International Journal of Applied Mathematics

Volume 37 No. 6 2024, 627–648
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
doi: http://dx.doi.org/10.12732/ijam.v37i6.5

A NEW BIMODAL DOUBLE DISTRIBUTION ON THE REAL LINE AND ITS APPLICATION

Hajar M. Alkhezi [1], M. E. Ghitany [1,§], Mai F. Alfahad [1], J. Mazucheli [2]

[1] Department of Statistics and Operations Research

Faculty of Science, Kuwait University, KUWAIT

[§] e-mail: me.ghitany@ku.edu.kw

[2] Department of Statistics

Universidade Estadual de Maringá

Maringá, PR, BRAZIL

Abstract

In this paper, we propose a new bimodal double distribution on the real line using random sign mixture transform and study its associated statistical inferences. Maximum likelihood estimation is used to estimate the underlying parameters. Monte Carlo simulation experiments are carried out to examine the performance of the estimators and the corresponding confidence intervals of the parameters. The proposed distribution is fitted to a bimodal real data set and is compared with other recently published bimodal double distributions.

MSC 2020: 62E15, 62F12

Key Words and Phrases: random sign mixture transform; bimodality; maximum likelihood estimation; simulations

Received: June 15, 2024

1. Introduction

The Chaudhry-Ahmad (CA) distribution is a two-parameter continuous probability distribution defined on the positive real line was introduced by [7]. Its probability density function (PDF) is

$$f_X(x;\alpha,\lambda) = 2\sqrt{\frac{\alpha}{\pi}}\ \exp\left[-\alpha\left(x-\frac{\lambda^2}{x}\right)^2\right], \quad x>0, \quad \alpha,\lambda>0. \tag{1}$$

The CA distribution can be obtained as the root reciprocal of the inverse-Gaussian (IG) distribution, that is, the distribution of the random variable (RV) $X=\frac{1}{\sqrt{W}}$ where W follows the IG distribution with PDF

$$f_W(w;\mu,\nu) = \sqrt{\frac{\nu}{2\pi}}\ w^{-3/2}\ \exp\left[-\frac{\nu(w-\mu)^2}{2\mu^2 w}\right], \quad w>0, \quad \mu,\nu>0 \tag{2}$$

where $\mu=\lambda^{-2}$ and $\nu=2\alpha$.

Recently, [11] presented basic properties of the CA distribution and fitted it to wind speed data from six weather stations distributed in the state of Tocantins in Brazil.

To extend its range of applications to data on the whole real line, for example, weather temperatures, stock returns and DNA microarray data, we introduce a double Chaudhry-Ahmad (DCA) distribution defined on the whole real line.

We follow the procedure presented by [3] to construct a DCA distribution using the random sign mixture transform (RSMT) given by

$$Z = YX_1-(1-Y)X_2, \tag{3}$$

where Y is a Bernoulli random variable (RV) with parameter β, X_1 and X_2 are non-negative RVs independent of Y. If X_1 and X_2 are independent and identically distributed (IID), we have what is called random sign transform (RST).

If X_1, X_2 are independent RVs from the sme family of distributions $\mathcal{F}$, then the distribution of Z is said to have *double* $\mathcal{F}$ distribution.

In the literature, many authors use the word double as the distribution of the absolute value and some of them use the word reflection. For example, [4] and [8] presented the double Weibull distribution. [9] studied the reflected version of the exponential distribution. The reflected version of the generalized Gamma was studied by [12]. In fact, these double distributions are the distributions of the RST when Y has a Bernoulli distribution with pararameter $\beta=0.5$.

Recently, [2] and [1] introduced double inverse-Gaussian and double log-normal distributions, respectively, using RSMT. These papers also reviewed previous publications in the literature using the RST/RSMT approach.

In our construction of the DCA distribution using RSMT given by equation (3), it is assumed that Y is a Bernoulli RV with parameter β, X_1 and X_2 are CA RVs independent of Y.

The contents of this paper are organized as follows. The main statistical properties of the DCA distribution are presented in Section 2. Section 3 gives explicit MLEs and their asymptotic distributions. Simulation studies are carried out to study the performance of the MLEs in Section 4. In Section 5, the proposed DCA distribution is fitted to a DNA microarray data set and compared with other recent published bimodal double distributions. Finally, conclusions and comments are presented in Section 6.

2. **Statistical properties**

In this section, we present the main statistical properties of the DCA distribution.

2.1. **Probability density function.** The PDF of DCA distribution is

$$f_Z(z) \;=\; \begin{cases} \overline{\beta}\; f_{X_2}(|z|;\alpha_2,\lambda_2), & z<0,\\ \beta\;\; f_{X_1}(z;\alpha_1,\lambda_1), & z\geq 0,\end{cases} \tag{4}$$

where, for $\alpha_j, \lambda_j > 0,\; j = 1, 2$,

$$f_{X_j}(x;\alpha_j,\lambda_j) = 2\sqrt{\frac{\alpha_j}{\pi}}\; \exp\left[-\alpha_j\left(x-\frac{\lambda_j^2}{x}\right)^2\right], \quad x>0, \tag{5}$$

are the PDFs of the CA distributions.

The DCA distribution has two modes given by

$$\text{Mode}(Z) = -\,\text{Mode}(X_2) \;\text{ and }\; \text{Mode}(X_1), \tag{6}$$

where, for $j = 1, 2$,

$$\text{Mode}(X_j) = \lambda_j, \tag{7}$$

are the modes of the CA distributions.

2.2. **Cumulative distribution function.** The cumulative distribution function (CDF) of DCA distribution is is given by

$$F_Z(z) = P(Z\leq z) \;=\; \begin{cases} \overline{\beta}\;\left[1-F_{X_2}(|z|;\alpha_2,\lambda_2)\right], & z<0,\\ \overline{\beta}+\beta\; F_{X_1}(z;\alpha_1,\lambda_1), & z\geq 0,\end{cases} \tag{8}$$

where, for $j = 1, 2$,

$$\begin{aligned} F_{X_j}(x;\alpha_j,\lambda_j) &= P(X_j \le x) \\ &= \Phi\left(\sqrt{2\alpha_j}\left(x - \frac{\lambda_j^2}{x}\right)\right) \\ &\quad - e^{4\alpha_j\lambda_j^2}\,\Phi\left(-\sqrt{2\alpha_j}\left(x + \frac{\lambda_j^2}{x}\right)\right),\ x > 0, \end{aligned} \tag{9}$$

are the CDFs of the CA distributions and

$$\Phi(a) = P(Z \le a) = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}}\, e^{-z^2/2}\, dz, \quad a \in \mathbb{R}, \tag{10}$$

is the CDF of the standard normal distribution.

