Austrian Journal of Statistics
July 2023, Volume 52, 1–24.
http://www.ajs.or.at/
doi:10.17713/ajs.v52i3.1434

Easily Changeable Kurtosis Distribution

Piotr Sulewski
Pomeranian University in Słupsk, Poland

Abstract

The goal of this paper is to introduce the easily changeable kurtosis (ECK) distribution. The uniform distribution appears as a special cases of the ECK distribution. The new distribution tends to the normal distribution. Properties of the ECK distribution such as PDF, CDF, modes, inflection points, quantiles, moments, moment generating function, Moors' measure, moments of order statistics, random number generator and the Fisher Information Matrix are derived. The unknown parameters of the ECK distribution are estimated by the maximum likelihood method. The Shannon, Renyi and Tsallis entropies are calculated. Illustrative examples of applicability and flexibility of the ECK distribution are given. The most important R codes are presented in the Appendix.

Keywords: normal distribution, modeling kurtosis, departure from normality, Monte Carlo method.

1. Introduction

The article presents a flexible, symmetric distribution defined in the finite domain. It is named as the easily changeable kurtosis (ECK) distribution. Its special cases is the uniform distribution. The ECK distribution tends to the normal distribution. The various properties of the ECK distribution are presented. Undoubtedly, symmetric distributions do not form such a big family as asymmetric distributions. Table 1 andnTable 2 present (in alphabetical order) symmetric distributions defined in the infinite and finite domains, respectively, with the formulas for the excess kurtosis $\overline{\gamma}_2$ and the number of modes. Instead of kurtosis γ_2, the paper analyzes the excess kurtosis $\overline{\gamma}_2 = \gamma_2 - 3$, which can be positive or negative.

The author would like to emphasize that the correctness of the formulas obtained in the Mathematica software has been checked using numerical methods.

As shown in Tables 1 and 2, we can divide the mentioned distributions into eleven groups, namely:

1. distribution with one mode (9-14,16-18,20,22-25,27,30-31,33-34),
2. distribution with at least two modes (1,4,6-8,19)
3. distribution with two modes (2-3,15,21,26,29,31)

Notation: $\phi(x, a, b)$ - PDF of $N(a, b)$; $\Phi(x, a, b)$ - CDF of $N(a, b)$.

Table 1: Symmetric distributions defined in the infinite domain. Excess kurtosis $\overline{\gamma}_2$ and modality

No.	Name	$\overline{\gamma}_2$	Range of $\overline{\gamma}_2$	Modality
1	Bimodal exponential power	$\frac{\Gamma\left(\frac{q+5}{p}\right)\Gamma\left(\frac{q+1}{p}\right)}{\Gamma[(q+3)/p]^2}-3, \begin{matrix} p\geq 1 \\ q\geq 0\end{matrix}$	$[-3,3]$ $[-1.2,3]$ $(q=0)$ <0 $(p>2, q=0)$	$q=0$ (1 mode) $q>0$ (2 modes)
2	Bimodal normal	$-4/3$		2 modes
3	Bimodal Laplace	$1/3$		2 modes
4	Bimodal power normal	Complicated, $p\in R$	$(-2,0)$ $(p>1)$ $(0,10.97)$ $(p<1)$	$p\leq 1$ (1 mode) $p>1$ (2 modes)
5	Cauchy	undefined		1 mode
6	Extended Normal	$\frac{-12p^2}{(3p+1)^2}-3, p>0$	$\left[-\frac{13}{3},-3\right]$	$p>0.5$ (2 modes) $p<0.5$ (1 mode)
7	Extended Laplace	$\frac{3(p^2+6p+1)}{(3p+1)^2}-3, p>0$	$\left[-\frac{8}{3},0\right]$	$p>0.5$ (2 modes) $p<0.5$ (1 mode)
8	Extended t	$\frac{36p-12p^2v+78p^2+6}{(v-4)(3p+1)^2}-3,$ $p>0, v=5,6,\ldots$	$\left[-\frac{13}{3},3\right]$	$p>0.5$ (2 modes) $p<0.5$ (1 mode)
9	Generalized normal	$\frac{\Gamma(5/p)\Gamma(1/p)}{\Gamma(3/p)^2}-3, (p>0)$	$(-1.2,0)$ $(p>2)$ >0 $(p<2)$	1 mode
10	Hyperbolic secant	2		1 mode
11	Laplace	3		1 mode
12	Logistic	6/5		1 mode
13	Normal	0		1 mode
14	Normal-exponential-gamma	$\frac{3p}{p-2}$ $(p>2)$	<3	1 mode
15	Plasticizing component	$\frac{\sqrt{\pi}\Gamma(0.5+2/p)}{\Gamma(0.5+1/p)^2}-3, p\geq 1$	$(-2,0)$	2 modes
16	t	$\frac{6}{v-4}$ $(v=5,6,\ldots)$	$(0,6]$	1 mode
17	Tukey $(p=0)$	1.2		1 mode
18	Tukey $p\in(-1/3,0]$	complicated $\left(p>-\frac{1}{4}\right)$	>0	1 mode
19	Uniform $U(a,b)$	$-6/5$		any value in (a,b)
20	Voigt	undefined		1 mode

4. distribution with an undefined excess kurtosis (5,20),
5. distribution with a complicated excess kurtosis formula (4,18,27,33),
6. distribution with a constant excess kurtosis value (2-3,10-13,17,19,21,25,28,30-32,34),
7. distribution with discrete excess kurtosis values (8,16,22,23),
8. distribution with excess kurtosis values on infinite interval (9,14,18,26,33),
9. distribution with excess kurtosis values on finite interval (1,4,6-8,15-16,22-24,27,29),
10. distribution with an existing discontinuous function $p=f(\overline{\gamma}_2)$, where p is the shape parameter (16,22-23).
11. distribution with an existing continuous function $p=f(\overline{\gamma}_2)$, where p is the shape parameter (14,24).

The proposed distribution, as it will be shown further in this paper, can be classified simultaneously into groups 1, 9 and 11, just like the Q-gaussian distribution defined in the finite domain. Both distributions are characterized by a simple excess kurtosis formula (see group 7) and their special cases are uniform and normal distributions. The excess kurtosis range

Table 2: Symmetric distributions defined in the finite domain. Excess kurtosis $\overline{\gamma}_2$ and modality

No.	Name	Domain	$\bar{\gamma}_2$	Range of $\bar{\gamma}_2$	Modality
21	Arcsine	$(a, a+b), b>0$	-1.5		2 modes
22	Bates	$[a, b]$ $b>0, n=1,2,\ldots$	$-\frac{6}{5n}$	$[-1.2,0)$	1 mode
23	Irwin–Hall	$[0, n], n=2,3,\ldots$	$\frac{-6}{5n}$	$[-1.2,0)$	1 mode
24	Q-gaussian	$\left[\frac{\pm 1}{\sqrt{p(1-q)}}\right], \begin{matrix} p>0 \\ q\in(0,1)\end{matrix}$	$\frac{6q-6}{7-5q}$	$[-0.857,0]$	1 mode
25	Raised cosine	$[a-b, a+b]$ $b>0$	$\frac{6(90-\pi^4)}{5(\pi^2-6)^2}$		1 mode
26	U-quadratic	$[a, b], b>a$	$\frac{3}{112}(b-a)^4$	>0	2 modes
27	Von Mises	$[-k\pi, k\pi], k\in Z$	complicated	$[-1.2,1.069]$	1 mode
28	Wigner semicircle	$[-R, R], R>0$	-1		1 mode
29	U-power	$[a-b, a+b]$ $b>0, k\in N^+$	$\frac{-8k^2-24k-6}{(2k+5)(2k+1)}$	$[-2,-1.81]$	2 modes
30	Sine	$[a, a+b]$ $a\in R, b>0$	$\frac{2(96-\pi^4)}{(\pi^2-8)^2}$		1 mode
31	Semicircle	$[-r, r], r>0$	-1		1 mode
32	U-shaped	$(-k, k), k>0$	-1.5		2 modes
33	Tukey (p>0)	$\left[-\frac{1}{p}, \frac{1}{p}\right], p>0$	complicated	>-1.243	1 mode
34	Triangle type I	$[-1,1]$	-0.6		1 mode

for the Q-gaussian distribution defined in the finite domain is $(-0.857, 0)$ and for the ECK distribution is $(-2, 0)$. For comparison, the excess kurtosis of the U-power distribution takes discrete values in the range $[-2, -1.81]$.

The ECK distribution can be used to model an excess kurtosis in the range $(-2, 0)$. This distribution can be extremely useful when you want to seamlessly test the goodness-of-fit tests (GoFTs) ability to detect deviations from normality caused by a negative excess kurtosis. The new proposition as a component of the mixed distribution will also be used when fitting the distributions to the data.

It should also be mentioned that there is a group of asymmetric distributions, which are symmetrical for certain parameter values, e.q. the truncated normal, Birnbaum-Saunders (Birnbaum and Saunders 1969), skew-normal (Azzalini 1985), beta, two-piece normal (Gibbons and Mylroie 1985) and two-piece power normal (Sulewski 2019b) distributions.

The author would like to emphasize that the correctness of the formulas obtained in the Mathematica software has been checked using numerical methods.

This paper is organized as follows. Section 2 presents the main properties of the ECK distribution such as PDF, CDF, modes, inflection points, quantiles, moments, moment generating function, Moors' measure, moments of order statistics, instructions to generate ECK random numbers and the Fisher Information Matrix. The estimation procedures are provided in Section 3, while the Entropy is presented in Section 4. The papers ends with applications and conclusions. The most important R codes are given in the Appendix.

2. Main properties of introduced distribution

2.1. Distribution and density functions

Definition 1. The distribution of the random variable X with PDF given by

$$f(x;a,p) = \frac{\left(1-\frac{x^2}{a^2}\right)^p}{aB(0.5,p+1)}, x \in \left\{ \begin{array}{ll} (-a,a) & \text{if } -1<p<0 \\ [-a,a] & \text{if } p \geq 0 \end{array} \right\} \quad (1)$$

is called the easily changeable kurtosis (ECK) distribution, where $a > 0$ is the scale parameter and $p > -1$ is the shape parameter. The $ECK(a > 0, p > -1)$ is symmetric around zero since, based on (1), $f(x;a,p) = f(-x;a,p)$ (Figure 2). The R codes of the dECK function for computing PDF are provided in the Appendix.

