^{-1},b\right)[a(k+1)+ci+l]}\psi_{a(k+1)+ci+l}(x;\underline{\varphi}). \tag{18}$$

Therefore, we have

$$f(x;a,b,c,\lambda,\underline{\varphi})=\sum_{k=0}^{\infty}\sum_{i=0}^{\infty}\sum_{l=0}^{\infty}\delta_{kil}\psi_{a(k+1)+ci+l}(x;\underline{\varphi}), \tag{19}$$

where $\delta_{kil}=\frac{(-1)^l\binom{b-1}{l}c\pi_{ki}}{B((k+1)ac^{-1},b)[a(k+1)+ci+l]}$.

We integrate Eq. (19) to get the corresponding CDF as

$$F(x;a,b,c,\lambda,\underline{\varphi})=\sum_{k=0}^{\infty}\sum_{i=0}^{\infty}\sum_{l=0}^{\infty}\delta_{kil}\Psi_{a(k+1)+ci+l}(x;\underline{\varphi}), \tag{20}$$

where

$$\Psi_{a(k+1)+ci+l}=\left[G(x;\underline{\varphi})\right]^{a(k+1)+ci+l} \tag{21}$$

is the CDF of the exp-G distribution with power parameter $a(k+1)+ci+l$ and

$$\psi_{a(k+1)+ci+l}=[a(k+1)+ci+l]\left[G(x;\underline{\varphi})\right]^{a(k+1)+ci+l-1}g(x;\underline{\varphi}) \tag{22}$$

is the PDF of exp-G distribution with power parameter $a(k+1)+ci+l$. □

2. SPECIAL MODELS

In this Section we present some special models of the McGP family. The PDF in Eq. (10) will be most tractable when the CDF $G(x;\underline{\varphi})$ and PDF $g(x;\underline{\varphi})$ have simple analytical expressions. These sub-models generalize several important distributions in the literature. We will consider the following distributions listed in Table 1: Exponential (Ex), Weibull (W), Pareto (Pa), Log-logistic (LL), Fréchet (Fr) and Lindley (Li).

TABLE 1
The PDF and CDF of base models of McGP family.

Model	$PDF: g(x;\underline{\varphi})$	$CDF: G(x;\underline{\varphi})$	Support
Ex	$\alpha\exp(-\alpha x)$	$1-\exp(-\alpha x)$	$(0,\infty)$
W	$\beta\alpha^{\beta}x^{\beta-1}\exp\left[-(\alpha x)^{\beta}\right]$	$1-\exp\left[-(\alpha x)^{\beta}\right]$	$(0,\infty)$
Pa	$\left(\frac{\alpha}{x}\right)\left(\frac{\theta}{x}\right)^{\alpha}$	$1-\left(\frac{\theta}{x}\right)^{\alpha}$	(θ,∞)
LL	$\beta\alpha^{-\beta}x^{\beta-1}\left[1+\left(\frac{x}{\alpha}\right)^{\beta}\right]^{-2}$	$\left[1+\left(\frac{\alpha}{x}\right)^{\beta}\right]^{-1}$	$(0,\infty)$
Fr	$\beta\alpha^{\beta}x^{-(\beta+1)}\exp\left[-\left(\frac{\alpha}{x}\right)^{\beta}\right]$	$\exp\left[-\left(\frac{\alpha}{x}\right)^{\beta}\right]$	$(0,\infty)$
Li	$\frac{\alpha^2}{1+\alpha}(1+x)\exp(-\alpha x)$	$1-\frac{1+\alpha+\alpha x}{1+\alpha}\exp(-\alpha x)$	$(0,\infty)$

2.1. *The McExP distribution*

The CDF and PDF of the McDonald-exponential Poisson (McExP) distribution are given, respectively, by

$$F(x)=\frac{1-\exp\left[-\lambda I_{G^c(x;\alpha)}(a,b)\right]}{[1-\exp(-\lambda)]} \tag{23}$$

and

$$f(x) = \frac{c\lambda\alpha\exp(-\alpha x)\exp[-\lambda I_{G^c(x;\alpha)}(a,b)]}{B(ac^{-1},b)[1-\exp(-\lambda)]} \tag{24}$$

$$\times[1-\exp\{-(\alpha x)\}]^{a-1}[1-\{1-\exp(-\alpha x)\}^{c}]^{b-1}, \tag{25}$$

where

$$I_{G^c(x;\alpha)}(a,b)=\frac{1}{B(ac^{-1},b)}\int_0^{[1-\exp(-\alpha x)]^c} w^{\frac{a}{c}-1}(1-w)^{b-1}\mathrm{d}w. \tag{26}$$

Plots of the PDF and CDF of the McExP distribution are displayed in Figure 1 for selected parameter values.

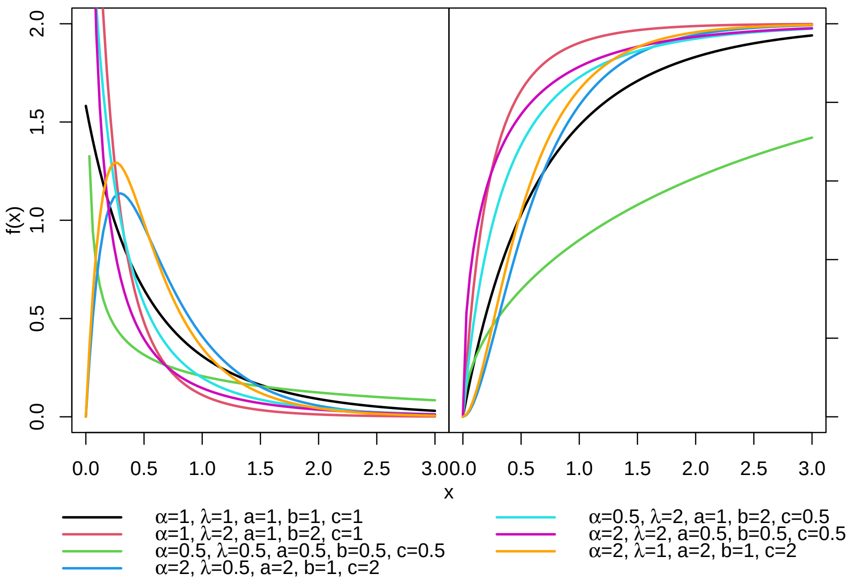

Figure 1 – PDF (left panel) and CDF (right panel) of McExP distribution.

2.2. *The McWP distribution*

The CDF and PDF of the McDonald-Weibull Poisson (McWP) distribution are given, respectively, by

$$F(x) = \frac{1-\exp\left[-\lambda I_{G^c(x;\alpha,\beta)}(a,b)\right]}{[1-\exp(-\lambda)]} \tag{27}$$

and

$$f(x) = \frac{c\lambda\beta\alpha^{\beta}x^{\beta-1}\exp[-(\alpha x)^{\beta}]\exp\left[-\lambda I_{G^c(x;\alpha,\beta)}(a,b)\right]}{B(ac^{-1},b)[1-\exp(-\lambda)]} \tag{28}$$

$$\times[1-\exp\{-(\alpha x)^{\beta}\}]^{a-1}\left[1-\left\{1-\exp(-(\alpha x))^{\beta}\right\}^{c}\right]^{b-1}, \tag{29}$$

where

$$I_{G^c(x;\alpha,\beta)}(a,b) = \frac{1}{B(ac^{-1},b)} \int_0^{\left[1-\exp\{-(\alpha x)^{\beta}\}\right]^c} w^{\frac{a}{c}-1}(1-w)^{b-1}\mathrm{d}w. \tag{30}$$

Plots of the PDF and CDF of the McWP distribution are displayed in Figure 2 for selected parameter values.