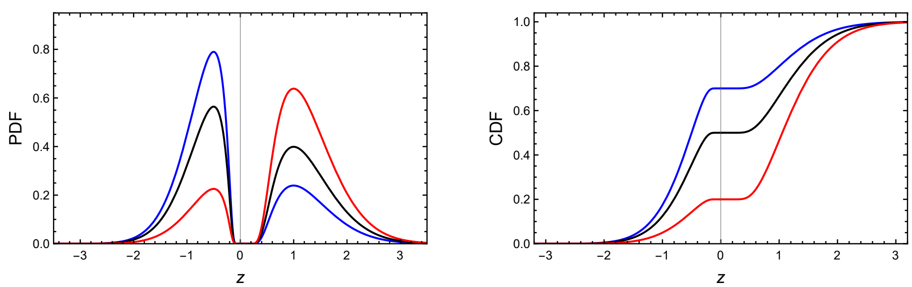

FIGURE 1. PDF and CDF of DCA distribution: $(\beta, \alpha_1, \lambda_1, \alpha_2, \lambda_2)$: $(0.3, 0.5, 1, 1, 0.5)$(blue), $(0.5, 0.5, 1, 1, 0.5)$(black), $(0.8, 0.5, 1, 1, 0.5)$(red).

Figure 1 shows the the PDF and CDF of the DCA distribution for selected values of the parameters.

2.3. **Moments and associated measures.** The rth raw moment of DCA distribution is given by

$$E(Z^r) = \beta\ E(X_1^r) + (-1)^r\ \overline{\beta}\ E(X_2^r), \qquad r \ge 1, \tag{11}$$

where, for $j = 1, 2$,

$$\begin{aligned} E(X_j^r) &= \int_0^\infty x^r\ f_{X_j}(x;\alpha_j,\lambda_j)dx \\ &= 2\sqrt{\frac{\alpha_j}{\pi}}\ \lambda_j^{r+1}\ e^{2\alpha_j\lambda_j^2}\ K_{\frac{r+1}{2}}(2\alpha_j\lambda_j^2), \end{aligned} \tag{12}$$

are the rth moments of the CA distributions and

$$K_\nu(c) = \frac{1}{2}\left(\frac{c}{2}\right)^\nu \int_0^\infty \frac{1}{t^{\nu+1}}\ \exp\left[-t - \frac{c^2}{4t}\right]\ dt, \quad c > 0,\ \nu \in \mathbb{R},$$

is the modified Bessel function of the second kind.

The mean, variance, skewness and kurtosis of DCA distribution can be obtained using these raw moments.

Figure 2 shows the mean, variance, skewness, and kurtosis of the DCA distribution as a function in β for selected values of the parameters $(\alpha_1, \lambda_1, \alpha_2, \lambda_2)$. Also, this figure shows that the skewness can be negative/positive, i.e. the DCA distribution can be skewed to the left/right.

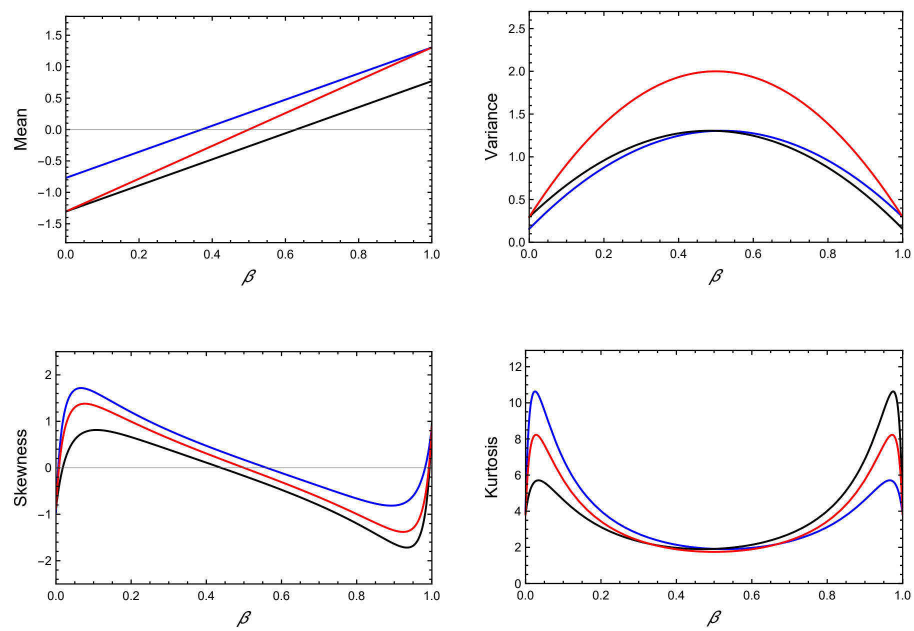

FIGURE 2. Mean, variance, skewness and kurtosis of DCA distribution as a function in β: $(\alpha_1, \lambda_1, \alpha_2, \lambda_2)$: $(0.5, 1, 1, 0.5)$(——), $(1, 0.5, 0.5, 1)$(——), $(0.5, 1, 0.5, 1)$(——).

2.4. **Harmonic mean.**

The harmonic mean of a RV V, is defined as $HM(V) = \frac{1}{E[1/V]}$, provided $E[1/V]$ exists.

LEMMA 2.1. *The harmonic mean of DCA distribution is given by*

$$HM(Z) = \frac{1}{\frac{\beta}{HM(X_1)} - \frac{\overline{\beta}}{HM(X_2)}} \tag{13}$$

where

$$HM(X_j) = \frac{1}{2\sqrt{\frac{\alpha_j}{\pi}}\ e^{2\alpha_j\lambda_j^2}\ K_0(2\alpha_j\lambda_j^2)}, \quad j = 1, 2, \tag{14}$$

are the harmonic means of the CA distributions.

P r o o f. Equation (13) is given in [1]. For a RV $X_j \sim CA(\alpha_j, \lambda_j)$ distribution, we have

$$\begin{aligned} \frac{1}{HM(X_j)} &= \int_0^\infty \frac{1}{x}\, f_{X_j}(x, \alpha_j, \lambda_j) dx \\ &= 2\sqrt{\frac{\alpha_j}{\pi}}\; e^{2\alpha\lambda_j^2}\; K_0(2\alpha_j\lambda_j^2). \end{aligned}$$

This completes the proof of the lemma. □

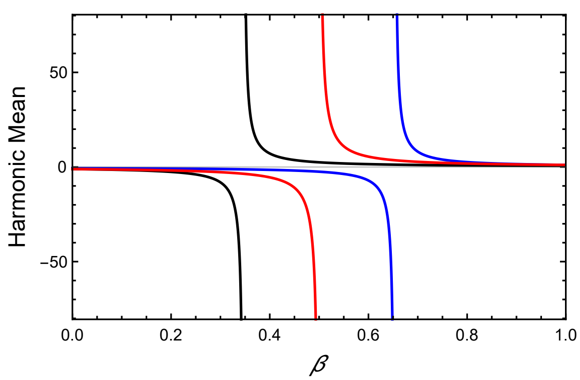

FIGURE 3. Harmonic mean of DCA distribution as a function in β: $(\alpha_1, \lambda_1, \alpha_2, \lambda_2) : (0.5, 1, 1, 0.5)$(——), $(1, 0.5, 0.5, 1)$(——), $(0.5, 1, 0.5, 1)$(——).