The standard deviation of the new proposal based on (28) equals $\sqrt{\frac{a^2}{2p+3}}$, therefore the $ECK(a,p)$ distribution tends to the normal distribution $N\left(0, a\sqrt{\frac{1}{2p+3}}\right)$.
Let M (2) be the similarity measure of these distributions (Sulewski 2019a). We have for $a > 0, p > -1$

$$M(a,p) = \int_{-a}^{a} \min\left[f(x;a,p), \phi\left(x, 0, a\sqrt{\frac{1}{2p+3}}\right)\right] dx \quad (2)$$

The similarity measure M takes values on (0,1) and if PDFs are identical then $M = 1$.
For example $M(4,100) = 0.998$, $M(4,120) = 0.999$. For more details see Figure 1.

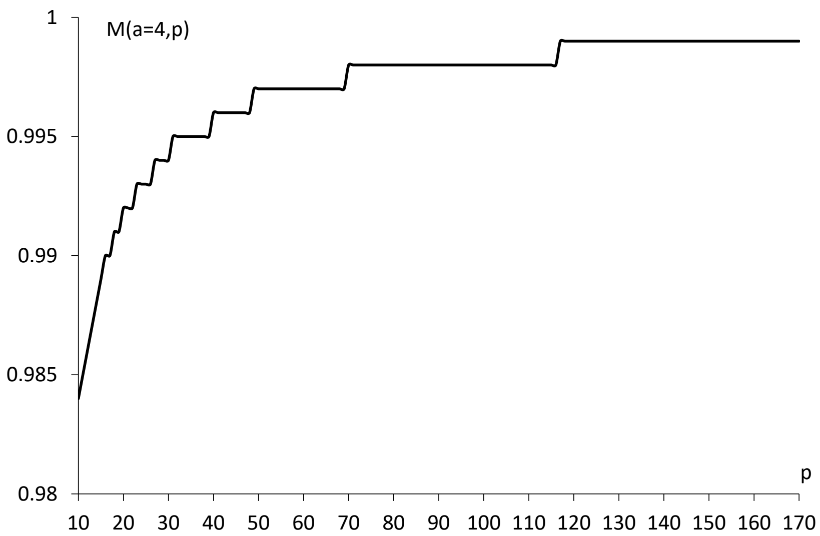
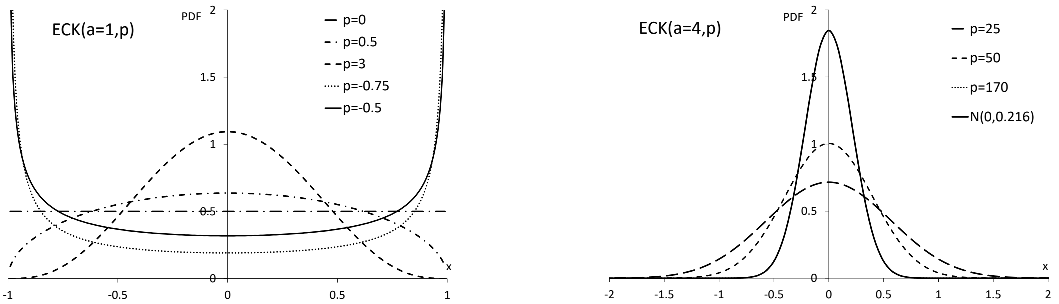


Figure 1: Similarity measure of the $ECK(a,p)$ and normal distributions

The $ECK(a, p = 0)$ is the uniform distribution $U(-a, a)$ (Figure 2, serie $p = 0$). The $ECK(a, p > 0)$ based on Theorem 4 is unimodal with mode equals 0 (see e.g. Figure 2, series $p = 0.5, p = 3$). The $ECK(a, -1 < p < 0)$ based on Theorem 4 is pseudo $(-a < x < a)$ bimodal with bathtub shape (see e.g. Figure 2, serie $p = -0.75, p = -0.5$). The $ECK(4, p = 120)$ is in 99.9% the normal distribution $N(0, 0.216)$ (Figure 2, serie $N(0, 0.216)$).



(a) $p \leq 0$ (b) $p > 0$

Figure 2: PDF of the $ECK(a,p)$ distribution for various parameter values

Theorem 1. *Let X follows the ECK $(a > 0, p > -1)$ with the PDF (1), then $\frac{X^2}{a^2}$ follows the beta distribution with the parameters 0.5 and $p+1$.*

Proof. Let $p \geq 0$. Support $[-a, a]$ can be presented in two intervals $[-a, 0] \cup (0, a]$, in which $\frac{x^2}{a^2}$ is strictly monotonic.The inverse function of $y = \frac{x^2}{a^2}$ on several intervals is given by

- for $x \in [-a, 0]$ we have $x = -a\sqrt{y}$ and $\frac{dx}{dy} = -\frac{a}{2\sqrt{y}}$,
- for $x \in (0, a]$ we have $x = a\sqrt{y}$ and $\frac{dx}{dy} = \frac{a}{2\sqrt{y}}$

We cane write the PDF of Y as

$$g(y; a, p) = \frac{a}{2\sqrt{y}} \left[f(-a\sqrt{y}; a, p) + f(a\sqrt{y}; a, p) \right].$$

We have from (1)

$$f(-a\sqrt{y}; a, p) = \frac{1}{aB(0.5, p+1)} (1-y)^p = f(a\sqrt{y}; a, p).$$

As a result of simple transformations

$$g(y; p) = \frac{1}{B(0.5, p+1)} y^{-0.5} (1-y)^p.$$

The above equation is the PDF of the beta distribution with the parameters 0.5 and $p+1$. The same result we obtain for $-1 < p < 0$ and support $(-a, a)$. The proof is complete. □

Theorem 2. *Let Y follows the beta distribution with the PDF $g(y; 0.5, p+1)$ then $X = \pm a\sqrt{Y}$ follows the $ECK(a > 0, p > -1)$ with the PDF*

$$f(x; a, p) = \frac{|x|}{a^2} g\left(\frac{x^2}{a^2}; 0.5, p+1\right). \tag{3}$$

Proof. Let $\pm a\sqrt{y}$ are strictly monotonic. The inverse function of $x = a\sqrt{y}$ and $x = -a\sqrt{y}$ is $y = \frac{x^2}{a^2}$ as well as $\left|\frac{dy}{dx}\right| = \frac{2|x|}{a^2}$. In this situation, we can write a PDF of X as

$$f(x; a, p) = \frac{|x|}{a^2} g\left(\frac{x^2}{a^2}; 0.5, p+1\right).$$

The proof is complete. □

Using the PDF of the beta distribution in (3), we get (1) obviously.

Theorem 3. *If $X \sim WCK(a > 0, p > -1)$ with the PDF $f(x; a, p)$ (1) then CDF of X is given by*

$$F(x; a, p) = 0.5 \left[1 + sgn(x)\, G\left(\frac{x^2}{a^2}; 0.5, p + 1 \right) \right], \tag{4}$$

where sgn and G are the signum function and CDF of the beta distribution, respectively.

Proof. Let $x > 0$. The WCK is a symmetric distribution so using (1) we obtain

$$F(x; a, p) = 0.5 + \frac{1}{aB(0.5, p + 1)} \int_0^x \left(1 - \frac{t^2}{a^2} \right)^p dt. \tag{5}$$

Substituting $u = \frac{t^2}{a^2}$, we obtain from (5)

$$F(x; a, p) = 0.5 + \frac{1}{2B(0.5, p + 1)} \int_0^{\frac{x^2}{a^2}} u^{-0.5} (1 - u)^p du. \tag{6}$$

Using the definition of an incomplete beta function we get

$$F(x; a, p) = 0.5 + \frac{1}{2B(0.5, p + 1)} B\left(\frac{x^2}{a^2}, 0.5, p + 1 \right) \tag{7}$$

and finally

$$F(x; a, p) = 0.5 + 0.5G\left(x^2/a^2; 0.5, p + 1 \right). \tag{8}$$

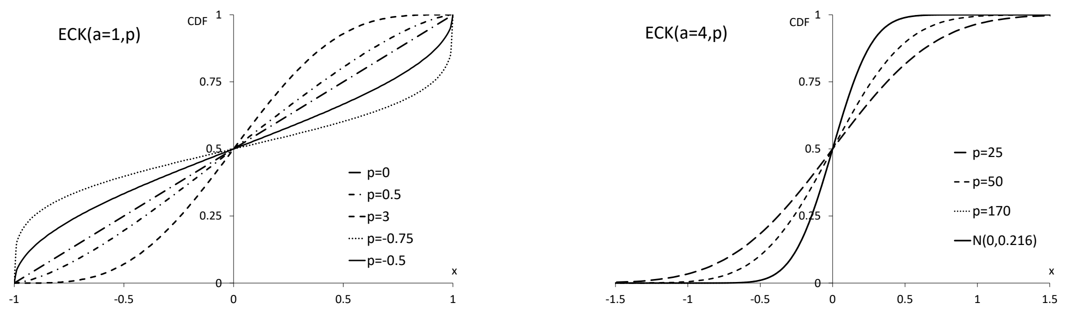
For $x < 0$ we get similarly

$$F(x; a, p) = 0.5 - 0.5G\left(x^2/a^2; 0.5, p + 1 \right). \tag{9}$$

The proof is complete. □

The R codes of the pECK function for computing CDF are provided in the Appendix.

Figure 3 plots CDF of the $ECK(a > 0, p > -1)$ distribution for some values of parameters. For $p = 0$ we obtain the straight line (uniform distribution). For $p > 0$ CDF is convex in $[-a, 0)$ and is concave in $(0, a]$. For $-1 < p < 0$ CDF is concave in $(-a, 0)$ and is convex in $(0, a)$. For $p = 170$ CDFs of ECK distribution and normal one coincide.