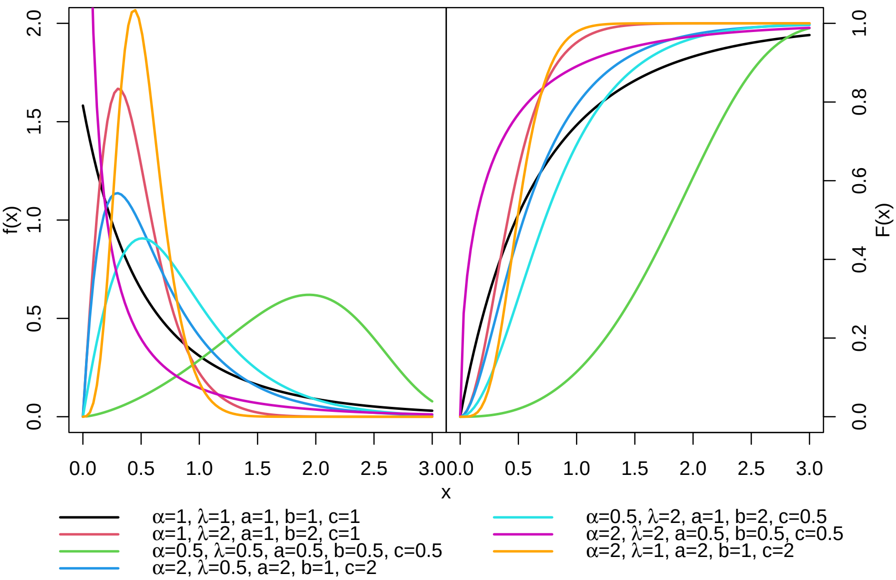

Figure 2 – PDF (left panel) and CDF (right panel) of McWP distribution.

2.3. *The McPaP distribution*

The CDF and PDF of the McDonald-Pareto Poisson (McPaP) distribution are given, respectively, by

$$F(x) = \frac{1-\exp\left[-\lambda I_{G^c(x;\alpha,\theta)}(a,b)\right]}{[1-\exp(-\lambda)]} \tag{31}$$

and

$$f(x) = \frac{c\lambda\alpha\theta^{\alpha}}{B(ac^{-1},b)x^{\alpha+1}} \frac{\left[1-(\frac{\theta}{x})^{\alpha}\right]^{a-1}\left[1-\left\{1-(\frac{\theta}{x})^{\alpha}\right\}^{c}\right]^{b-1}}{[1-\exp(-\lambda)]} \exp\left[-\lambda I_{G^{c}(x;\alpha,\theta)}(a,b)\right], \quad (32)$$

where

$$I_{G^{c}(x;\alpha,\beta)}(a,b) = \frac{1}{B(ac^{-1},b)} \int_{0}^{\left[1-\left(\frac{\theta}{x}\right)^{\alpha}\right]^{c}} w^{\frac{a}{c}-1}(1-w)^{b-1}\mathrm{d}w. \quad (33)$$

Plots of the PDF and CDF of the McPaP distribution are displayed in Figure 3 for selected parameter values.

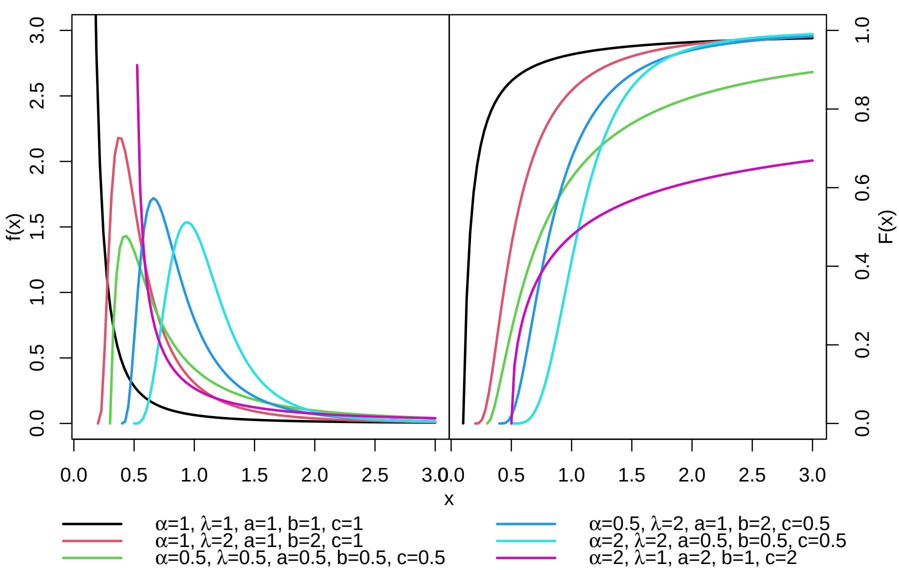

Figure 3 – PDF (left panel) and CDF (right panel) of McPaP distribution.

2.4. The McFrP distribution

The CDF and PDF of the McDonald Fréchet Poisson (McFrP) distribution are given, respectively, by

$$F(x) = \frac{1-\exp\left[-\lambda I_{G^c(x;\alpha,\beta)}(a,b)\right]}{[1-\exp(-\lambda)]} \tag{34}$$

and

$$f(x) = \frac{c\lambda\beta\alpha^{\beta}\exp\left[-a(\frac{\alpha}{x})^{\beta}\right]}{B(ac^{-1},b)x^{\beta+1}} \frac{\left\{1-\exp[-c\left(\frac{\alpha}{x}\right)^{\beta}]\right\}^{b-1}}{[1-\exp(-\lambda)]} \exp\left[-\lambda I_{G^c(x;\alpha,\beta)}(a,b)\right], \tag{35}$$

where

$$I_{G^c(x;\alpha,\beta)}(a,b) = \frac{1}{B(ac^{-1},b)} \int_0^{\exp[-c\left(\frac{\alpha}{x}\right)^{\beta}]} w^{\frac{a}{c}-1}(1-w)^{b-1}\mathrm{d}w. \tag{36}$$

Plots of the PDF and CDF of the McFrP distribution are displayed in Figure 4 for selected parameter values.