Figure 3 shows the harmonic mean of the DCA distribution as a function in β for selected values of the parameters.

2.5. **Entropies.** Entropies are measures of a system's variation, instability, or uncertainty.

For a RV V with PDF $f_V(v)$, the following are two well known entropies:

1. Tsallis entropy: (see [15])

$$T_s(V) = \frac{1}{s-1}\{1 - E[f_V^{s-1}(V)]\}, \quad 0 < s \neq 1.$$

2. Shannon entropy: (see [14])

$$H(V) = E[-\ln f_V(V)]\} = \lim_{s\to 1} T_s(V).$$

LEMMA 2.2. *Tsallis entropy of DCA distribution is given by*

$$T_s(Z) = T_s(Y) + \beta^s\ T_s(X_1) + \overline{\beta}^{\,s}\ T_s(X_2), \tag{15}$$

where

$$T_s(Y) = \frac{1 - \beta^s - \overline{\beta}^{\,s}}{s-1}, \tag{16}$$

is Tsallis entropy of Bernoulli distribution and

$$T_s(X_j) = \frac{1}{s-1}\left\{1 - \frac{1}{\sqrt{s}}\ \left(2\sqrt{\frac{\alpha_j}{\pi}}\right)^{s-1}\right\}, \qquad j = 1, 2, \tag{17}$$

are Tsallis entropies of CA distributions.

P r o o f. Equations (15) and (16) are given in [2]. For a RV $X_j \sim CA(\alpha_j, \lambda_j)$ distribution, we have

$$\begin{aligned} 1 - (s-1)T_s(X_j) &= \int_0^\infty f_X^s(x; \alpha_j, \lambda_j) dx \\ &= \int_0^\infty 2^s \left(\frac{\alpha_j}{\pi}\right)^{s/2} \exp\left[-s\ \alpha_j \left(x - \frac{\lambda_j^2}{x}\right)^2\right] dx \\ &= \frac{1}{\sqrt{s}} \left(2\sqrt{\frac{\alpha_j}{\pi}}\right)^{s-1}. \end{aligned}$$

This completes the proof of the lemma. □

LEMMA 2.3. *Shannon entropy of DCA distribution is given by*

$$H(Z) = H(Y) + \beta\ H(X_1) + \overline{\beta}\ H(X_2), \tag{18}$$

where

$$H(Y) = -\beta\ \ln(\beta) - \overline{\beta}\ \ln(\overline{\beta}), \tag{19}$$

is Shannon entropy of Bernoulli distribution and

$$H(X_j) = \frac{1}{2} - \ln\left(2\sqrt{\frac{\alpha_j}{\pi}}\right), \qquad j = 1, 2, \tag{20}$$

are Shannon entropies of CA distributions.

P r o o f. Equations (18) and (19) are given in [3]. For a RV $X_j \sim CA(\alpha_j, \lambda_j)$ distribution, using L'Hôpital's rule, we have

$$\begin{aligned} H(X_j) &= \lim_{s\to 1} T_s(X_j) \\ &= \lim_{s\to 1} -\left(2\sqrt{\frac{\alpha_j}{\pi}}\right)^{s-1} \left\{\frac{1}{\sqrt{s}}\ln\left(2\sqrt{\frac{\alpha_j}{\pi}}\right) - \frac{1}{2s^{3/2}}\right\} \\ &= \frac{1}{2} - \ln\left(2\sqrt{\frac{\alpha_j}{\pi}}\right). \end{aligned}$$

This completes the proof of the lemma. □

Figure 4 shows Tsallis and Shannon entropies of DCA distribution as a function in β for selected values of the parameters. This figure also shows that Tsallis and Shannon entropies of DCA distribution can be negative/positive.

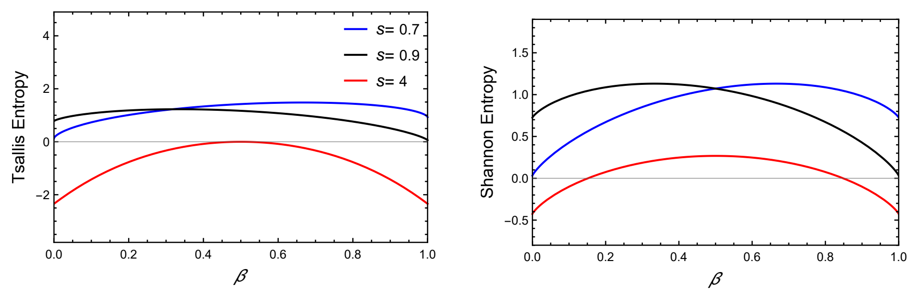

FIGURE 4. Tsallis and Shannon entropies of DCA distribution as a function in β: $(\alpha_1, \lambda_1, \alpha_2, \lambda_2)$: $(0.5, \lambda_1, 2, \lambda_2)$(——), $(2, \lambda_1, 0.5, \lambda_2)$(——), $(5, \lambda_1, 5, \lambda_2)$(——), for all $\lambda_1, \lambda_2 > 0$.

2.6. **Extropies.** Recently, alternative measures of uncertainty, called extropy and weighted extropy, are proposed in the literature.

For a RV V with PDF $f_V(v)$, the following are two well known extropies:

1. Extropy: (see [10])

$$J(V) = -\frac{1}{2}E[f_V(V)]\} = \frac{1}{2}[T_2(V) - 1].$$

2. Weighted extropy: (see [5])

$$J^w(V) = -\frac{1}{2}E[V f_V(V)]\}.$$

The authors of [5] gave example of two RVs having the same extropy but different weighted extropy.

Lemma 2.4. *The extropy of DCA distribution is given by*

$$J(Z) = \beta^2 \ J(X_1) + \overline{\beta}^2 \ J(X_2), \tag{21}$$

where, for $j = 1, 2$,

$$J(X_j) = -\sqrt{\frac{\alpha_j}{2\pi}}, \tag{22}$$

are extropies of CA distributions.

P r o o f. Equations (21) is given in [2]. For a RV $X_j \sim CA(\alpha_j, \lambda_j)$ distribution, we have

$$J(X_j) = \frac{1}{2}[T_2(X_j) - 1] = -\sqrt{\frac{\alpha_j}{2\pi}}.$$

This completes the proof of the lemma. □

Lemma 2.5. *The weigted extropy of DCA distribution is given by*

$$J^w(Z) = \beta^2 \ J^w(X_1) - \overline{\beta}^2 \ J^w(X_2), \tag{23}$$

where, for $j = 1, 2$,

$$J^w(X_j) = -\frac{2}{\pi} \ \alpha_j \lambda_j^2 \ e^{4\alpha_j \lambda_j^2} \ K_1(4\alpha_j \lambda_j^2), \tag{24}$$

are the weighted extropies of CA distributions.