(a) $p \leq 0$ (b) $p > 0$

Figure 3: CDF of the $ECK(a, p)$ distribution for various parameter values

Theorem 4. *The $ECK(a > 0, p > -1)$ distribution with PDF given by (2) is identifiable in a parameter space $v = (a, p)$.*

Proof. Let $v_1 = (a_1, p_1)$ and $v_2 = (a_2, p_2)$. Let us suppose that $f_{v_1}(x) = f_{v_2}(x)$ for all x from support. This condition based on (4) implies that

$$0.5 + 0.5sgn\,(x)\,G\left(\frac{x^2}{a_1^2}; 0.5, p_1+1\right) = 0.5 + 0.5sgn\,(x)\,G\left(\frac{x^2}{a_2^2}; 0.5, p_2+1\right), \tag{10}$$

where G is a CDF of the beta distribution.

From (10) we have

$$G\left(\frac{x^2}{a_1^2}; 0.5, p_1+1\right) = G\left(\frac{x^2}{a_2^2}; 0.5, p_2+1\right). \tag{11}$$

The proof is complete because the beta distribution is identifiable. □

2.2. Modes and inflection points

Theorem 5. *Let $X \sim ECK(a>0, p>-1)$. If $p=0$ then modal values $x_m \in [-a,a]$ (case of uniform distribution). If $p>0$ then $x_m = 0$. If $-1<p<0$ then the $ECK(a,p)$ distribution is pseudo bimodal with modes $x_m(-a), x_m(a)$. The $f(x;a,p>0)$ (2) is monotonically increasing on the interval $(-a,0)$ and monotonically decreasing on the interval $(0,a)$. The $f(x;a,-1<p<0)$ (2) is monotonically decreasing on the interval $(-a,0)$ and monotonically increasing on the interval $(0,a)$.*

Proof. Let $p \geq 0$ then PDF of the ECK distribution for any $x \in [-a,a]$ is given by

$$f(x;a,p) = \frac{\Gamma(p+1.5)}{a\sqrt{\pi}\Gamma(p+1)}\left(1-\frac{x^2}{a^2}\right)^p. \tag{12}$$

Let $p=0$ then $f(x;a,0) = \frac{\Gamma(1.5)}{a\sqrt{\pi}} = \frac{0.5\Gamma(0.5)}{a\sqrt{\pi}} = \frac{0.5\sqrt{\pi}}{a\sqrt{\pi}} = \frac{1}{2a}$ is constant in [-a,a].

Let $p>0$ then

$$\frac{d}{dx}f(x;a,p) = \frac{p\Gamma(p+1.5)}{a\sqrt{\pi}\Gamma(p+1)}\left(1-\frac{x^2}{a^2}\right)^{p-1}\left(-\frac{2x}{a^2}\right). \tag{13}$$

As a result of simple transformations $x_m = 0$ and (13) is positive on the interval $(-a,0)$ and negative on the interval $(0,a)$.

Let $-1<p<0$ then PDF (12) is defined for any $x \in (-a,a)$. As a result of simple transformations, (13) is negative on the interval $(-a,0)$ and positive on the the interval $(0,a)$. For x values very close to $-a$ and a PDF (12) has locally maximum values. The author of this paper denotes these values as $x_m(-a), x_m(a)$ and proposed distribution defines as pseudo bimodal with modes at these points. The proof is complete. □

Theorem 6. *Let $X \sim ECK(a>0, p>-1)$. The inflection points of the $f(x;a,p)$ (1) for $p>1$ are given by means of the following formulas*

$$x_1 = -\frac{a}{\sqrt{2p-1}}, x_2 = \frac{a}{\sqrt{2p-1}}. \tag{14}$$

Proof. We can write (13) as

$$\frac{d}{dx}f(x;a,p) = \frac{-2p\Gamma(p+1.5)}{a^3\sqrt{\pi}\Gamma(p+1)}x\left(1-\frac{x^2}{a^2}\right)^{p-1}. \tag{15}$$

Let $A = \frac{-2p\Gamma(p+1.5)}{a^3\sqrt{\pi}\Gamma(p+1)}$ then the second derivative of (12) is given by

$$\frac{d^2}{dx^2}f(x;a,p) = A\left\{\left(1-\frac{x^2}{a^2}\right)^{p-1} + x(p-1)\left(1-\frac{x^2}{a^2}\right)^{p-2}\left(\frac{-2x}{a^2}\right)\right\}, \tag{16}$$

$$\frac{d^2}{dx^2} f(x;a,p) = A\left(1-\frac{x^2}{a^2}\right)^{p-2}\left\{1-\frac{x^2}{a^2}+x(p-1)\left(\frac{-2x}{a^2}\right)\right\}, \tag{17}$$

thus

$$\frac{d^2}{dx^2} f(x;a,p) = 1-\frac{x^2}{a^2}-\frac{2x^2(p-1)}{a^2} \tag{18}$$

Based on (18) we have for $p > 0.5$

$$a^2 - x^2 - 2px^2 + 2x^2 = 0,$$

$$a^2 - x^2(2p-1) = 0,$$

$$\left(a - x\sqrt{2p-1}\right)\left(a + x\sqrt{2p-1}\right) = 0,$$

$$x_1 = \frac{a}{\sqrt{2p-1}} \vee x_2 = \frac{-a}{\sqrt{2p-1}}.$$

For $a > 0 \wedge p > 0.5$ is $x_1 < x_2$, hence taking into account the domain $(-a, a)$ we obtain

$$x_1 = -\frac{a}{\sqrt{2p-1}} \wedge x_1 > -a, \quad x_2 = \frac{a}{\sqrt{2p-1}} \wedge x_2 < a. \tag{19}$$

Solving the above inequalities, we have finally $p > 1$. The proof is complete. □

2.3. Quantiles

Theorem 7. *Let $X \sim ECK(a > 0, p > -1)$. The q-th $(0 < q < 1)$ quantile x_q is the solution of the following equation*

$$0.5\left[1 + sgn(x_q)\, I_{x_q^2/a^2}(0.5, p+1)\right] - q = 0, \tag{20}$$

where sgn and I are the signum function and CDF of the beta distribution, respectively.
The proposed distribution is symmetrical then $x_p = -x_{1-p}$, obviously and $x_{0.5} = 0$.

Proof. Obtaining (20), based on the quantile definition, is trivial.
The quantile x_q can be computed by numerical methods. The R codes of the qECK function for computing the quantile x_q are provided in the Appendix. □

2.4. Moments, moment generating function and Moors' measure

Theorem 8. *The k-th $(k = 0, 1, 2, \dots)$ non-central moments of the $ECK(a > 0, p > -1)$ distribution are given by*

$$\alpha_k = \frac{a^k + (-a)^k}{2B(0.5, p+1)} B\left(\frac{k+1}{2}, p+1\right). \tag{21}$$

Proof. The k-th $(k = 0, 1, 2, \dots)$ non-central moments of the ECK distribution, based on (1), are defined as

$$\alpha_k = \frac{\int_0^a x^k\left(1-\frac{x^2}{a^2}\right)^p dx + \int_{-a}^0 x^k\left(1-\frac{x^2}{a^2}\right)^p dx}{aB(0.5, p+1)} = \frac{I_1 + I_2}{aB(0.5, p+1)}. \tag{22}$$

To solve the integrals I_1 and I_2, we need the following formula (Gradshteyn and Ryzhik (2014))

$$\int_0^1 x^{c-1}\left(1-x^\lambda\right)^{d-1} dx = \frac{1}{\lambda} B\left(\frac{c}{\lambda}, d\right). \tag{23}$$

Substituting $t = \frac{x}{a}$ we have

$$I_1 = a^{k+1} \int_0^1 t^k \left(1 - t^2\right)^p dt \tag{24}$$

and based on (23)

$$I_1 = \frac{a^{k+1}}{2} B\left(\frac{k}{2} + \frac{1}{2}, p + 1\right). \tag{25}$$

Substituting $u = \frac{-x}{a}$ we get

$$I_2 = (-a)^{k+1} \int_0^1 u^k \left(1 - u^2\right)^p du \tag{26}$$

and based on (23)

$$I_2 = \frac{(-a)^{k+1}}{2} B\left(\frac{k}{2} + \frac{1}{2}, p + 1\right). \tag{27}$$

Substituting (25) and (27) to (22) we obtain (21).The proof is complete. □

Theorem 9. *The non-central moments $\alpha_k k = (1, 3, ...)$, variance μ_2 and excess kurtosis $\overline{\gamma}_2$ of the $ECK(a > 0, p > -1)$ distribution are given by*

$$\alpha_k = 0 \; (k = 1, 3, ...) \, , \mu_2 = \frac{a^2}{2p + 3}, \overline{\gamma}_2 = \frac{-6}{2p + 5}. \tag{28}$$

Proof. The proof $\alpha_k = 0 \, (k = 1, 3, ...)$, based on (21), is trivial. The first non-central moment equals zero, so the non-central moments $\alpha_k \, (k = 0, 1, ...)$ are equal to the central moments $\mu_k \, (k = 0, 1, ...)$. From (21), we have

$$\mu_2 = \alpha_2 = \frac{a^2 B\,(1.5, p + 1)}{B\,(0.5, p + 1)}, \tag{29}$$

$$\mu_4 = \alpha_4 = \frac{a^4 B\,(2.5, p + 1)}{B\,(0.5, p + 1)}. \tag{30}$$

As a result of simple transformation, using the properties $\Gamma\,(x + 1) = x\Gamma\,(x)\,, \Gamma\,(0.5) = \sqrt{\pi}$, we get

$$\mu_2 = \frac{a^2}{2p + 3}, \tag{31}$$

$$\mu_4 = \frac{3a^4}{(2p + 3)\,(2p + 5)}. \tag{32}$$

The excess kurtosis is defined as

$$\overline{\gamma}_2 = \gamma_2 - 3 = \frac{\mu_4}{\mu_2^2} - 3 \tag{33}$$

and substituting (31) and (32) to (33), we have

$$\overline{\gamma}_2 = \frac{\mu_4}{\mu_2^2} - 3 = \frac{3a^4}{(2p + 3)\,(2p + 5)} \frac{(2p + 3)^2}{a^4} - 3 = \frac{3\,(2p + 3)}{2p + 5} - 3 = \frac{-6}{2p + 5}. \tag{34}$$

The proof is complete. □

Changing the shape parameter values p, we can obtain the excess kurtosis of the $ECK(a > 0, p > -1)$ distribution in the range $(-2, 0)$ (Figure 4). As the shape parameter p increases, the excess kurtosis increases from -2 to 0.