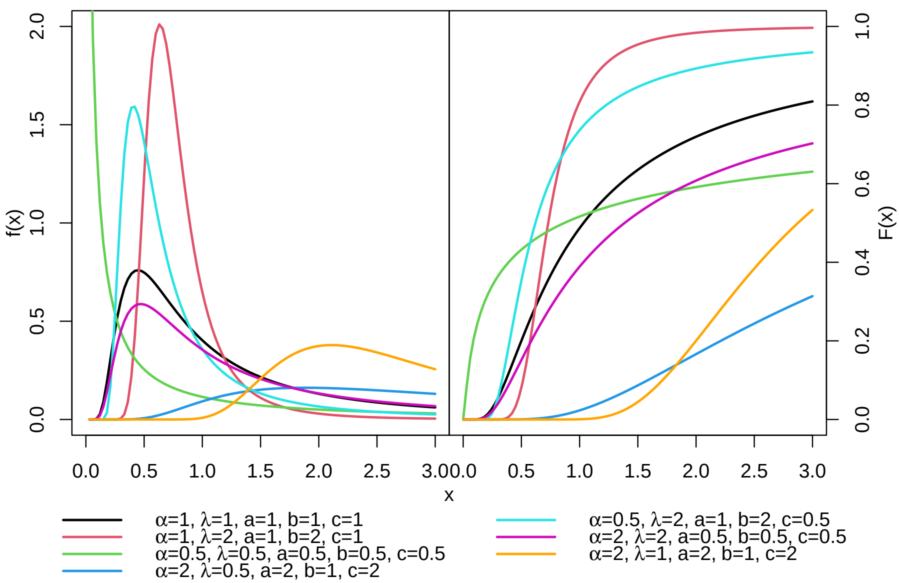

Figure 4 – PDF (left panel) and CDF (right panel) of McFrP distribution.

2.5. *The McLLP distribution*

The CDF and PDF of the McDonald log-logistic Poisson (McLLP) distribution are given, respectively, by

$$F(x)=\frac{1-\exp\left[-\lambda I_{G^c(x;\alpha,\beta)}(a,b)\right]}{[1-\exp(-\lambda)]} \tag{37}$$

and

$$f(x) = \frac{c\lambda\beta\alpha^{-\beta}x^{\beta-1}\left[1+\left(\frac{x}{\alpha}\right)^{\beta}\right]^{-2}}{B(ac^{-1},b)[1-\exp(-\lambda)]}\frac{\left[1-\left[1+\left(\frac{\alpha}{x}\right)^{\beta}\right]^{-c}\right]^{b-1}}{\left[1+\left(\frac{\alpha}{x}\right)^{\beta}\right]^{a-1}}\exp\left[-\lambda I_{G^c(x;\alpha,\beta)}(a,b)\right], \tag{38}$$

where

$$I_{G^c(x;\alpha,\beta)}(a,b)=\frac{1}{B(ac^{-1},b)}\int_0^{\left[1+\left(\frac{\alpha}{x}\right)^{\beta}\right]^{-c}} w^{\frac{a}{c}-1}(1-w)^{b-1}\mathrm{d}w. \tag{39}$$

Plots of the PDF and CDF of the McLLP distribution are displayed in Figure 5 for selected parameter values.

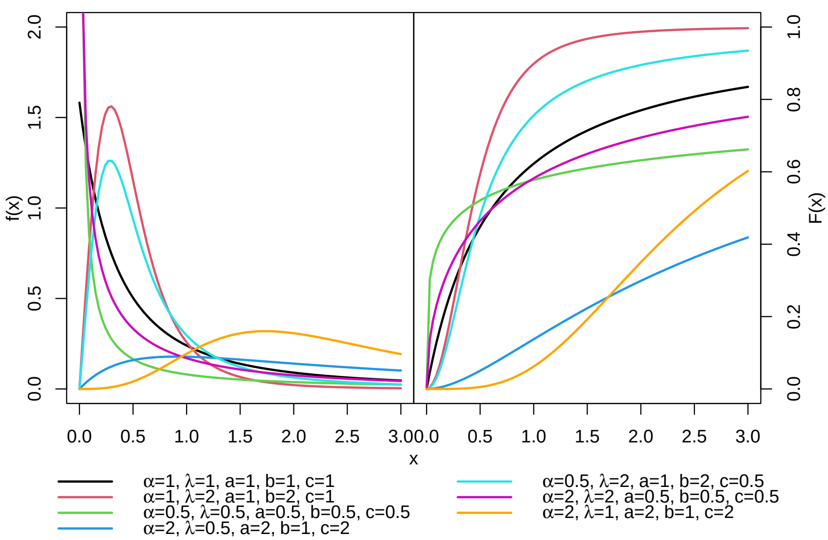

Figure 5 – PDF (left panel) and CDF (right panel) of McLLP distribution.

2.6. *The McLiP distribution*

The CDF and PDF of the McDonald-Lindley Poisson (McLiP) distribution are given, respectively, by

$$F(x)=\frac{1-\exp\left[-\lambda I_{G^c(x;\alpha)}(a,b)\right]}{[1-\exp(-\lambda)]} \tag{40}$$

and

$$f(x) = \frac{c\lambda\alpha^2(1+x)\exp(-\alpha x)}{(1+\alpha)B(ac^{-1},b)}\frac{\exp\left[-\lambda I_{G^c(x;\alpha)}(a,b)\right]}{[1-\exp(-\lambda)]} \tag{41}$$

$$\times\left[1-\frac{1+\alpha+\alpha x}{1+\alpha}\exp(-\alpha x)\right]^{a-1}\left[1-\left(1-\frac{1+\alpha+\alpha x}{1+\alpha}\exp(-\alpha x)\right)^c\right]^{b-1} \tag{42}$$

where

$$I_{G^c(x;\alpha)}(a,b)=\frac{1}{B(ac^{-1},b)}\int_0^{[1-\frac{1+\alpha+\alpha x}{1+\alpha}\exp(-\alpha x)]^c} w^{\frac{a}{c}-1}(1-w)^{b-1}dw. \tag{43}$$

Plots of the PDF and CDF of the McLiP distribution are displayed in Figure 6 for selected parameter values.

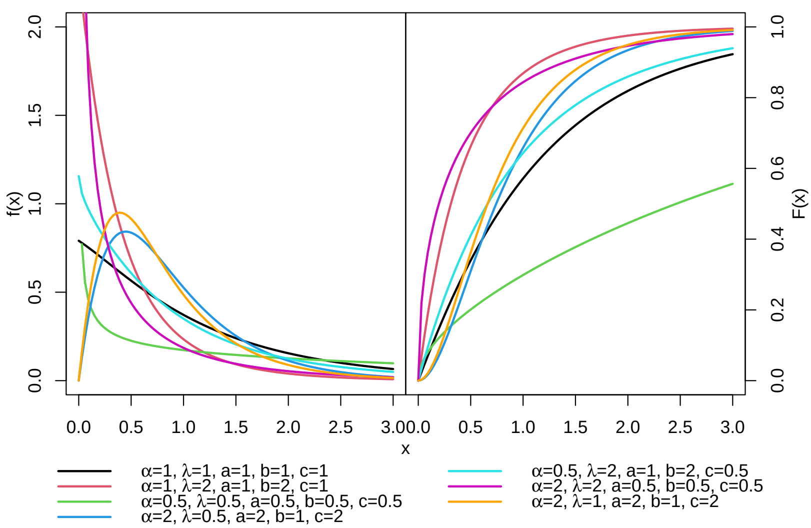

Figure 6 – PDF (left panel) and CDF (right panel) of McLiP distribution.