P r o o f. Since

$$\begin{aligned} J^w(Z) &= -\frac{1}{2} \int_{-\infty}^{\infty} z \ f_Z^2(z) dz \\ &= -\frac{1}{2} \left\{ \beta^2 \int_0^{\infty} z \ f_{X_1}^2(z) dz + \overline{\beta}^2 \int_{-\infty}^{0} z \ f_{X_2}^2(|z|) dz \right\} \\ &= -\frac{1}{2} \left\{ \beta^2 \int_0^{\infty} z \ f_{X_1}^2(z) dz - \overline{\beta}^2 \int_0^{\infty} x \ f_{X_2}^2(x) dx \right\} \\ &= \beta^2 \ J^w(X_1) - \overline{\beta}^2 \ J^w(X_2), \end{aligned}$$

equation (23) follows. For a RV $X_j \sim CA(\alpha_j, \lambda_j)$ distribution, we have

$$\begin{aligned} J^w(X_j) &= -\frac{1}{2} \int_0^{\infty} x \ f_{X_j}^2(x; \alpha_j, \lambda_j) dx \\ &= -\frac{2}{\pi} \ \alpha_j \lambda_j^2 \ e^{4\alpha_j \lambda_j^2} \ K_1(4\alpha_j \lambda_j^2). \end{aligned}$$

This completes the proof of the lemma. □

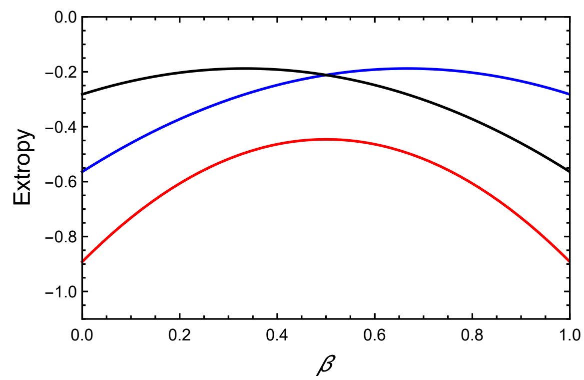

FIGURE 5. Extropy of DCA distribution as a function in β: $(\alpha_1, \lambda_1, \alpha_2, \lambda_2)$: $(0.5, \lambda_1, 2, \lambda_2)$(——), $(2, \lambda_1, 0.5, \lambda_2)$(——), $(5, \lambda_1, 5, \lambda_2)$(——), for all $\lambda_1, \lambda_2 > 0$.

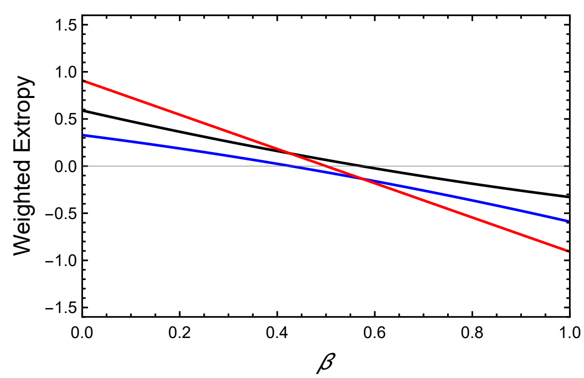

FIGURE 6. Weighted extropy of DCA distribution as a function in β: $(\alpha_1, \lambda_1, \alpha_2, \lambda_2)$: $(0.5, 2, 2, 0.5)$(——), $(2, 0.5, 0.5, 1)$(——), $(5, 1, 5, 1)$(——).

Figure 2.6 shows the extropy of DCA distribution as a function in β for selected values of the parameters. This figure also shows that the extropy of DCA distribution is always negative. On the other hand, Figure 2.6 shows that the weighted extropy of DCA distribution can be positive/negative.

3. Maximum likelihood estimation

In this section, we derive the MLEs of the parameters of DCA distribution and their asymptotic distributions.

Let $z_1, z_2, \ldots, z_n$ be a random sample from DCA$(\beta, \alpha_1, \lambda_1, \alpha_2, \lambda_2)$ distribution.The log-likelihood function is given by

$$\begin{aligned} \ln L(\beta, \alpha_1, \lambda_1, \alpha_2, \lambda_2) \;=\; & \sum_{i=1}^{n} \ln[\beta\ f_{X_1}(z_i; \alpha_1, \lambda_1)]\, \nVdash_{\{z_i>0\}} \\ & + \sum_{i=1}^{n} \ln[\overline{\beta}\ f_{X_2}(|z_i|; \alpha_2, \lambda_2)]\, \nVdash_{\{z_i<0\}}, \end{aligned} \tag{25}$$

where $\nVdash_A = 1(0)$ if A is true (false) is the indicator function.

The MLEs of the parameters $(\beta, \alpha_1, \lambda_1, \alpha_2, \lambda_2)$ are:

$$\widehat{\beta} \;=\; n_1/n, \tag{26}$$

$$\widehat{\alpha}_1 \;=\; a_1/\left(2(a_1 b_1 - 1)\right), \qquad \widehat{\lambda}_1 = 1/\sqrt{a_1}, \tag{27}$$

$$\widehat{\alpha}_2 \;=\; a_2/\left(2(a_2 b_2 - 1)\right), \qquad \widehat{\lambda}_2 = 1/\sqrt{a_2} \tag{28}$$

where

$$n_1 = \sum_{i=1}^{n} \nVdash_{(z_i>0)}, \qquad n_2 = \sum_{i=1}^{n} \nVdash_{(z_i<0)}, \qquad n_1 + n_2 = n, \tag{29}$$

$$a_1 = \frac{1}{n_1} \sum_{i=1}^{n} z_i^{-2}\, \nVdash_{(z_i>0)}, \qquad b_1 = \frac{1}{n_1} \sum_{i=1}^{n} z_i^{2}\, \nVdash_{(z_i>0)}, \tag{30}$$

$$a_2 = \frac{1}{n_2} \sum_{i=1}^{n} z_i^{-2}\, \nVdash_{(z_i<0)}, \qquad b_2 = \frac{1}{n_2} \sum_{i=1}^{n} z_i^{2}\, \nVdash_{(z_i<0)}. \tag{31}$$

The Fisher information matrix about $(\beta, \alpha_1, \lambda_1, \alpha_2, \lambda_2)$ is given by

$$\mathbf{I}(\beta, \alpha_1, \lambda_1, \alpha_2, \lambda_2) = \operatorname{diag}\left(I_Y(\beta), \beta\ \mathbf{I}_{X_1}(\alpha_1, \lambda_1), \overline{\beta}\ \mathbf{I}_{X_2}(\alpha_2, \lambda_2)\right), \tag{32}$$

where $I_Y(\beta) = \frac{1}{\beta\, \overline{\beta}}$ is the Fisher information about β and, for $j = 1, 2$,

$$\mathbf{I}_{X_j}(\alpha_j, \lambda_j) = \operatorname{diag}\left(\frac{1}{2\alpha_j^2}, 8\alpha_j\right), \tag{33}$$

are the Fisher information matrices about $(\alpha_j,\ \lambda_j), j = 1, 2$.