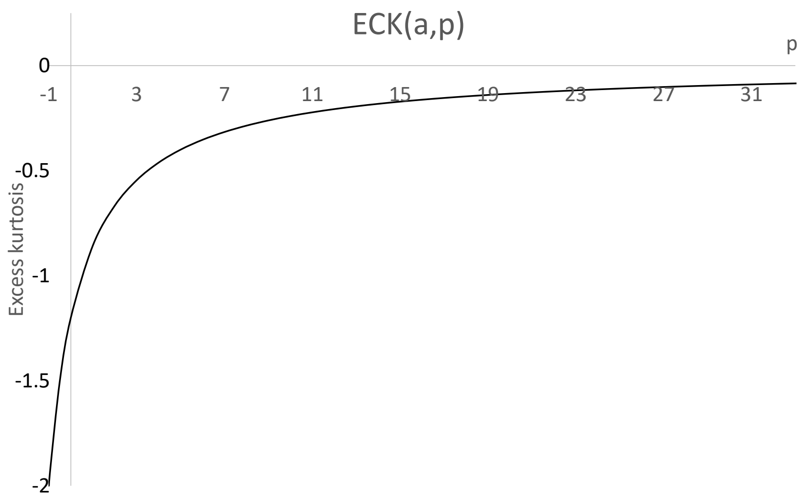


Figure 4: Excess kurtosis as a function of the shape parameter p

Theorem 10. *The moment generating function of the $ECK(a>0, p>-1)$ distribution is given by*

$$M_X(t;a,p) = {}_0F_1\left(p+1.5;\frac{a^2t^2}{4}\right), \tag{35}$$

where ${}_0F_1(a;x)$ is the confluent hypergeometric function.

Proof. Based on (1) we have

$$M_X(t;a,p) = \frac{\Gamma(p+1.5)}{a\sqrt{\pi}\Gamma(p+1)}\int_{-a}^{a} e^{tx}\left(1-\frac{x^2}{a^2}\right)^p dx. \tag{36}$$

Formula (36) can be written using a power series (Wolfram (1988))

$$M_X(t;a,p) = \sum_{k=0}^{\infty}\frac{(at)^{2k}}{4^k k!\,(p+1.5)_k} = \sum_{k=0}^{\infty}\frac{\left(\frac{a^2t^2}{4}\right)^k}{(p+1.5)_k\,k!} =_0 F1\left(p+1.5;\frac{a^2t^2}{4}\right), \tag{37}$$

where $(p+1.5)_k$ is the Pochhammer symbol

$$x = \left\{\begin{array}{ll} 1 & \text{if } k=0 \\ (p+1.5)(p+2.5)(p+3.5)\ldots(p+k+0.5) & \text{if } k>0. \end{array}\right\}.$$

The proof is complete. □

Moors (1988) proposed a measure based on quantiles in the form

$$T = \frac{x_{7/8} - x_{5/8} + x_{3/8} - x_{1/8}}{x_{6/8} - x_{2/8}}, \tag{38}$$

where x_q is the solution of (20). The measure (38) is a quantile alternative for kurtosis and exists even for distribution for which no moments exist. Of course, this measure does not

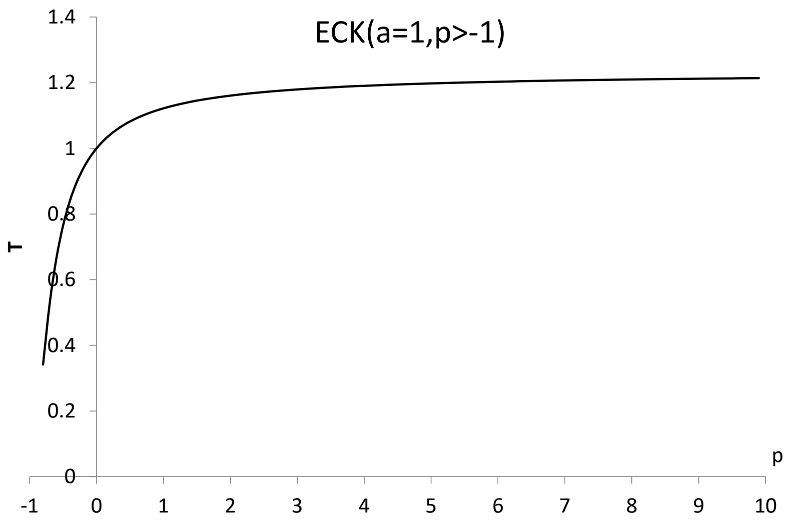


Figure 5: Moors' measure T as a function of the shape parameter p.

depend on the scale parameter a. Figure 5 shows the measure T as a function of the shape parameter p. The $T(p)$ is strictly increasing function for the initial p values.

2.5. Moments of order statistics

Theorem 11. *Let $X_{i,n}$ be the i-th order statistic $(X_{1,n} \leq X_{2,n} \leq ... \leq X_{n,n})$ in a sample of size n from the $ECK(a > 0, p > -1)$ distribution. The k-th moment of the i-th order statistic $X_{i,n}$ is defined as*

$$E\left(X_{i,n}^k\right) = \int_{-\infty}^{\infty} x^k f_{i,n}(x; a, p) \tag{39}$$

and

$$f_{i,n}(x; a, p) = \frac{n!\left(a^2 - x^2\right)^p F(x; a, p)^{i-1}\left[1 - F(x; a, p)\right]^{n-i}}{(i-1)!\,(n-i)!a^{2p+1}B(0.5, p+1)}, \tag{40}$$

where $F(x; a, p)$ is CDF (4).

Proof. The proof based on the definition of PDF of the order statistics is trivial. □

Figure 6 shows PDF of the $X_{5i,30}(i = 1, \ldots, 5)$ of the $ECK(a > 0, p > -1)$ distribution

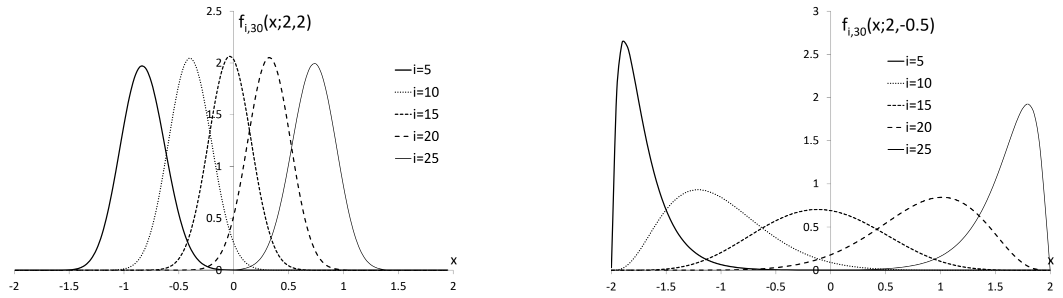
2.6. Random number generator

Theorem 12. *Let $X \sim ECK(a > 0, p > -1), R \sim U(0, 1)$. The random number generator of X, using the inverse CDF method, is given by*

$$x = \begin{cases} a\sqrt{G^{-1}(2R-1; 0.5, p+1)} & \text{if } R \geq 0.5 \\ -a\sqrt{G^{-1}(1-2R; 0.5, p+1)} & \text{if } R < 0.5 \end{cases} \tag{41}$$

where G^{-1} is the quantile function of the beta distribution.

The proof, using inversion of the CDF (4), is trivial.
The R codes of the rECK function for generating n values of X are presented in the Appendix.



(a) $a = 2, p = 2$ (b) $a = 2, p = -0.5$

Figure 6: PDF of the $X_{5i,30}(i = 1, \ldots, 5)$ of the $ECK(a > 0, p > -1)$ distribution

2.7. Fisher information matrix

Theorem 13. *The Fisher information matrix* $I_{i,j}(i, j = 1, 2)$ *for the* $ECK(a > 0, p > -1)$ *distribution is given by*

$$I_{11} = \frac{2p^2 + p}{2a^2 (p - 1)} + \frac{1}{a} + \frac{1}{a^2}, \tag{42}$$

$$I_{12} = I_{21} = (B - A) \left(\frac{a + 2}{2a}\right) + \left(\frac{a + 2}{2}\right) \ln (a) + \frac{2apI_1 - I_2}{a}, \tag{43}$$

$$I_{22} = (A - B) [A - B + 2I_2 - 4 \ln (a)] + I_3 + 4 \ln (a) [I_2 + \ln (a)], \tag{44}$$

where

$$A = \Psi (p + 1.5), B = \Psi (p + 1), \tag{45}$$

$$I_1 = E \left[\frac{x}{x^2 - a^2} \ln \left(a^2 - x^2\right)\right], I_2 = E \left[\ln \left(a^2 - x^2\right)\right], I_3 = E \left[\ln^2 \left(a^2 - x^2\right)\right] \tag{46}$$

and Ψ *is the digamma function.*

Proof. First, we need to take the logarithm. From (1) we have

$$\ln [f (x|a, p)] = \ln [\Gamma (p + 1.5)] + p \ln \left(1 - \frac{x^2}{a^2}\right) - \ln (a) - \ln \left(\sqrt{\pi}\right) - \ln [\Gamma (p + 1)]. \tag{47}$$

Second, we need to calculate the partial derivatives

$$\frac{d \ln [f (x|a, p)]}{da} = -\frac{1}{a} + \frac{2px}{x^2 - a^2}, \frac{d \ln [f (x|a, p)]}{dp} = \Psi (p + 1.5) + \ln \left(1 - \frac{x^2}{a^2}\right) - \Psi (p + 1), \tag{48}$$

where Ψ is the digamma function. Hence, we get the Fisher score in the form

$$\boldsymbol{h} (x; a, p) = \begin{bmatrix} -\frac{1}{a} + \frac{2px}{x^2 - a^2} \\ \Psi (p + 1.5) + \ln \left(1 - \frac{x^2}{a^2}\right) - \Psi (p + 1) \end{bmatrix}. \tag{49}$$