3. Mathematical properties

In this Section we provide some structural and mathematical properties of the MGP distribution including the quantile function, moments, probability weighted moment, and components of reliability model.

3.1. *Quantile function*

The quantile function of a distribution is the real solution of $F(x_q) = q$ for $0 \leq q \leq 1$. The quantile function is obtained by inverting Eq. (9) provided that a closed form expression is available. Setting

$$F(x) = \frac{1-\exp[-\lambda H(x;a,b,c,\underline{\varphi})]}{[1-\exp(-\lambda)]} = \frac{1-\exp\left[-\lambda I_{G^c(x;\underline{\varphi})}(a,b)\right]}{[1-\exp(-\lambda)]} = q, \tag{44}$$

we have

$$I_{G^c(x;\underline{\varphi})}(a,b) = -\lambda^{-1}\ln[1-q+q\exp(-\lambda)]. \tag{45}$$

Since the quantile function for the beta distribution has no closed form, we make use of the beta distribution's quantile function `qbeta` in statistical software `R`. Therefore,

$$G(x;\underline{\varphi}) = \left\{\texttt{qbeta}\left(-\lambda^{-1}\ln[1-q+q\exp(-\lambda)],a,b\right)\right\}^{1/c}. \tag{46}$$

Given the CDF $G(x;\underline{\varphi})$ of a random variable X, it is easy to obtain the quantiles of the McGP distribution from the equation above. That is,

$$x_q = G^{-1}\left[\left\{\texttt{qbeta}(-\lambda^{-1}\ln[1-q+q\exp(-\lambda)],a,b),\underline{\varphi}\right\}^{1/c}\right]. \tag{47}$$

Further, one can use Eq. (47) to obtain the median, as well as octiles and then the measure of Bowley's skewness and Moors kurtosis. These measures are quartile alternatives to the traditional skewness and kurtosis, and are more robust.

3.2. Moments

The r-th order moment of the McGP distribution is given by

$$\begin{aligned}
E(X^r) &= \int_{-\infty}^{\infty} x^r f(x;a,b,c,\lambda,\underline{\varphi})\mathrm{d}x \\
&= \int_{-\infty}^{\infty} x^r \sum_{k=0}^{\infty}\sum_{i=0}^{\infty}\sum_{l=0}^{\infty} \delta_{kil}\psi_{a(k+1)+ci+l}(x;\underline{\varphi})\mathrm{d}x \\
&= \sum_{k=0}^{\infty}\sum_{i=0}^{\infty}\sum_{l=0}^{\infty} \delta_{kil}\int_{-\infty}^{\infty} x^r \psi_{a(k+1)+ci+l}(x;\underline{\varphi})\mathrm{d}x \\
&= \sum_{k=0}^{\infty}\sum_{i=0}^{\infty}\sum_{l=0}^{\infty} \delta_{kil}\int_{-\infty}^{\infty} x^r [a(k+1)+ci+l]\Big[G(x;\underline{\varphi})\Big]^{a(k+1)+ci+l-1} g(x;\underline{\varphi})\mathrm{d}x \\
&= \sum_{k=0}^{\infty}\sum_{i=0}^{\infty}\sum_{l=0}^{\infty} \delta_{kil}\int_{-\infty}^{\infty} x^r [a(k+1)+ci+l]\Big[G(x;\underline{\varphi})\Big]^{a(k+1)+ci+l-1} g(x;\underline{\varphi})\mathrm{d}x \\
&= \sum_{k=0}^{\infty}\sum_{i=0}^{\infty}\sum_{l=0}^{\infty} \delta_{kil}[a(k+1)+ci+l]\int_{-\infty}^{\infty} x^r \Big[G(x;\underline{\varphi})\Big]^{a(k+1)+ci+l-1} g(x;\underline{\varphi})\mathrm{d}x \\
&= \sum_{k=0}^{\infty}\sum_{i=0}^{\infty}\sum_{l=0}^{\infty} \delta_{kil}[a(k+1)+ci+l]\zeta_{r,a(k+1)+ci+l-1}, \qquad (48)
\end{aligned}$$

where $\zeta_{r,s} = \int_{-\infty}^{\infty} x^r G^s(x;\underline{\varphi})g(x;\underline{\varphi})\mathrm{d}x$.

3.3. Reliability analysis

The usefulness of the reliability function $R(x)$, the hazard rate function $h(x)$, and the reversed hazard rate function $r(x)$ is well-known in literature. For an McGP random variable X, these quantities are given respectively by

$$R(x) = \frac{\exp\Big[-\lambda I_{G^c(x;\underline{\varphi})}(a,b)\Big] - \exp[-\lambda]}{[1-\exp(-\lambda)]}, \qquad (49)$$

$$h(x) = \frac{c\lambda g(x;\underline{\varphi})\Big[G(x;\underline{\varphi})\Big]^{a-1}\Big[1-G^c(x;\underline{\varphi})\Big]^{b-1}}{B(ac^{-1},b)\Big[1-\exp\Big\{-\lambda(1-I_{G^c(x;\underline{\varphi})}(a,b))\Big\}\Big]} \qquad (50)$$

and

$$r(x) = \frac{c\lambda g(x;\underline{\varphi})\Big[G(x;\underline{\varphi})\Big]^{a-1}\Big[1-G^c(x;\underline{\varphi})\Big]^{b-1}}{B(ac^{-1},b)\Big[\exp\Big\{\lambda I_{G^c(x;\underline{\varphi})}(a,b)\Big\}-1\Big]}. \qquad (51)$$

4. Parameter estimation

4.1. Method of maximum likelihood

In this Section the parameter estimation using the method of maximum likelihood is described. Let $X_1, X_2, \ldots, X_n$ be a random sample from the McGP distribution with parameters λ, a, b, c and $\underline{\varphi}$. Let $\underline{\Theta} = (a, b, c, \lambda, \underline{\varphi}^{\intercal})^{\intercal}$ be a $(p+4)\times 1$ parameter vector, where $\underline{\varphi}$ is a $(p \times 1)$ baseline parameter vector. For determining the MLE of $\underline{\Theta}$, the likelihood function may be expressed as

$$
\begin{aligned}
L &= c^n \lambda^n \prod_{i=1}^{n} g(x_i;\underline{\varphi}) \exp\left(-\sum_{i=1}^{n} \lambda I_{G^c(x_i;\underline{\varphi})}(a,b)\right) \left\{B(ac^{-1},b)\right\}^{-n} [1-\exp(-\lambda)]^{-n} \\
&\quad \times \prod_{i=1}^{n} \left\{\left[G(x_i;\underline{\varphi})\right]^{a-1} [1-G^c(x_i;\underline{\varphi})]^{b-1}\right\}. \qquad (52)
\end{aligned}
$$

Therefore, the log-likelihood function $\ell = \ln(L)$ reduces to

$$
\begin{aligned}
\ell &= n\ln(c) + n\ln(\lambda) + \sum_{i=1}^{n} \ln\left\{g(x_i;\underline{\varphi})\right\} - \lambda \sum_{i=1}^{n} I_{G^c(x_i;\underline{\varphi})}(a,b) + n\ln\Gamma(ac^{-1}+b) \\
&\quad - n\ln\Gamma(ac^{-1}) - n\ln\Gamma(b) - n\ln\{1-\exp(-\lambda)\} + (a-1)\sum_{i=1}^{n} \ln\left\{G(x_i;\underline{\varphi})\right\} \\
&\quad + (b-1)\sum_{i=1}^{n} \ln\left\{1-G^c(x_i;\underline{\varphi})\right\}. \qquad (53)
\end{aligned}
$$