Moreover, the asymptotic distribution of the MLEs is given by:

As $n \to \infty$,

$$\sqrt{n}\begin{bmatrix}\widehat{\beta}-\beta\\ \widehat{\alpha}_1-\alpha_1\\ \widehat{\lambda}_1-\lambda_1\\ \widehat{\alpha}_2-\alpha_2\\ \widehat{\lambda}_2-\lambda_2\end{bmatrix} \xrightarrow{d} MVN\left(\mathbf{0}, \mathbf{I}^{-1}(\beta,\alpha_1,\lambda_1,\alpha_2,\lambda_2)\right) \tag{34}$$

where $\xrightarrow{d}$ denotes *convergence in distribution*, MVN stands for *multivariate normal distribution* and

$$\mathbf{I}^{-1}(\beta,\alpha_1,\lambda_1,\alpha_2,\lambda_2) = \operatorname{diag}\left(\beta\,\overline{\beta}, \frac{2\alpha_1^2}{\beta}, \frac{1}{8\alpha_1\beta}, \frac{2\alpha_2^2}{\overline{\beta}}, \frac{1}{8\alpha_2\overline{\beta}}\right). \tag{35}$$

4. **Simulation study**

The purpose of this section is to perform simulation study to evaluate the finite-sample behaviour of the MLEs of the parameters of the proposed DCA distribution. The simulation study was done using the `R` language [13]. Such study was repeated $M = 10,000$ times. In each of the M repetitions, a random sample of size $n = 50, 100, \ldots, 500$ is drawn from DCA distribution with true parameters $(\beta, \alpha_1, \lambda_1, \alpha_2, \lambda_2) =$ (0.35, 2, 1, 1, 2), (0.5, 1, 2, 2, 1), (0.7, 0.5, 1, 1, 0.5), using the following algorithm:

1. Generate $Y_i \sim Bernoulli(\beta),\ i = 1, 2, \ldots, n$;
2. Generate $W_{1,i} \sim \mathrm{IG}(\lambda_1^{-2}, 2\alpha_1),\ i = 1, 2, \ldots, n$;
3. Set $X_{1,i} = \frac{1}{\sqrt{W_{1,i}}},\ i = 1, 2, \ldots, n$;
4. Generate $W_{2,i} \sim \mathrm{IG}(\lambda_2^{-2}, 2\alpha_2),\ i = 1, 2, \ldots, n$;
5. Set $X_{2,i} = \frac{1}{\sqrt{W_{2,i}}},\ i = 1, 2, \ldots, n$;
6. Set $Z_i = Y_i\, X_{1,i} - (1 - Y_i)\, X_{2,i},\ i = 1, 2, \ldots, n$.

The behaviours of the MLEs of the parameters of the proposed DCA distribution are evaluated in terms of the following measures:

1. (Average) Bias of the MLEs:

$$Bias(\widehat{\theta}) = \frac{1}{M}\sum_{j=1}^{M}(\widehat{\theta}_j - \theta), \qquad \theta = \beta, \alpha_1, \lambda_1, \alpha_2, \lambda_2,$$

where $\widehat{\theta}_j$ is the MLE of θ in the jth simulation repetition.

2. Mean square error (MSE) of the MLEs:

$$MSE(\widehat{\theta}) = \frac{1}{M}\sum_{j=1}^{M}(\widehat{\theta}_j - \theta)^2.$$

3. Coverage probability (CP) of the 95% confidence interval of each parameter:

$$CP(\theta) = \frac{1}{M}\sum_{j=1}^{M} \nVdash_{\{\widehat{\theta}_j - 1.96\ S.E.(\widehat{\theta}_j)\ <\theta<\ \widehat{\theta}_j + 1.96\ S.E.(\widehat{\theta}_j)\}}.$$

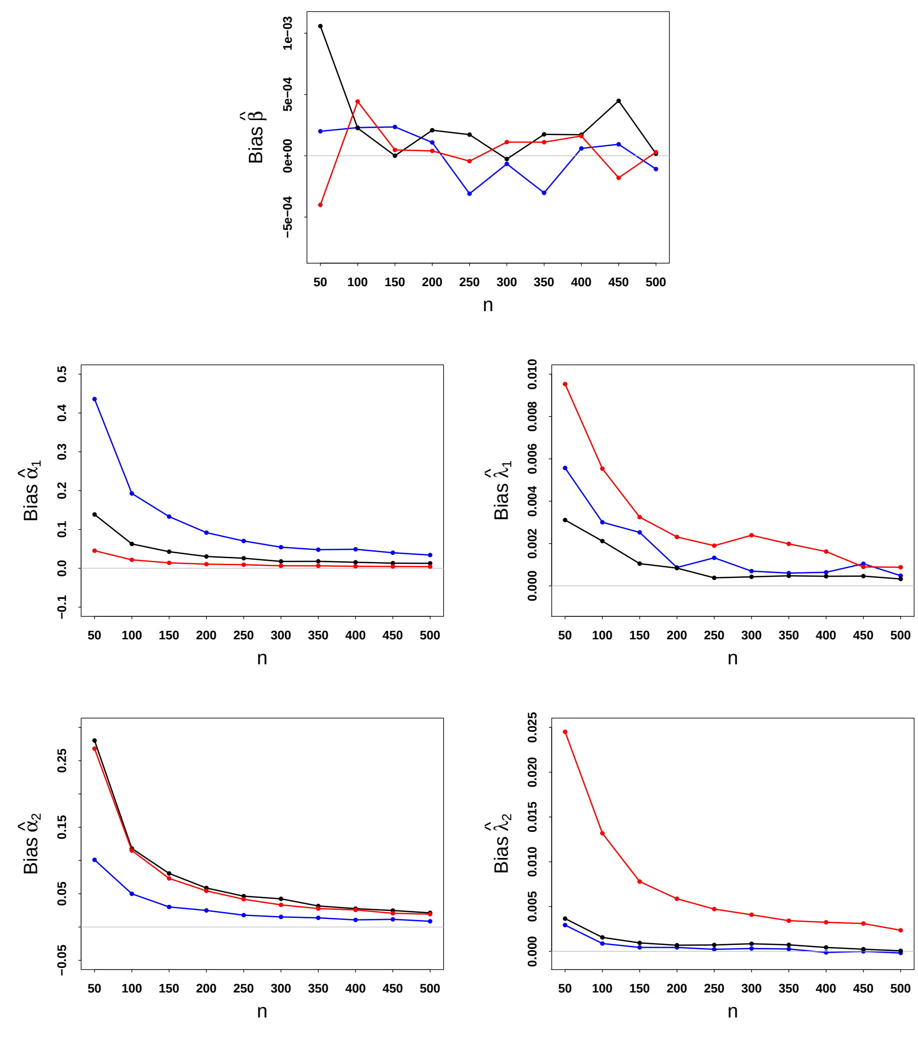

FIGURE 7. Bias of the MLEs of the parameters of DCA distribution: $(\beta, \alpha_1, \lambda_1, \alpha_2, \lambda_2)$: $(0.35, 2, 1, 1, 2)$(——), $(0.5, 1, 2, 2, 1)$(——), $(0.7, 0.5, 1, 1, 0.5)$(——).