Let $\boldsymbol{u} (x; a, p) = \boldsymbol{h} (x; a, p) \boldsymbol{h} (x; a, p)^T$ and $A = \Psi (p + 1.5), B = \Psi (p + 1)$ then

$$u_{11} = \left(\frac{2px}{x^2 - a^2} - \frac{1}{a}\right)^2, u_{22} = \left[A - B + \ln \left(1 - \frac{x^2}{a^2}\right)\right]^2. \tag{50}$$

$$u_{12} = u_{21} = \left(\frac{2px}{x^2 - a^2} - \frac{1}{a}\right) \left[A - B + \ln \left(1 - \frac{x^2}{a^2}\right)\right]. \tag{51}$$

Let $I_{i,j}= E\left[u_{i,j}\right](i,j=1,2)$ then

$$I_{11}=4p^2E\left[\frac{x^2}{\left(x^2-a^2\right)^2}\right]-\frac{4p}{a}E\left[\frac{x^2}{x^2-a^2}\right]+\frac{1}{a^2}, \tag{52}$$

$$I_{12}=\frac{2p\left(A-B\right)}{\left(E\left[\frac{x^2}{x^2-a^2}\right]\right)^{-1}}+2pE\left[\frac{x^2}{x^2-a^2}\ln\left(1-\frac{x^2}{a^2}\right)\right]+\frac{B-A-E\left[\ln\left(1-\frac{x^2}{a^2}\right)\right]}{a}, \tag{53}$$

$$I_{22}=\left(A-B\right)^2+2\left(A-B\right)E\left[\ln\left(1-\frac{x^2}{a^2}\right)\right]+E\left[\ln^2\left(1-\frac{x^2}{a^2}\right)\right]. \tag{54}$$

To write the Fisher Information Matrix in a simpler form we use (23), (29) and integration by substitution $\frac{x}{a}$.

We have:

$E\left[\frac{x^2}{x^2-a^2}\right]=\frac{-1}{4p}$,

$E\left[\frac{x^2}{(x^2-a^2)^2}\right]=\frac{2p+1}{8a^2p(p-1)}$,

$E\left[\frac{x}{x^2-a^2}\ln\left(1-\frac{x^2}{a^2}\right)\right]=E\left[\frac{x}{x^2-a^2}\ln\left(a^2-x^2\right)\right]-2\ln\left(a\right)E\left[\frac{x^2}{x^2-a^2}\right]=$

$=E\left[\frac{x}{x^2-a^2}\ln\left(a^2-x^2\right)\right]+\frac{\ln(a)}{2p}=I_1+\frac{\ln(a)}{2p}$,

$E\left[\ln\left(1-\frac{x^2}{a^2}\right)\right]=E\left[\ln\left(a^2-x^2\right)\right]-2\ln\left(a\right)=I_2-2\ln\left(a\right)$,

$E\left[\ln^2\left(1-\frac{x^2}{a^2}\right)\right]=E\left\{\left[\ln\left(a^2-x^2\right)-2\ln\left(a\right)\right]^2\right\}=E\left[\ln^2\left(a^2-x^2\right)\right]-4\ln\left(a\right)E\left[\ln\left(a^2-x^2\right)\right]+$ $4\ln^2\left(a\right)=I_3-4\ln\left(a\right)I_2+4\ln^2\left(a\right)$,

where I_1, I_2, I_3 are defined in (46).

Substituting above calculations into formulas (52, 53, 54), as a result of simple transformations, we get

$I_{11}=\frac{2p^2+p}{2a^2(p-1)}+\frac{1}{a}+\frac{1}{a^2}$,

$I_{12}=I_{21}=\left(B-A\right)\left(\frac{a+2}{2a}\right)+\left(\frac{a+2}{2}\right)\ln\left(a\right)+\frac{2apI_1-I_2}{a}$

$I_{22}=\left(A-B\right)\left[A-B+2I_2-4\ln\left(a\right)\right]+I_3+4\ln\left(a\right)\left[I_2+\ln\left(a\right)\right]$.

The proof is complete. □

3. Maximum likelihood estimation

Let $x_1^*, x_2^*, ..., x_n^*$ be a random sample size n from the $ECK(a>0, p>-1)$ distribution. Our target is to estimate the unknown values of the parameters a, p. The likelihood function based on (1) is given by

$$L=\prod_{i=1}^{n}f\left(x_i^*;a,p\right)=\frac{\Gamma\left(p+1.5\right)}{a\sqrt{\pi}\Gamma\left(p+1\right)}\prod_{i=1}^{n}\left(1-\frac{x_i^{*2}}{a^2}\right)^p, \tag{55}$$

then the log-likelihood function is defined as

$$l=\ln\left(L\right)=n\ln\left[\frac{\Gamma\left(p+1.5\right)}{a\sqrt{\pi}\Gamma\left(p+1\right)}\right]+p\sum_{i=1}^{n}\ln\left(1-\frac{x_i^{*2}}{a^2}\right) \tag{56}$$

and

$$\frac{dl}{da}=-\frac{n}{a}-\frac{2p}{a}\sum_{i=1}^{n}\frac{x_i^{*2}}{a^2-x_i^{*2}}=0, \tag{57}$$

$$\frac{dl}{dp} = n\Psi\left(p+1.5\right) - n\Psi\left(p+1\right) + \sum_{i=1}^{n} \ln\left(1 - \frac{x_i^{*2}}{a^2}\right) = 0, \tag{58}$$

where Ψ is the digamma function. The maximum likelihood estimates (MLEs) are solutions of the system equations (57) and (58). From (52) we get

$$p = -\frac{n}{2\sum_{i=1}^{n} \frac{x_i^*}{a^2 - x_i^{*2}}}. \tag{59}$$

Substituting (59) into (58) we obtain the nonlinear equation in the form

$$\Psi\left(1.5 - \frac{n}{2\sum_{i=1}^{n} \frac{x_i^*}{a^2 - x_i^{*2}}}\right) - \Psi\left(1 - \frac{n}{2\sum_{i=1}^{n} \frac{x_i^*}{a^2 - x_i^{*2}}}\right) + \frac{1}{n}\sum_{i=1}^{n} \ln\left(1 - \frac{x_i^{*2}}{a^2}\right) = 0. \tag{60}$$

Solving (60) with numerical method we have $\widehat{a}$ and from (59)

$$\widehat{p} = -\frac{n}{2\sum_{i=1}^{n} \frac{x_i^*}{\widehat{a}^2 - x_i^{*2}}}. \tag{61}$$

The biases and the root mean squared errors (RMSEs) of the MLEs are shown in Table 3. The simulation study was performed with 10^3 samples using sample sizes of $50, 100, 200, 500$. The samples were drawn from the $ECK(1,p)$, where $p = (1,2,3)$. We observe that the estimates approach true values when the sample size increases, it implies the consistency of the estimates. The biases of the $\widehat{a}$ and $\widehat{p}$ diminish for large samples and are smaller for $\widehat{a}$ than for $\widehat{p}$. The RMSEs increase with the value of p.

Table 3: Biases and RMSEs of the MLEs from the $ECK(1,p)$

		$\widehat{a}$		$\widehat{p}$	
p	n	Bias	RMSE	Bias	RMSE
1	50	-0.171	0.111	-0.210	0.895
	100	-0.080	0.073	-0.167	0.529
	200	-0.060	0.042	-0.149	0.311
	500	-0.059	0.023	-0.103	0.172
2	50	-0.135	0.195	-0.417	1.885
	100	-0.052	0.157	-0.163	1.558
	200	-0.021	0.102	-0.071	0.893
	500	-0.015	0.056	-0.075	0.468
3	50	-0.189	0.247	-0.691	2.644
	100	-0.100	0.193	-0.402	2.127
	200	-0.027	0.163	-0.015	1.758
	500	-0.008	0.095	0.000	1.036

To examine the accuracy of the coverage probability of the asymptotic confidence intervals (CIs), another simulation study was performed with 10^3 samples using sample sizes of $50, 100, 200, 500$. The study focused on the parameters a, p and samples drawn from the $ECK(a = 1, p = 4)$. The coverage probabilities of the obtained 95% CIs for $a = 1, p = 4$ reported in Table 4 are very close to the nominal level. The results suggested that the obtained standard errors and hence the asymptotic CIs are reliable.

Table 4: Coverage probability for the standard asymptotic 95% CIs

Sample size n	Coverage probability	
	a	p
50	0.941	0951
100	0.940	0.935
200	0.952	0.941
500	0.965	0.963

4. Shannon, Renyi and Tsallis entropies

Theorem 14. *Let $f(x;a,p)$ be PDF (1). The Shannon entropy S, Renyi entropy of order α R_α and Tsallis entropy of order α T_α of the $ECK(a>0,p>-1)$ are given as*

$$S\left(a,p\right)=0.5\ln\left[aB\left(0.5,p+1\right)\right]-\frac{pI_4}{a^{2p+1}B\left(0.5,p+1\right)}+p\ln\left(a\right), \tag{62}$$

$$R_\alpha\left(a,p\right)=\frac{1}{1-\alpha}\ln\left\{\frac{B\left(0.5,p\alpha+1\right)}{2a^{\alpha-1}\left[B\left(0.5,p+1\right)\right]^\alpha}\right\}, \tag{63}$$

$$T_\alpha\left(a,p\right)=\frac{B\left(0.5,p\alpha+1\right)}{2\left(\alpha-1\right)a^{\alpha-1}\left[B\left(0.5,p+1\right)\right]^\alpha}, \tag{64}$$

where $I_4=\int_{-a}^{a}\ln\left(a^2-x^2\right)\left(a^2-x^2\right)^p$.