The components of the score vector are obtained by taking the partial derivatives of the log-likelihood function with respect to the parameters. $\underline{U}(\underline{\Theta}) = \frac{\partial \ell}{\partial \underline{\Theta}} = (\frac{\partial \ell}{\partial \lambda}, \frac{\partial \ell}{\partial a}, \frac{\partial \ell}{\partial b}, \frac{\partial \ell}{\partial c}, \frac{\partial \ell}{\partial \underline{\varphi}})^{\intercal}$ are given by

$$
\begin{aligned}
U_\lambda &= \frac{n}{\lambda} - \frac{n\exp(-\lambda)}{[1-\exp(-\lambda)]} - \sum_{i=1}^{n} I_{G^c(x_i,\underline{\varphi})}(a,b), \\
U_a &= \frac{n}{c}\left[\psi(ac^{-1}+b) - \psi(ac^{-1})\right] - \lambda \sum_{i=1}^{n} I^a_{G^c(x_i,\underline{\varphi})}(a,b) + \sum_{i=1}^{n} \ln\left\{G(x_i,\underline{\varphi})\right\}, \\
U_b &= n\left[\psi(ac^{-1}+b) - \psi(b)\right] - \lambda \sum_{i=1}^{n} I^b_{G^c(x_i,\underline{\varphi})}(a,b) + \sum_{i=1}^{n} \ln\left\{1-G^c(x_i,\underline{\varphi})\right\}, \\
U_c &= \frac{n}{c} + \frac{na}{c^2}\left[\psi(ac^{-1}) - \psi(ac^{-1}+b)\right] - \lambda \sum_{i=1}^{n} I^c_{G^c(x_i,\underline{\varphi})}(a,b) \\
&\quad - c(b-1)\sum_{i=1}^{n} \frac{G^{c-1}(x_i)}{1-G^c(x_i,\underline{\varphi})} g(x_i,\underline{\varphi}) \qquad (54)
\end{aligned}
$$

and, for $r = 1, 2, ..., p$,

$$
\begin{aligned}
U_{\underline{\varphi}_r} &= \sum_{i=1}^{n} \frac{g'_{\varphi_r}\left(x_i;\underline{\varphi}\right)}{g\left(x_i;\underline{\varphi}\right)} + (a-1)\sum_{i=1}^{n} \frac{G'_{\varphi_r}\left(x_i;\underline{\varphi}\right)}{G\left(x_i;\underline{\varphi}\right)} - (b-1)\sum_{i=1}^{n} \frac{{G^c_{\varphi_r}}'\left(x_i;\underline{\varphi}\right)}{1-G^c\left(x_i;\underline{\varphi}\right)} \\
&\quad -\lambda \sum_{i=1}^{n} I^{\varphi_r}_{G^c(x_i;\underline{\varphi})}(a,b), \qquad (55)
\end{aligned}
$$

where $\psi(.)$ is the digamma function defined as $\psi(x) = \frac{\mathrm{d}}{\mathrm{d}x}(\ln\Gamma(x))$, and we have

$$
\begin{aligned}
I^{a}_{G^c(x_i;\underline{\varphi})}(a,b) &= \frac{\partial I_{G^c(x_i;\underline{\varphi})(a,b)}}{\partial a}, && (56)\\
I^{b}_{G^c(x_i;\underline{\varphi})}(a,b) &= \frac{\partial I_{G^c(x_i;\underline{\varphi})}(a,b)}{\partial b}, && (57)\\
I^{c}_{G^c(x_i;\underline{\varphi})}(a,b) &= \frac{\partial I_{G^c(x_i;\underline{\varphi})}(a,b)}{\partial c}, && (58)\\
I^{\varphi_r}_{G^c(x_i;\underline{\varphi})}(a,b) &= \frac{\partial I_{G^c(x_i;\underline{\varphi})}(a,b)}{\partial \underline{\varphi}_r}. && (59)
\end{aligned}
$$

The maximum likelihood estimate (MLE) of $\underline{\Theta}$ is $\widehat{\underline{\Theta}} = (\widehat{\lambda}, \widehat{a}, \widehat{b}, \widehat{c}, \widehat{\underline{\varphi}}^{\intercal})^{\intercal}$, setting the nonlinear system of equations $U_\lambda = U_a = U_b = U_c = U_{\varphi_r} = 0$ and solving them simultaneously. These equations can be solved numerically using iterative methods such as Newton-Raphson type algorithms. For interval estimation of the model parameters, we derive the observed information matrix $J(\underline{\Theta})$. Under standard regularity conditions, when $n \to \infty$, the distribution of $\widehat{\underline{\Theta}}$ can be approximated by a multivariate normal $N_p(0, J(\widehat{\underline{\Theta}})^{-1})$ distribution to construct approximate confidence intervals for the parameters. Here, $J(\widehat{\underline{\Theta}})$ is the total observed information matrix evaluated at $\widehat{\underline{\Theta}}$.

4.2. Method of ordinary and weighted least-squares estimation

Let $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ be the order statistics of a random sample of size n from the McGP distribution with parameters λ, a, b, c and $\underline{\varphi}$. The least square estimators of the λ, a, b, c and $\underline{\varphi}$ can be obtained by minimizing

$$
\sum_{i=1}^{n} \left[\frac{1-\exp[-\lambda H(x_{(i)}; a, b, c, \underline{\varphi})]}{[1-\exp(-\lambda)]} - \frac{i}{n+1} \right]^2 \qquad (60)
$$

with respect to the unknown parameters λ, a, b, c, and $\underline{\varphi}$.

The weighted least squares estimator of the unknown parameters is obtained by minimizing

$$\sum_{i=1}^{n} \frac{(n+1)^2(n+2)}{n-i+1}\left[\frac{1-\exp[-\lambda H(x_{(i)};a,b,c,\underline{\varphi})]}{[1-\exp(-\lambda)]}-\frac{i}{n+1}\right]^2 \tag{61}$$

with respect to the unknown parameters λ, a, b, c, and $\underline{\varphi}$.