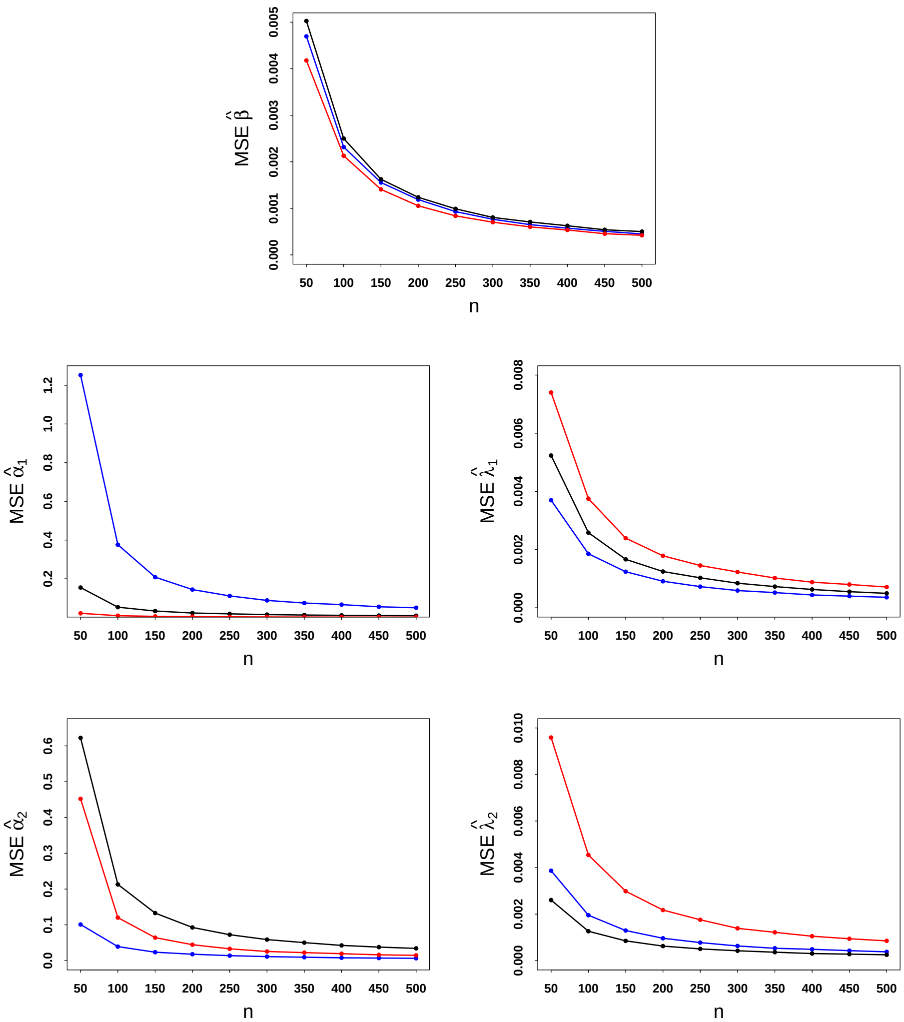

FIGURE 8. MSE of the MLEs of the parameters of DCA distribution: $(\beta, \alpha_1, \lambda_1, \alpha_2, \lambda_2)$: $(0.35, 2, 1, 1, 2)$(——), $(0.5, 1, 2, 2, 1)$(——), $(0.7, 0.5, 1, 1, 0.5)$(——).

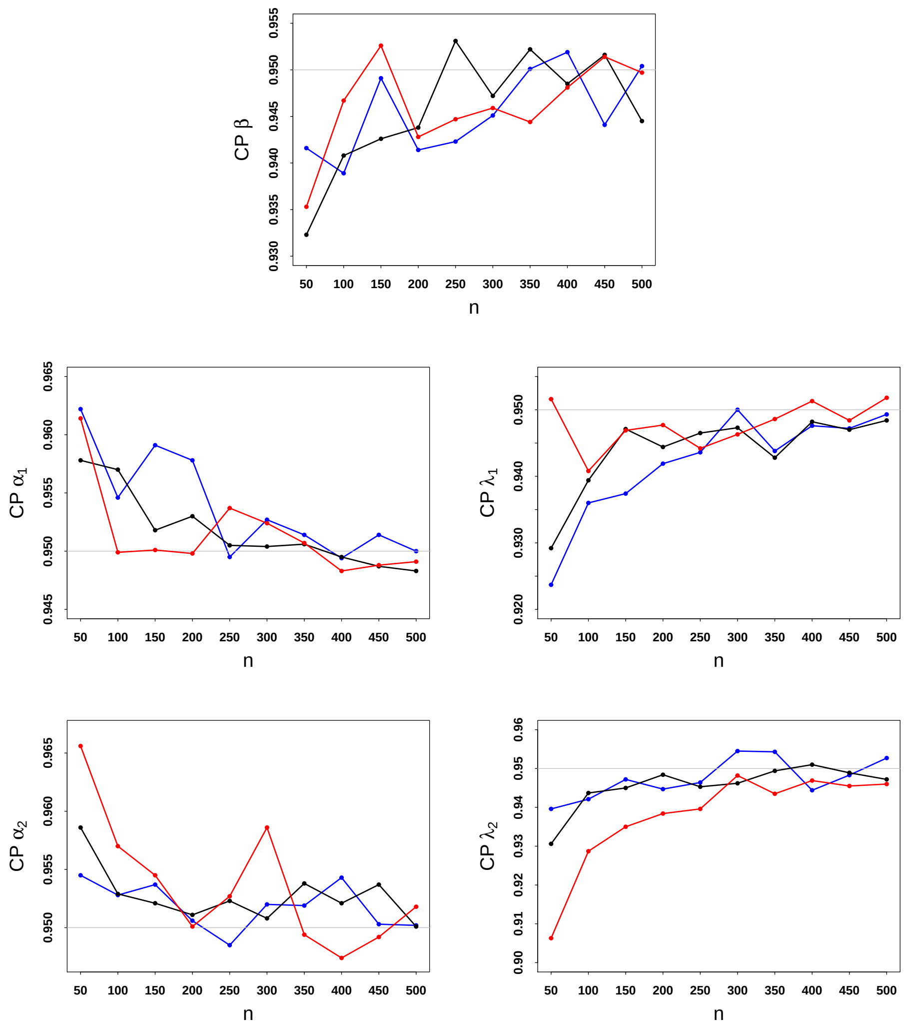

FIGURE 9. CP of the 95% confidence intervals of the parameters of DCA distribution: $(\beta, \alpha_1, \lambda_1, \alpha_2, \lambda_2)$: $(0.35, 2, 1, 1, 2)$(——), $(0.5, 1, 2, 2, 1)$(——), $(0.7, 0.5, 1, 1, 0.5)$(——).