Proof. The Shannon entropy is defined as (Shannon (1948); Tabass, Borzadaran, and Amini (2016))

$$S\left(a,p\right)=-\int_{-a}^{a}f\left(x;a,p\right)\ln\left[f\left(x;a,p\right)\right]dx, \tag{65}$$

then

$$S\left(a,p\right)=\frac{-1}{aB\left(0.5,p+1\right)}\int\limits_{-a}^{a}\left(1-\frac{x^2}{a^2}\right)^p\left[\ln\left(\frac{1}{aB\left(0.5,p+1\right)}\right)+p\ln\left(1-\frac{x^2}{a^2}\right)\right]dx=I_5+I_6, \tag{66}$$

where

$$I_5=\frac{1}{aB\left(0.5,p+1\right)}\ln\left[aB\left(0.5,p+1\right)\right]\int_{-a}^{a}\left(1-\frac{x^2}{a^2}\right)^p dx, \tag{67}$$

$$I_6=\frac{-p}{aB\left(0.5,p+1\right)}\int_{-a}^{a}\left(1-\frac{x^2}{a^2}\right)^p\ln\left(1-\frac{x^2}{a^2}\right). \tag{68}$$

The formula (67), based on the integration by substituting $\left(\frac{x}{a}\right)$ and (23), is as follows

$$I_5=\frac{\ln\left[aB\left(0.5,p+1\right)\right]}{aB\left(0.5,p+1\right)}\frac{aB\left(0.5,p+1\right)}{2}=0.5\ln\left[aB\left(0.5,p+1\right)\right]. \tag{69}$$

We can write the integral in (68) as

$$\frac{1}{a^{2p}}\int_{-a}^{a}\left(a^2-x^2\right)^p\ln\left(a^2-x^2\right)-2\ln\left(a\right)\int_{-a}^{a}\left(1-\frac{x^2}{a^2}\right)^p=\frac{1}{a^{2p}}I_4+I_7, \tag{70}$$

where

$$I_7=-2\ln\left(a\right)\int_{-a}^{a}\left(1-\frac{x^2}{a^2}\right)^p=-a\ln\left(a\right)B\left(0.5,p+1\right). \tag{71}$$

The formula (68) using (70) and (71) is given by

$$I_6 = \frac{-p}{aB\left(0.5, p+1\right)} \left[\frac{1}{a^{2p}} I_4 - a\ln\left(a\right) B\left(0.5, p+1\right)\right] = \frac{-pI_4}{a^{2p+1} B\left(0.5, p+1\right)} + p\ln\left(a\right). \quad (72)$$

Substituting (69) and (72) to (66), we get (62).

The Renyi entropy of order α is defined as (Renyi (1961),Tabass *et al.* (2016))

$$R_\alpha\left(a, p\right) = \frac{1}{1-\alpha} \ln\left[\int_{-a}^{a} f\left(x; a, p\right)^\alpha dx\right] \left(\alpha > 0, \alpha \neq 1\right), \quad (73)$$

then based on (1)

$$R_\alpha\left(a, p\right) = \frac{1}{1-\alpha} \ln\left\{\frac{1}{a^\alpha \left[B\left(0.5, p+1\right)\right]^\alpha} \int_{-a}^{a} \left(1 - \frac{x^2}{a^2}\right)^{p\alpha} dx\right\}. \quad (74)$$

Using the integration by substituting and (23) we obtain (63).

The Tsallis entropy of order α is defined as (Tsallis (1988),Tabass *et al.* (2016))

$$T_\alpha\left(a, p\right) = \frac{1}{\alpha - 1} \int_{-a}^{a} \left[f\left(x; a, p\right)\right]^\alpha dx - 1 \left(\alpha > 0, \alpha \neq 1\right), \quad (75)$$

then based on (1)

$$T_\alpha\left(a, p\right) = \frac{1}{\left(\alpha - 1\right)\left[aB\left(0.5, p+1\right)\right]^\alpha} \int_{-a}^{a} \left(1 - \frac{x^2}{a^2}\right)^{p\alpha} dx. \quad (76)$$

Using the integration by substituting and (23) we obtain (64).

The proof is complete. □

The Renyi and Tsallis entropies converge to the Shannon entropy.

Figure 7 shows the Shannon, Renyi and Tsallis entropies for the $ECK(a = 1, p > -1)$ distribution. The Shannon and Renyi entropies increase for $p \in (-1, 0)$ and decrease for $p > 0$. The Tsallis entropy decreases for $p \in (-1, 0)$ and increases for $p > 0$. The higher a value, the higher S value. The higher α value, the lower R_α value.

5. Application

This section is divided into two subsections. We present examples of the applicability and flexibility of the $ECK(a > 0, p > -1)$. The first subsection is devoted to the GoFTs, the second one deals with fitting distributions to data.

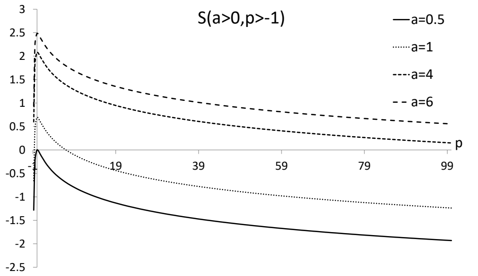
5.1. Comparison of goodness-of-fit tests

As it was mentioned in the Introduction, the advantage of the ECK distribution is e.g. a simple formula that allows you to change an excess kurtosis. The distribution can be extremely useful when you want to seamlessly test the GoFTs ability to detect deviations from normality caused by a negative excess kurtosis.

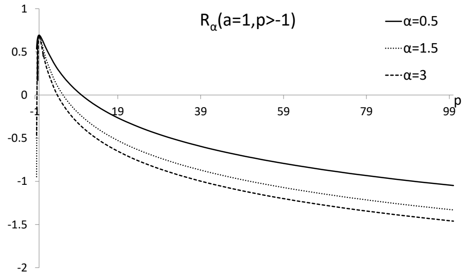
The excess kurtosis $\overline{\gamma}_2$ of the $ECK(a > 0, p > -1)$ distribution has a very simple form. From (31) we have

$$\overline{\gamma}_2 = \frac{-6}{2p+5} \Leftrightarrow p = -\frac{5\overline{\gamma}_2 + 6}{2\overline{\gamma}_2}. \quad (77)$$

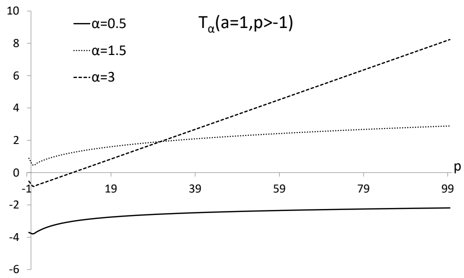
Based on the inequality $p > -1$ we obtain $\overline{\gamma}_2 \in (-2; 0)$ (see Figure 8, Table 5).



(a) Shannon



(b) Renyi



(c) Tsallis

Figure 7: Shannon, Renyi and Tsallis entropies, the $ECK(a > 0, p > -1)$ distribution

Let $x_{(1)}, x_{(2)}, ..., x_{(n)}$ be an ordered random sample of size n. Seven GoFTs were selected to be subjects of the Monte Carlo simulation. Five of them as being very popular GoFTs have been implemented in the R software. These tests are: Shapiro-Wilk (SW), Kolmogorov-Smirnov (KS), Cramer-von Mises (CVM), Anderson-Darling (AD) and Shapiro - Francia (SF). Two tests not implemented yet, probably for their novelty, are: H_n (Torabi, Montazeri, and Grane (1961)) and LF_m (Sulewski (2019a)) tests. The H_n test statistic is defined as

$$H_n = \frac{1}{n}\sum_{i=1}^{n} h\left[\frac{1+\Phi\left(\frac{x_{(i)}-\overline{x}}{s},0,1\right)}{1+\frac{i}{n}}\right], h\left(x\right) = \left(\frac{x-1}{x+1}\right)^2, \tag{78}$$

where $\overline{x}$ and s^2 are the sample mean and sample variance, respectively.

The LF_m test statistic is given by

$$LF_m = \max\left|\frac{i-\overline{\alpha}}{n-\overline{\alpha}-\overline{\beta}+1} - \Phi\left(\frac{x_{(i)}-\overline{x}}{s},0,1\right)\right|, \left(\overline{\alpha},\overline{\beta} \geq 1\right). \tag{79}$$

If an alternatively distribution is both symmetric and of negative excess kurtosis $\overline{\alpha} = \overline{\beta} = 0$ are recommended.

The similarity measure M (2) of $N(0.0216)$ and $ECK(a = 4, p = 170)$, as was mentioned in Section 2.1, is 0.999. Figure 8 shows PDF of the $N(0, 0.216)$ and $ECK(a = 4, p)$ distributions involved in the Monte Carlo simulation. In the legend of this figure values of similarity measures M (2) of these distributions are given. If p increases, the similarity measure M also increases.

Phase 1: In this phase the aim is to investigate to what degree selected GoFTs listed in Table 5 are able to distinct between $N(0, 0.216)$ and $ECK(a = 4, p)$ distributions. In other words the aim is to determine powers of GoFTs being under discussion when samples come from

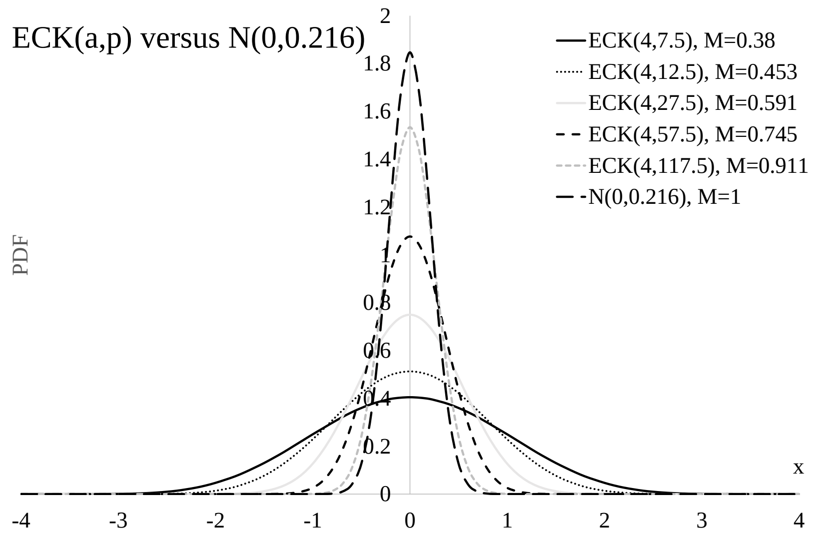


Figure 8: The $ECK(a, p)$ distribution with values of the similarity measure M to the $N(0, 0.216)$

$ECK(a = 4, p)$ general populations. Table 5 shows how the ECK distribution tends to the Normal distribution in terms of kurtosis as its shape parameter increases.