4.3. *Method of maximum product spacings estimation*

The Maximum Product of Spacings (MPS) estimation method was introduced by Cheng and Amin (1979), Cheng and Amin (1983), as an alternative to the maximum likelihood estimation (MLE) method for the estimation of parameters of continuous univariate distributions. It is based on maximization of the geometric mean of spacing of the data. The geometric mean of the differences is given by

$$G(a,b,c,\lambda,\underline{\varphi}) = \left(\prod_{i=1}^{n+1} D_i\right)^{\frac{1}{n+1}}, \tag{62}$$

where D_i is defined as

$$D_i = F(x_{i:n}|a,b,c,\lambda,\underline{\varphi}) - F(x_{i-1:n}|a,b,c,\lambda,\underline{\varphi}) = \int_{x_{i-1}}^{x_i} f(x;a,b,c,\underline{\varphi})\mathrm{d}x, \tag{63}$$

for $i = 1,2,\ldots,n+1$, such that $F(x_{(0)};a,b,c,\lambda,\underline{\varphi}) = 0$ and $F(x_{(n+1)};a,b,c,\lambda,\underline{\varphi}) = 1$. The maximum product spacing estimators of a, b, c, λ, and $\underline{\varphi}$ are obtained by maximizing

$$H(a,b,c,\lambda,\underline{\varphi}) = \log\left[G(a,b,c,\lambda,\underline{\varphi})\right] = \frac{1}{n+1}\sum_{i=1}^{n+1}\log\left[D_i(a,b,c,\lambda,\underline{\varphi})\right]. \tag{64}$$

The estimates of $a, b, c, \lambda, \underline{\varphi}$ are obtained by solving the following non-linear equations:

$$\frac{\partial H}{\partial a} = \frac{1}{n+1}\sum_{i=1}^{n+1}\frac{1}{D_i(a,b,c,\lambda,\underline{\varphi})}\times \times\left[\frac{\partial F(x_{i:n}|a,b,c,\lambda,\underline{\varphi})}{\partial a}-\frac{\partial F(x_{i-1:n}|a,b,c,\lambda,\underline{\varphi})}{\partial a}\right]=0 \tag{65}$$

$$\frac{\partial H}{\partial b} = \frac{1}{n+1}\sum_{i=1}^{n+1}\frac{1}{D_i(a,b,c,\lambda,\underline{\varphi})}\times \times\left[\frac{\partial F(x_{i:n}|a,b,c,\lambda,\underline{\varphi})}{\partial b}-\frac{\partial F(x_{i-1:n}|a,b,c,\lambda,\underline{\varphi})}{\partial b}\right]=0 \tag{66}$$

$$\frac{\partial H}{\partial c} = \frac{1}{n+1}\sum_{i=1}^{n+1}\frac{1}{D_i(a,b,c,\lambda,\underline{\varphi})}\times \times\left[\frac{\partial F(x_{i:n}|a,b,c,\lambda,\underline{\varphi})}{\partial c}-\frac{\partial F(x_{i-1:n}|a,b,c,\lambda,\underline{\varphi})}{\partial c}\right]=0 \tag{67}$$

$$\frac{\partial H}{\partial \lambda} = \frac{1}{n+1}\sum_{i=1}^{n+1}\frac{1}{D_i(a,b,c,\lambda,\underline{\varphi})}\times \times\left[\frac{\partial F(x_{i:n}|a,b,c,\lambda,\underline{\varphi})}{\partial \lambda}-\frac{\partial F(x_{i-1:n}|a,b,c,\lambda,\underline{\varphi})}{\partial \lambda}\right]=0 \tag{68}$$

$$\frac{\partial H}{\partial \underline{\varphi}} = \frac{1}{n+1}\sum_{i=1}^{n+1}\frac{1}{D_i(a,b,c,\lambda,\underline{\varphi})}\times \times\left[\frac{\partial F(x_{i:n}|a,b,c,\lambda,\underline{\varphi})}{\partial \underline{\varphi}}-\frac{\partial F(x_{i-1:n}|a,b,c,\lambda,\underline{\varphi})}{\partial \underline{\varphi}}\right]=0 \tag{69}$$

4.4. *Method of percentile estimation*

We know that the cdf is given by

$$F(x;a,b,c,\lambda,\underline{\varphi}) = \frac{1-\exp[-\lambda H(x;a,b,c,\underline{\varphi})]}{[1-\exp(-\lambda)]} = \frac{1-\exp\left[-\lambda I_{G^c(x;\underline{\varphi})}(a,b)\right]}{[1-\exp(-\lambda)]}, \tag{70}$$

which yields

$$x = G^{-1}\Big[\Big\{\texttt{qbeta}(-\lambda^{-1}\ln\big[1-F(x;a,b,c,\lambda,\underline{\varphi})+ +F(x;a,b,c,\lambda,\underline{\varphi})\exp(-\lambda)\big],a,b),\underline{\varphi}\Big\}^{1/c}\Big]. \tag{71}$$

Let $X_{(i)}$ be the i-th order statistic, and let $p_i = \frac{i}{n+1}$. The estimates for a, b, c, λ, and $\underline{\varphi}$ are found by minimizing the following expression with respect to a, b, c, λ, and $\underline{\varphi}$

$$\sum_{i=1}^{n} \Big\{x_{(i)} - G^{-1}\Big[\big\{\texttt{qbeta}(-\lambda^{-1}\ln[1-F(x;a,b,c,\lambda,\underline{\varphi})+ +F(x;a,b,c,\lambda,\underline{\varphi})\exp(-\lambda)],a,b),\underline{\varphi}\big\}^{1/c}\Big]\Big\}^2. \tag{72}$$

5. Applications

In this Section we consider some applications of the McGP family to model real data sets and illustrate their flexibility. We will include three different types of examples to illustrate the usefulness and effectiveness of this family. The measures of goodness of fit including the Akaike Information Criterion (AIC), Consistent Akaike Information Criterion (CAIC), Bayesian Information Criterion (BIC), Hannan-Quinn Information Criterion (HQIC), Anderson-Darling (A^*) Cramér-von Mises (W^*), and KS statistic are computed to compare the fitted models. One can employ the Likelihood Ratio Test (LRT) to contrast the adaptability of McGP distribution over the other distributions. The required computations are carried out in the `R` programming language using the `AdequacyModel` script of the `R`-package by Marinho *et al.* (2016).

Example 3. *The purpose of this Example is to illustrate that an additional parameter will produce significantly better results than its sub-model. It is of continued interest to generalize a probability distribution by adding different parameters. In this Example we consider the failure times of* 50 *components (per* 1000*h) to see how the McWP distribution is used to model failure times. This is an Example that shows how the additional parameter can help to improve the fitting of the data under study. This data set has been used by several authors including Murthy* et al. *(2004), Khan* et al. *(2019), among others, to illustrate the goodness of other competing models.*

We shall compare the fits of the proposed McWP distribution with those of other competitive models, namely: Beta Weibull Poisson (BWP) by Aryal *et al.* (2019), Exponentiated Generalized Weibull Poisson (EGWP) by Aryal and Yousof (2017). The estimated values of the parameters for the McWP, BWP and EGWP distributions for this data are given in Table 2.

TABLE 2
MLEs and their standard errors (in parenthesis) for component failure data.

Model	Estimates					
	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\lambda}$	$\widehat{a}$	$\widehat{b}$	$\widehat{c}$
McWP	1.463	1.161	0.452	0.683	0.078	7.413
	(0.002)	(0.003)	(1.290)	(0.168)	(0.024)	(2.974)
BWP	0.096	5.454	2.768	0.106	0.206	–
	(0.017)	(0.041)	(1.671)	(0.016)	(0.206)	–
EGWP	0.131	5.167	0.031	0.110	2.031	–
	(0.029)	(0.073)	(0.052)	(0.018)	(1.113)	–

The statistics of the fitted models are presented in Table 3. We note from Table 3 that the McWP gives the lowest values of the AIC, BIC, A*, and W* as compared to the other generalizations of the Weibull distribution. Therefore, the McWP distribution provides the best fit among the competing models for the failure time data.