The results of the simulation study are reported in Figures 7–9. These results are summarized below.

(1) Figure 7 shows that the *absolute* biases of the MLEs of the parameters are small and tend to zero for large n.
(2) Figure 8, shows that the MSE of the MLEs of the parameters are small and decrease as n increases.
(3) Figure 9 shows that the coverage probability of 95% confidence interval of each parameter is close to the nominal level of 95%.

The above conclusions show that the MLEs of the parameters of the DCA distribution are well behaved for point estimation and confidence intervals.

5. **Application**

Here, we fit the proposed DCA distribution to a real data set from DNA microarray reported by [6]. The considered data, labelled as "SID 377353, ESTs [5':, 3':AA055048]", consists of 118 observations.

Some descriptive statistics of the DNA microarray data are summarized in Table 1. Note that the skewness value is negative, indicating that the data is left-skewed.

TABLE 1. Descriptive statistics of DNA microarray data.

Min.	Q1	Median	Q3	Max.	Mean	St. dev.
-0.3390	-0.1138	0.0130	0.0780	0.2840	-0.0133	0.1293

For comparing the proposed DCA distribution with other bimodal double distributions, we consider double inverse-Gaussion (DIG) distribution introduced by [2] and double log-normal (DLN) distribution introduced by [1].

(i) The PDF of the DIG distribution is

$$f_Z(z) \quad = \quad \begin{cases} \overline{\beta}\ f_{X_2}(|z|;\nu_2,\lambda_2), & z<0,\\ \beta\ \ f_{X_1}(z;\nu_1,\lambda_1), & z\geq 0,\end{cases}$$

where, $\overline{\beta}=1-\beta,\ 0<\beta<1,\ \nu_j,\lambda_j>0,\ j=1,2,$

$$f_{X_j}\,(x;\nu_j,\lambda_j)=\sqrt{\frac{\lambda_j}{2\pi}}\ x^{-3/2}\ \exp\left[-\frac{\lambda_j\,(x-\nu_j)^2}{2\nu_j^2 x}\right],\quad x>0, \tag{36}$$

are the PDFs of the IG distributions.

(ii) The PDF of the DLN distribution is

$$f_Z(z) \quad = \quad \begin{cases} \overline{\beta}\ f_{X_2}(|z|;\mu_2,\sigma_2), & z<0, \\ \beta\ \ f_{X_1}(z;\mu_1,\sigma_1), & z\geq 0, \end{cases}$$

where, $\overline{\beta}=1-\beta,\ 0<\beta<1,\ \mu_j\in\mathbb{R},\ \sigma_j>0,\ j=1,2,$

$$f_{X_j}\left(x;\mu_j,\sigma_j\right)=\frac{1}{\sqrt{2\pi}\ \sigma_j\ x}\ \exp\left[-\frac{\left(\ln(x)-\mu_j\right)^2}{2\sigma_j^2}\right],\quad x>0, \tag{37}$$

are the PDFs of LN distributions.

Table 2 gives the MLEs, their standard errors (S.E.), estimated log-likelihoods, Akaike information criterion (AIC) and Bayesian information criterion (BIC) of the fitted DIG, DLN and DCA distributions.

TABLE 2. Summary of the three fitted models for DNA microarray data.

Model	Parameter	MLE	S.E.	$\ln\widehat{L}$	AIC	BIC
DIG	β	0.542	0.046	39.249	-68.499	-54.645
	ν_1	0.087	0.017			
	λ_1	0.036	0.006			
	ν_2	0.132	0.018			
	λ_2	0.126	0.024			
DLN	β	0.542	0.046	64.829	-119.659	-105.805
	μ_1	-2.812	0.127			
	σ_1	1.016	0.090			
	μ_2	-2.224	0.104			
	σ_2	0.764	0.074			
DCA	β	0.542	0.046	78.573	-147.146	-133.293
	α_1	42.771	7.561			
	λ_1	0.008	0.007			
	α_2	22.866	4.401			
	λ_2	0.028	0.010			

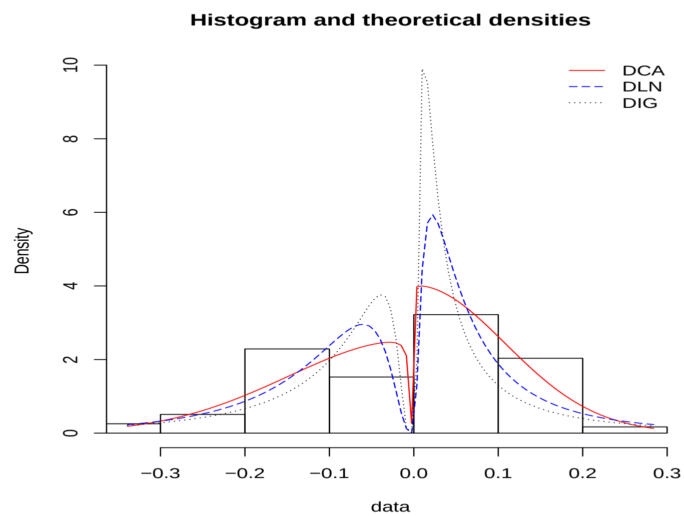

FIGURE 10. Histogram and theoretical densities of the three fitted models.

Table 3 gives the Kolmogrov-Smirnov (KS), Anderson-Darling (AD) and Cramér-von Mises (CVM) goodness-of-fit tests of the considered three models.

TABLE 3. Goodness-of-fit tests of fitted models

	KS		AD		CVM	
Model	statistic	p-value	statistic	p-value	statistic	p-value
DIG	0.126	0.046	3.285	0.020	0.545	0.030
DLN	0.065	0.709	0.851	0.446	0.103	0.570
DCA	0.078	0.468	0.594	0.653	0.109	0.543

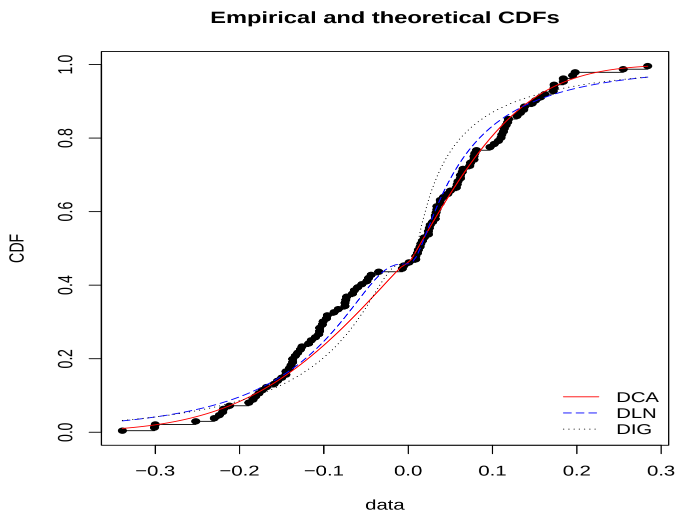

FIGURE 11. Empirical and theoretical CDFs of the three fitted models.