Table 5: ECK distribution versus $N(0, 0.216)$

$ECK\ (a = 4, p)$						$N(0, 0.216)$
p	7.5	12.5	27.5	57.5	117.5	
$\overline{\gamma}_2$	-0.3	-0.2	-0.1	-0.05	-0.025	0

For the aim to be accomplished, critical values $cv_{0.05}$ ascribed to the GoFTs (where $\alpha = 0.05$ is the test confidence level) were needed. These $cv_{0.05}$ values were estimated with the Monte Carlo method. Seven large scale experiments were performed each of which devoted to one of the GoFT. Each experiment consisted of generating 10^6 samples of sizes $n = 30$ and $n = 50$. The samples followed the $N(0, 0.216)$ distribution. Each sample was tested for normality. Obtained in this way values of test statistics (denoted $Q_i\ (i = 1, 2, ..., m)$ were collected an then ranked. Critical values were assessed according to the formula $cv_{0.05} = Q_{[0.95m]}$.

Table 6 presents obtained $cv_{0.05}$ critical values. Tables 7 and 9, in turn, present relevant test powers when samples come from the ECK general populations. The scale parameter was set to 4. Values of the shape parameter were listed in Table 5.

Table 6: Critical values $cv_{0.05}$ related to GoFTs considered

GoFT	$n = 30$	$n = 50$
LF	0.1589	0.1244
CVM	0.1241	0.1248
AD	0.7314	0.7396
SW	0.9863	0.9902
SF	0.9880	0.9915
H$_n$	0.0005	0.0003
LF$_m$	0.1394	0.1131

Table 7: Powers of tests when the $ECK(a = 4, p)$ is the actual population distribution not the Normal one, $n = 30$

	Excess kurtosis of the ECK				
GoFT	-0.3	-0.2	-0.1	-0.05	-0.025
LF	0.045	0.046	0.047	0.048	0.049
CVM	0.044	0.045	0.047	0.048	0.049
AD	0.043	0.044	0.046	0.047	0.049
SW	0.047	0.050	0.051	0.051	0.051
SF	0.055	0.054	0.051	0.051	0.051
H_n	0.049	0.049	0.051	0.051	0.053
LF_m	0.049	0.048	0.048	0.049	0.050

Table 8: Powers of tests when the $ECK(a = 4, p)$ is the actual population distribution not the Normal one, $n = 50$

	Excess kurtosis of the ECK				
GoFT	-0.3	-0.2	-0.1	-0.05	-0.025
LF	0.047	0.047	0.050	0.049	0.050
CVM	0.047	0.045	0.049	0.048	0.048
AD	0.047	0.045	0.048	0.048	0.048
SW	0.047	0.049	0.050	0.052	0.050
SF	0.055	0.054	0.052	0.051	0.049
H_n	0.050	0.049	0.052	0.050	0.050
LF_m	0.050	0.048	0.051	0.049	0.049

Powers of the considered tests are close to confidence level we set. It means that in Phase 1 we revealed that considered GoFTs are unable to distinct between the negatively skewed distribution (not only slightly but even moderately i.e. $\overline{\gamma}_2 = -0.3$) and the Normal distribution.

The following question is sure to be asked. What about numerous distributions similar to the new proposition. There were some investigations performed not being so in-depth like presented above, rather shallow ones. On their basis, however, one can tentatively say that the above conclusions relate to mentioned distributions too.

Phase 2. In this phase the aim is to investigate to what degree an undetected negative kurtosis impacts the performance of two basic test related to parameters of the Normal distribution, namely Student t-test and Fisher–Snedecor F test. For the aim to be accomplished we employ the Monte Carlo method and determine empirical CDFs of t and F test statistics in the case when samples come from $ECK(4, 7.5)$ general population. Then we compare these empirical CDFs with "true" t and F CDFs i.e. which hold for in the case when samples come from the Normal general population. These comparisons will be simply Kolmogorov-Smirnov GoFT. Since theoretical distributions are t and F distribution their parameters are known because they are so called degrees-of freedom equal to $n - 1$. It causes that K-S test can be applied in its classic, in other words, pre-Lilliefors form. What advocates the use of K-S is that it is powerful when sample is very large and parameters are not estimated from the sample but exactly known. Procedures that return values of "true" t and F CDFs are implemented in many computational environments including R software.

Let $x_{1,1}, x_{1,2}, ..., x_{1,n}$ and $x_{2,1}, x_{2,2}, ..., x_{2,n}$ be two samples of sizes n drawn from particular general populations. Let us remember that t and F test statistics have the following forms:

$$\dot{t} = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{s_{x1}^2 + s_{x2}^2}{n}}}, \dot{F} = \frac{s_{x1}^2}{s_{x2}^2}, \tag{80}$$

where $\overline{x_1}, \overline{x_2}$ are the sample means and s_{x1}, s_{x2} are the sample standard deviations.

The course of action was as follows:

Step 1: 10^4 pairs of samples both of size $n = 100$ were drawn from $ECK(4, 7.5)$ general population.

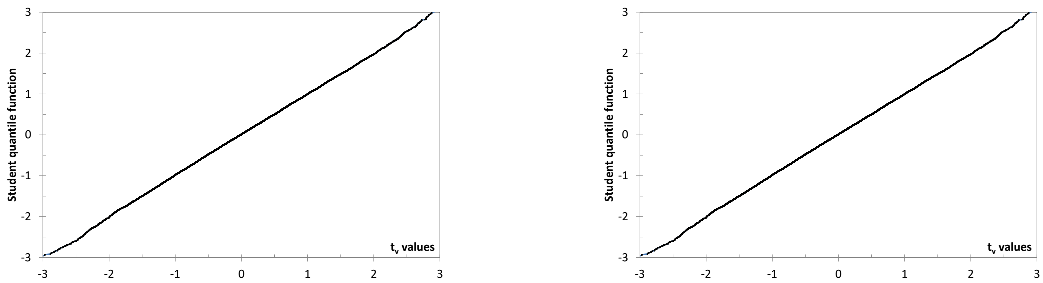
Step 2: These pairs of samples were consecutively, converted into pairs of $\dot{t}_v$ statistics and $\dot{F}_v$ statistics, $v = 1, 2, ..., m$.

Step 3: Sets of values of $\dot{t}_v$ and $\dot{F}_v$ statistics were stored in two matrices named T and F.

Step 4: The matrices were sorted in ascending order and served to determine two empirical CDFs namely $\Theta_t\left(\dot{t}_v\right)$ and $\Theta_F\left(\dot{F}_v\right)$.

Step 5: Probability papers were employed to check whether the above empirical CDFs fit the Student and Fisher-Snedecor distributions.

Figure 9 show empirical CDFs of Step 4 plotted on the Student and Snedecor probability papers. These probability papers were constructed in the same way as the Normal probability is constructed and commonly used by practitioners over the World. It turns out that the empirical distribution in question perfectly fit straight lines that relevant theoretical distributions. Thus, we can conclude that Student and Fisher-Snedecor tests may be applied even as population distributions are of negative excess kurtosis, slight or indeed moderate one.



(a) Student

(b) Snedecor

Figure 9: Empirical CDFs of Step 4 plotted on the Student and Snedecor probability paper

5.2. Fitting distributions to data

As it is well known, symmetric distributions have limited use in fitting the distributions to data (e.g. normal distribution). However, the situation looks much better when we use their mixture (e.g. compound normal distribution).

In this subsection, we present real data example to demonstrate a flexibility of the $ECK(a > 0, p > -1)$ distribution in the mixed variant. PDF of the compound ECK (CECK) distribution is given by

$$CECK\left(a, p_1, p_2, \omega\right) = \omega ECK\left(a, p_1\right) + (1 - \omega)\, ECK\left(a, p_2\right) \tag{81}$$

The estimation of the model parameters is carried out by the maximum likelihood method. To avoid local maxima of the logarithmic likelihood function, the optimization routine is run 100 times with several different starting values that are widely scattered in the parameter space.

Real data example The real data presents temperature dynamics of beaver Castor canadensis in north-central Wisconsin (Reynolds (1994)). Body temperature was measured by telemetry every 10 minutes from one period of less than a day. The data consists of 114 observations of the variable “measured body temperature in degrees Celsius” and are available in the R software with code beaver1[3].

The models selected for comparison with the $CECK\left(a, p_1, p_2, \omega\right)$ are:

- the compound normal (CN): $f_{CN}\left(x; a_1, b_1, a_2, b_2, \omega\right) = \omega\phi\left(x; a_1, b_1\right) + (1-\omega)\,\phi\left(x; a_2, b_2\right)$

- the compound Laplace (CL): $f_{CL}\left(x; a_1, b_1, a_2, b_2, \omega\right) = \omega f_L\left(x; a_1, b_1\right) + (1-\omega)\, f_L\left(x; a_2, b_2\right)$, where $f_L\left(x; a, b\right) = \frac{1}{2b}\exp\left[\exp\left(-\frac{|x-a|}{b}\right)\right]$

- the compound Cauchy (CC): $f_{CC}\left(x; a_1, b_1, a_2, b_2, \omega\right) = \omega f_C\left(x; a_1, b_1\right) + (1-\omega)\, f_C\left(x; a_2, b_2\right)$, where $f_C\left(x; a, b\right) = \frac{1}{\pi b\left[1+\left(\frac{x-a}{b}\right)^2\right]}$

- the compound Bates (CB): $f_{CB}\left(x; a, b, n_1, n_2, \omega\right) = \omega f_B\left(x; a, b, n_1\right) + (1-\omega)\, f_B\left(x; a, b, n_2\right)$, where

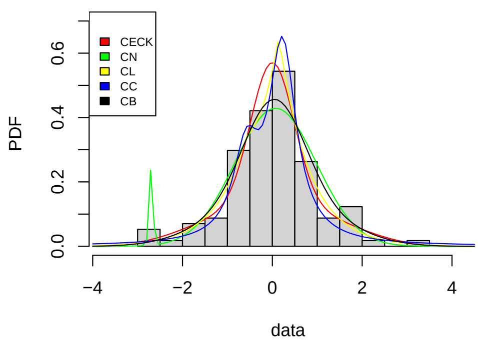
  $f_B\left(x; a, b, n\right) = \frac{n(b-a)^{-1}}{2(n-1)!}\sum_{k=0}^{n}(-1)^k C_n^k sgn\left(n\frac{x-a}{b-a}-k\right)\left(n\frac{x-a}{b-a}-k\right)^{n-1}$ $(a < b, n = 1, 2, ...)$.