TABLE 3
The AIC, CAIC, HQIC, BIC, W, A*, K-S statistics, and p-value for failure time data.*

Model	Goodness of fit criteria								
	$-\ell$	AIC	CAIC	BIC	HQIC	W*	A*	K-S	p-value
McWP	97.89	207.78	209.74	219.26	212.15	0.07	0.49	0.08	0.89
BWP	100.01	210.03	211.39	219.59	213.67	0.11	0.76	0.12	0.39
EGWP	100.58	211.16	212.52	220.72	214.80	0.12	0.80	0.13	0.35

One can compute the maximized unrestricted and restricted log-likelihood functions to construct the likelihood ratio (LR) test statistic for testing the models. For example, to test whether McWP is a significantly better fit than BWP, we perform the following test.

$$H_0 \;:\; \text{BWP distribution is appropriate}$$
$$H_a \;:\; \text{McWP distribution is appropriate}$$

The LR test statistic for testing H_0 versus H_a is

$$\omega = 2(\ell(\underline{\hat{\varphi}}, x) - \ell(\underline{\hat{\varphi}_0}, x)), \tag{73}$$

where $\underline{\hat{\varphi}}$ and $\underline{\hat{\varphi}_0}$ are the MLEs under H_a and H_0, respectively. The statistic ω is asymptotically distributed as χ^2_k, where k is the length of the parameter vector of interest. Note that for the subject data $\omega = 4.24$ with p-value $= 0.0395$. Therefore, at 0.05 level of significance, the McWP is superior to the BWP for the subject data.

Plots comparing the McWP distribution with the BWP distribution and the EGWP distribution for this data are displayed in Figure 7. It is evident that the McWP fits better than the other competitive distributions.

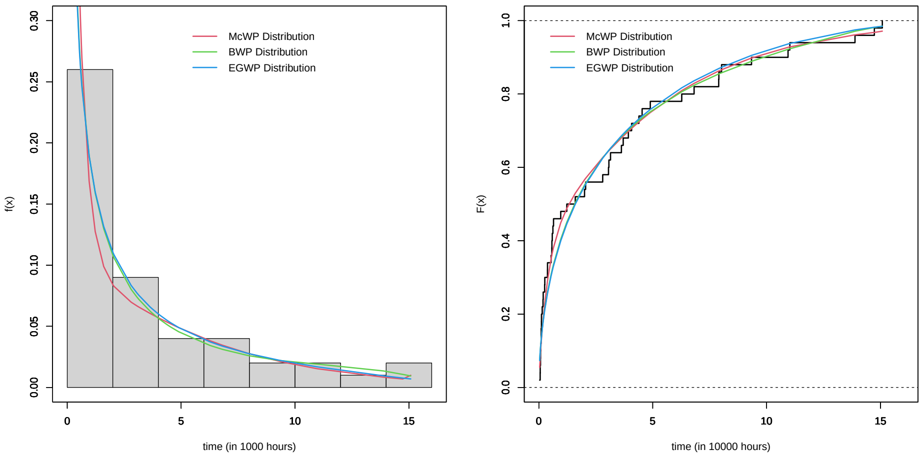

Figure 7 – Fitted PDF (left panel) and CDF (right panel) of McWP, BWP, and EGWP distribution for repair time data.

EXAMPLE 4. *The purpose of this Example is to illustrate that McWP can be used to model count data. The Poisson distribution is often used to model rare events. Major earthquakes (seismic intensity of* 7.0 *and over in rector scale) are considered to be rare. Assuming that earthquakes are independent, we can forecast the number of earthquakes per year using the Poisson distribution. In Table 4 the numbers of major earthquakes from the United States Geological Survey (*`https://www.usgs.gov`*) between 1900 and 2018 are reported.*

TABLE 4
Number of major earthquakes from the United States Geological Survey between 1900 and 2018.

13	14	8	10	16	26	32	27	18	32	36	24	22	23	22	18	25
21	21	14	8	11	14	23	18	17	19	20	22	19	13	26	13	14
22	24	21	22	26	21	23	24	27	41	31	27	35	26	28	36	39
21	17	22	17	19	15	34	10	15	22	18	15	20	15	22	19	16
30	27	29	23	20	16	21	21	25	16	18	15	18	14	10	15	8
15	6	11	8	7	18	16	13	12	13	20	15	16	12	18	15	16
13	15	16	11	11	18	12	17	24	20	16	19	12	19	16	7	17

The estimated maximum likelihood estimate (MLE) and standard error (in parenthesis) of the parameter of the Poisson distribution using the earthquake data is 19.06 (0.40) with AIC value of 860.54. Similarly, the MLEs and their standard errors (in parenthesis) of the McDonald's Weibull Poisson Distribution are given by $\hat{\alpha} = 0.05(0.03)$, $\hat{\beta} = 2.19(0.67)$, $\hat{\lambda} = 2.20(1.98)$, $\hat{a} = 2.11(1.39)$, $\hat{b} = 0.94(0.78)$ and $\hat{c} = 0.66(5.64)$ with an AIC value of 800.10. Further, the Kolmogorov-Smirnov test statistic (D) is 0.05 with p-value equal to 0.87. It is evident that the earthquake data can be modeled very well using the McWP distribution.

In Figures 8 and 9, we present a comparison of empirical and theoretical PDF and CDF of the Poisson distribution and the McWP distribution, respectively. We observe that the continuous approximation to the McWP of this discrete phenomenon depicts a better fit and low mean squared error. One can perform the likelihood ratio test as in Example 3 to prove that the McWP fits the earthquake data better than other competing models including the Poisson distribution.

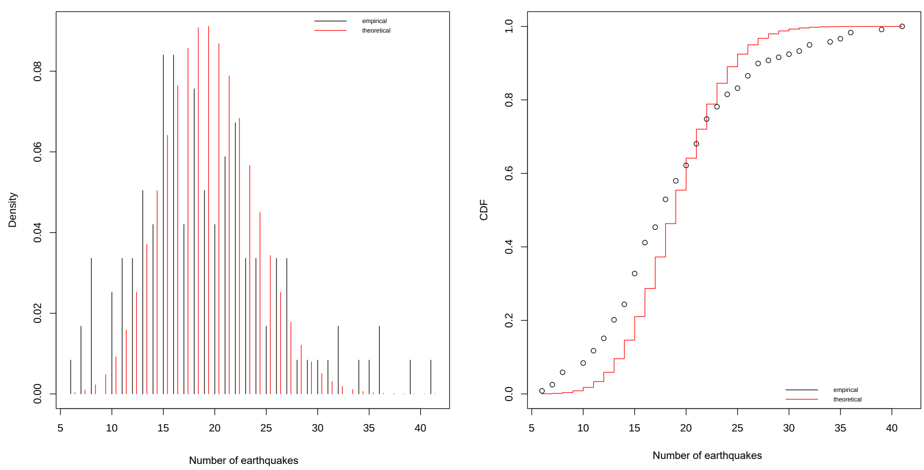

Figure 8 – Fitted Poisson distribution for earthquake data.