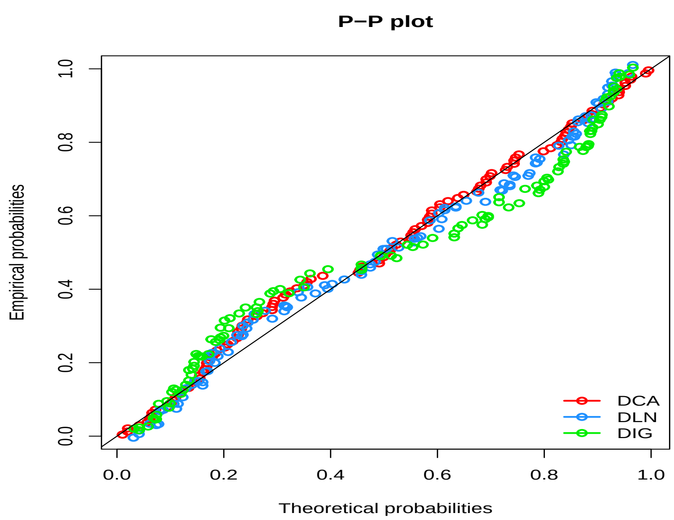

FIGURE 12. P-P plots of the three fitted models.

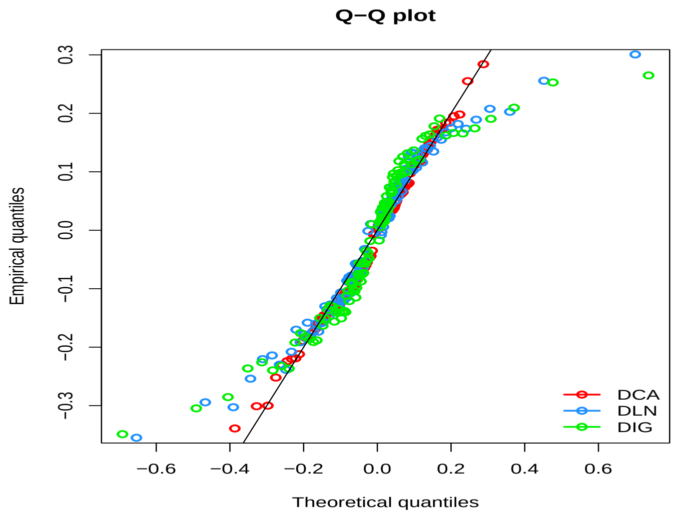

FIGURE 13. Q-Q plots of the three fitted models.

Remarks.

(1) DCA model has the smallest AIC and smallest BIC, see Table 2.

(2) DCA and DLN models (DIG model) are not rejected (is rejected) by all three goodness-of-fit tests, see Table 3.

(3) DCA model has the best performance in four diagnostic plots, namely (i) Histogram and theoretical densities, (ii) Empirical and theoretical CDFs, (iii) Percentile-Percentile (P-P), and (iv) Quantile-Quantile (Q-Q), see Figures 10 - 13.

Based on the last remarks, we conclude that DCA model is the most suitable for modeling the considered data.

6. Conclusion and comments

We proposed a new bimodal distribution on the real line, referred to as the double Chaudhry-Ahmad distribution. We derived several properties of the proposed distribution including density function, cumulative distribution function, moments, harmonic mean, Tsallis and Shannon entropies, extropy and weighted extropy. The model parameters were estimated by maximum likelihood approach. Monte Carlo simulation results indicate satisfactory performance of the maximum likelihood estimates. The proposed distribution is applied to a real set from DNA microarray data and is compared with recently introduced double inverse-Gaussian and double lognormal distributions. The proposed model can be extended by adding location and scale parameters to obtain more flexibility for modeling bimodal data on the whole real line.

References

[1] M. F. Alfahad, M. E. Ghitany, A. N. Alothman, S. Nadarajah, A bimodal extension of the log-normal distribution on the real line with an application to DNA microarray data. *Mathematics*, **11**, No 15 (2023), Art. 3360.

[2] A. Almutairi, M. E. Ghitany, A. N. Alothman, R. C. Gupta, Double inverse-Gaussian distributions and associated inference. *Journal of the Indian Society for Probability and Statistics*, **24** (2023), 157-182.

[3] E. Aly, A unifieded approach for developing Laplace-type distributions. *Journal of the Indian Society for Probability and Statistics*, **19** (2018), 245-269.

[4] N. Balakrishnan, S. Kocherlakota, On the double Weibull distribution: Order statistics and estimation. *Sankhya: The Indian Journal of Statistics* Series B, **47** (1985), 161-178.

[5] S. Bansal, N. Gupta, Weighted extropies and past extropy of order statistics and k-record values. *Communications in Statistics-Theory and Methods*, **51**, No 17 (2021), 1-24.

[6] M. Cankaya, Asymmetric bimodal exponential power distribution on the real line. *Entropy*, **20** (2018), 1-19.

[7] M. A. Chaudhry, M. Ahmad, On a probability function useful in size modelling. *Canadian Journal of Forest Research*, **23** (1993), 1679-1683.

[8] A. V. Dattatreya Rao, V. L. Narasimham, Linear estimation in double Weibull distribution. *Sankhya: The Indian Journal of Statistics* Series B, **51** (1989), 24-64.

[9] Z. Govindarajulu, Best linear estimates under symmetric censoring of the parameters of a double exponential population. *Journal of the American Statistical Association*, **61** (1966), 248-258.

[10] F. Lad, G. Sanfilippo, G. Agro, Extropy: Complementary dual of entropy. *Statistical Science*, **30**, No 1 (2015), 40–58.

[11] J. Mazucheli, A. F. Menezes, S. Dey, S. Nadarajah, Improved parameter estimation of the Chaudhry and Ahmad distribution with climate applications. *Chilean Journal of Statistics*, **11** (2020), 137-150.

[12] A. Plucinska, On certain problems connected with a division of a normal population into parts. *Zastosow. Mat.*, **8** (1965), 117-125.

[13] R Core Team R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*, Vienna, Austria, 2023.

[14] C. E. Shannon, A mathematical theory of communication. *Bell System Technical Journal*, **27**, No 3 (1948), 379-423.

[15] C. Tsallis, Possible generalization of Boltzmann-Gibbs statistics. *Journal of Statistical Physics*, **52** (1988), 479-487.