Table 9 presents the MLEs, log-likelihood function l, AIC, BIC and HQIC for the data set. Models are sorted by AIC values. Figure 10 presents histograms, estimated PDFs and CDFs of the analyzed models.

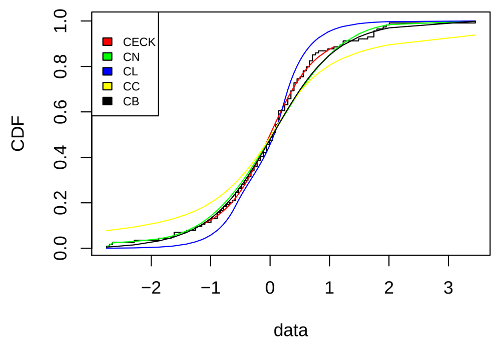
Table 9: Results of estimation. Information criteria. Real data example.

Model		$\widehat{\Theta}$	-l	AIC	BIC	HQIC
CECK	$\widehat{a}$	5.010	155.1653	318.3307	329.2755	322.7726
	$\widehat{p_1}$	5.412				
	$\widehat{p_2}$	55.361				
	$\widehat{\omega}$	0.484				
CN	$\widehat{a_1}$	-2.6999	154.7443	319.4887	333.1697	325.041
	$\widehat{b_1}$	0.0420				
	$\widehat{a_2}$	0.0708				
	$\widehat{b_2}$	0.9060				
	$\widehat{\omega}$	0.0255				
CL	$\widehat{a_1}$	-0.5798	155.3905	320.781	334.462	326.3333
	$\widehat{b_1}1$	0.7221				
	$\widehat{a_2}$	0.1437				
	$\widehat{b_2}2$	0.6631				
	$\widehat{\omega}$	0.2007				
CC	$\widehat{a_1}$	-0.5813	163.3858	336.7715	350.4525	342.3239
	$\widehat{b_1}$	0.3533				
	$\widehat{a_2}$	0.2263				
	$\widehat{b_2}$	0.3686				
	$\widehat{\omega}$	0.2938				
CB	$\widehat{a}$	-4.8025	157.4469	324.8938	338.5748	330.4462
	$\widehat{b}$	4.8952				
	$\widehat{n_1}$	18				
	$\widehat{n_2}$	5				
	$\widehat{\omega}$	0.5084				

Table 10 shows p-values for the KS, AD and CVM GoFTs calculated as follows. First, we obtain the values of the KS, AD and CvM test statistics (denoted ST) for true values of parameters $\widehat{\Theta}$ based on the sample $\left(x_{(1)}, x_{(2)}, ..., x_{(n)}\right)$. In the next step we simulate 10^4 samples $\left(x'_{(1)}, x'_{(2)}, ..., x'_{(n)}\right)$ from the given distribution with true values of parameters $\widehat{\Theta}$. For each sample, we calculate the values of the KS, AD and CvM test statistics (denoted ST^s). Finally, the p-value is calculated as $p \approx \#\left\{i : ST_i^s > ST\right\} 10^{-4}$. The CECK model is the best in terms of the AIC, BIC and HQIC values (see Table 6). This model has the highest p-values (see Table 7). Therefore, the CECK model fits better than the other models analyzed in this case.



(a) PDF



(b) CDF

Figure 10: Estimated PDF and CDF of analyzed distributions. Real data example

Table 10: TS values and p-values of the KS, AD and CvM tests. Real data example

Model	KS test		AD test		CVM test	
	TS	p-value	TS	p-value	TS	p-value
CECK	0.0406	0.9696	0.1984	0.9906	0.0285	0.9813
CN	0.0842	0.3193	0.7851	0.4905	0.1292	0.4616
CL	0.0872	0.2826	4.4573	0.0050	0.2858	0.1489
CC	0.1027	0.1382	2.9308	0.0264	0.3670	0.0835
CB	0.0798	0.3703	0.6034	0.6447	0.0990	0.5885

6. Conclusions

The paper presents the easily changeable kurtosis (ECK) distribution, the special cases of which is the uniform distribution. The new distribution, for large values of the parameter p, is similar (not identical) to the normal distribution.

The ECK belongs to the family of distributions with one mode, excess kurtosis values on the finite interval, the existing continuous function $p = f(\overline{\gamma}_2)$, where p is the shape parameter. The obtained results demonstrate that the ECK distribution can be extremely useful when we want to seamlessly test the GoFT's ability to detect deviations from normality by modeling negative excess kurtosis. We revealed that considered GoFTs are unable to distinct between the negatively skewed distribution (not only slightly but even moderately) and the Normal distribution. One can tentatively say that the above conclusion relates to numerous distributions similar to the new proposition.

Student and Fisher-Snedecor tests may be applied even as population distributions are of negative excess kurtosis, slight or indeed moderate one.

Real data example demonstrates that the $ECK(a, p)$ distribution in the mixed variant is flexible and competitive model that deserves to be added to the existing distributions in data modeling.

The information presented in the article shows that the proposed distribution deserves to be added to the symmetric distribution family.

Appendix

```
dECK=function(x,a,p)
{
l=exp(p*log(1-x*x/a/a)) #the sign of power is hard to edit in latex
m=a*beta(0.5,p+1)
if(p>=0) return(ifelse(abs(x)<=a,l/m,0))
if(p<0) return(ifelse(abs(x)<a,l/m,0))
}

pECK=function(x,a,p)
{
w=0.5+0.5*sign(x)*pbeta(x*x/a/a,0.5,p+1)
return(w)
}

q1=function(x,a,p,q)
{
w=pECK(x,a,p)-q
return(w)
}

qECK=function(a,p,q)
{
return(uniroot(q1, c(-a+0.001,a-0.001), tol = 0.0001, a=a, p=p, q=q)$root)
}

rECK=function (n,a,p)
{
rnd=numeric(n)
for (i in 1:n)
{
R=runif(n,0,1)
if (R>=0.5)
rnd[i]=a*sqrt(qbeta(2*R-1,0.5,p+1))
if (R<0.5)
rnd[i]=-a*sqrt(qbeta(1-2*R,0.5,p+1))
}
return(sort(rnd))
}
```

References

Azzalini A (1985). "A Class of Distributions which Includes the Normal Ones." *Scandinavian Journal of Statistics*, **12**(2), 171–178. URL https://www.jstor.org/stable/4615982.

Birnbaum ZW, Saunders SC (1969). "A New Family of Life Distributions." *Journal of Applied Probability*, **6**(2), 319–327. doi:https://doi.org/10.2307/3212003.

Gibbons JF, Mylroie S (1985). "Estimation of Impurity Profiles in Ion-implanted Amorphous Targets Using Joined Half-Gaussian Distributions." *Applied Physics Letters*, **22**(11), 568–569. doi:https://doi.org/10.1063/1.1654511.

Gradshteyn IS, Ryzhik IM (2014). *Table of Integrals, Series, and Products.* Academic Press.

Moors JJA (1988). "A Quantile Alternative for Kurtosis." *Journal of the Royal Statistical Society: Series D (The Statistician)*, **37**(1), 25–32. doi:https://doi.org/10.2307/2348376.

Renyi A (1961). "On Measures of Entropy and Information." *In Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, **4**(1), 547–561.

Reynolds PS (1994). *Time-series Analysis of Beaver Body Temperatures. Chapter 11 of Lange, N., Billard, L., Conquest, L., Ryan, L., Brillinger, D. and Greenhouse J. (Eds.), Case Studies in Biometry.* Wiley, New York.

Shannon CE (1948). "A Mathematical Theory of Communication." *Bell System Technical Journal*, **27**, 379–423. doi:10.1002/j.1538-7305.1948.tb01338.x.

Sulewski P (2019a). "Modified Lilliefors Goodness-of-fit Test for Normality." *Communications in Statistics - Simulation and Computation.* doi:https://doi.org/10.1080/03610918.2019.1664580.

Sulewski P (2019b). "Two-Piece Power Normal Distribution." *Communications in Statistics – Theory and Method*, **50**(11). doi:https://doi.org/10.1080/03610926.2019.1674871.

Tabass MS, Borzadaran GM, Amini M (2016). "Renyi Entropy in Continuous Case Is Not the Limit of Discrete Case." *Mathematical Sciences and Applications e-Notes*, **4**(1), 113–117. doi:https://doi.org/10.36753/mathenot.421418.

Torabi H, Montazeri NH, Grane A (1961). "A Test of Normality Based on the Empirical Distribution Function." *SORT*, **40**(1), 55–88. URL https://raco.cat/index.php/SORT/article/view/310069.

Tsallis C (1988). "Possible Generalizations of Boltzmann-Gibbs Statistics." *Journal of Statistical Physics*, **52**, 479–487. doi:https://doi.org/10.1007/BF01016429.

Wolfram (1988). *Hypergeometric0F1, Wolfram Language function.* URL https://reference.wolfram.com/language/ref/Hypergeometric0F1.html.

Affiliation:

Piotr Sulewski
Institute of Exact and Technical Sciences
Pomeranian University
Arciszewskiego Street 22, 76-200 Słupsk, Poland
E-mail: piotr.sulewski@apsl.edu.pl
URL: https://sulewski.apsl.edu.pl/

Austrian Journal of Statistics http://www.ajs.or.at/
published by the Austrian Society of Statistics http://www.osg.or.at/
Volume 52 *Submitted:* 2021-11-15
July 2023 *Accepted:* 2022-03-27