EXAMPLE 5. *The purpose of this Example is to illustrate that having a higher number of parameters in a given distribution does not assure a better fit for given data. In this example we consider a data set that fits the McWP distribution very well. However, almost similar goodness of fit could be achieved using a sub-model of the McWP, namely the BWP. Even though the addition of a parameter makes the McWP more flexible, it may not produce significantly different results than the BWP. Hence, having a higher number of parameters in a given distribution does not assure a better fit for given data. The data represents the waiting times (in minutes) before service of* 100 *bank customers. Readers are referred to*

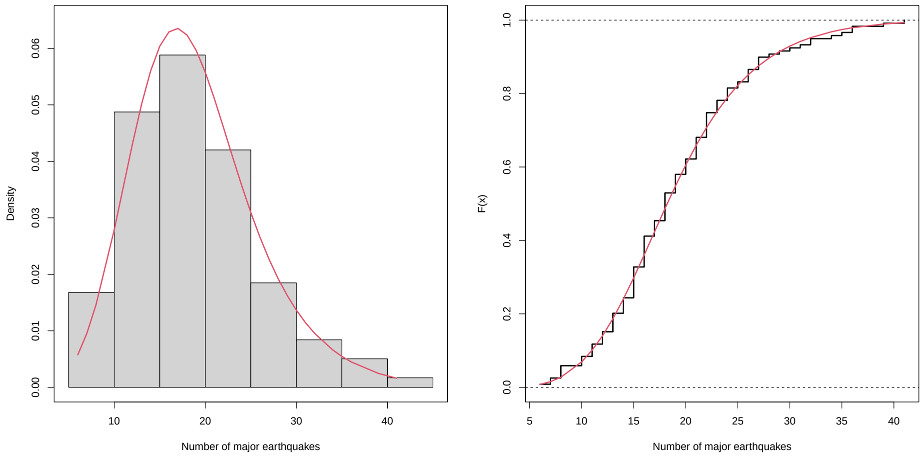

Figure 9 – The McWP distribution for earthquake data.

Ghitany et al. (2008) for details about this study. This data set has been studied by many authors including Ghitany et al. (2008), Al-Mutairi et al. (2013), among others.

The estimated values of the parameters for the McWP and BWP distributions using this data are given in the Table 5.

TABLE 5
MLEs and their standard errors (in parenthesis) for waiting time data.

Model	Estimates					
	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\lambda}$	$\widehat{a}$	$\widehat{b}$	$\widehat{c}$
McWP	0.263	1.147	0.379	2.046	0.375	1.326
	(0.738)	(2.076)	(7.008)	(4.390)	(2.074)	(8.384)
BWP	0.361	1.272	1.806	0.186	0.489	–
	(0.672)	(2.118)	(4.489)	(0.969)	(9.274)	–

To compare the models we compute the values of test statistics. The statistics of the fitted McWP and BWP are presented in the Table 6.

TABLE 6
The AIC, CAIC, HQIC, BIC, W^*, A^*, K-S statistics and p-value for failure time data.

Model	Goodness of fit criteria								
	$-\ell$	AIC	CAIC	BIC	HQIC	W^*	A^*	K-S	p-value
McWP	317.04	646.08	646.98	661.71	652.40	0.02	0.12	0.04	0.99
BWP	317.04	644.07	644.71	657.10	649.34	0.02	0.13	0.04	0.99

6. Conclusions

In this paper we have studied a new class of distributions that is being referred to as the McDonald generalized Poisson (McGP) family. The McGP family is defined by using the genesis of the zero truncated Poisson distribution and the McDonald distribution. Many mathematical and statistical properties and special cases of the McGP have been explored. The parameter estimation procedure using different methods has been discussed. We have analyzed the McGP family with a number of examples to illustrate that it is quite flexible to model real-life data. However, it has been observed that fitting of a probability distribution to a specific data is a local phenomenon. That is, a particular generalization might not end up with a better fit just because it possesses a larger number of parameters. In summary, probability distributions with a larger number of parameters are worth exploring to analyze a given data set. Nevertheless, specific data might be fitted by the base distribution as good as by the generalized distribution.

Acknowledgements

The authors are grateful to the editor and anonymous reviewers for their constructive comments and valuable suggestions which certainly improved the presentation and quality of the paper. The authors also would like to thank the Association of Nepalese Mathematicians in America (ANMA) for organizing the Workshop on Collaborative Research in Mathematical Sciences where the authors met and initiated this project.

References

Z. Ahmad, G. Hamedani, N. S. Butt (2019). *Recent developments in distribution theory: A brief survey and some new generalized classes of distributions*. Pakistan Journal of Statistics and Operation Research, 15, no. 1, pp. 87–110.

D. K. Al-Mutairi, M. E. Ghitany, D. Kundu (2013). *Inferences on stress-strength reliability from Lindley distributions*. Communications in Statistics - Theory and Methods, 42, no. 8, pp. 1443–1463.

G. R. Aryal, S. B. Chhetri, H. Long, A. A. Akinsete (2019). *On the beta-G Poisson family*. Annals of Data Science, 6, no. 3, pp. 361–389.

G. R. ARYAL, H. M. YOUSOF (2017). *The exponentiated generalized-G poisson family of distributions*. Stochastics and Quality Control, 32, no. 1, pp. 7–23.

R. C. CHENG, N. A. K. AMIN (1979). *Product of spacings estimation with applications to the lognormal distribution*. University of Wales IST, Math Report 79.

R. C. CHENG, N. A. K. AMIN (1983). *Estimating parameters in continuous univariate distributions with a shifted origin*. Journal of the Royal Statistical Society, Series B, 45, pp. 394–403.

F. N. DAVID, N. L. JOHNSON (1952). *The truncated Poisson*. Biometrics, 8, pp. 275–285.

M. E. GHITANY, B. ATIEH, S. NADARAJAH (2008). *Lindley distribution and its application*. Mathematics and Computers in Simulation, 78, pp. 493–506.

I. S. GRADSHTEYN, I. M. RYZHIK (2000). *Table of Integrals, Series, and Products*. 6th Edition, Academic Press, San Diego.

M. S. KHAN, R. KING, I. L. HUDSON (2019). *Transmuted exponentiated Weibull distribution*. Journal of Applied Probability and Statistics, 14, no. 2, pp. 37–51.

P. MARINHO, M. BOURGUIGNON, C. R. B. DIAS (2016). *Adequacy of Probabilistic Models and Generation of Pseudo-Random Numbers, R package-AdequacyModel*. URL http://cran.r-project.org/web/packages/AdequacyModel/AdequacyModel.pdf.

A. MARSHALL, I. OLKIN (2007). *Life Distributions. Structure of Nonparametric, Semiparametric and Parametric Families*. Springer-Verlag, New York.

J. B. MCDONALD (1984). *Some generalized functions the size distribution of income*. Econometrica, 52, no. 3, pp. 647–663.

D. MURTHY, M. XIE, R. JIANG (2004). *Weibull Models*. John Wiley & Sons, New York.

M. TAHIR, G. CORDEIRO (2016). *Compounding of distributions: a survey and new generalized classes*. Journal of Statistical Distributions and Applications, 3, no. 13, pp. 1–35.