Article

Not peer-reviewed version

# An Extension of the Akash Distribution: Properties, Inference and Application

Yolanda M. Gómez , Luis Firinguetti-Limone * , Diego I Gallardo , Héctor W. Gómez

Posted Date: 19 October 2023

doi: 10.20944/preprints202310.1196.v1

Keywords: Akash distribution; Kurtosis; Maximum likelihood estimation; Slash distribution

*Preprints* (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 October 2023 doi:10.20944/preprints202310.1196.v1



*Article*

# An Extension of the Akash Distribution: Properties, Inference and Application

**Yolanda M. Gómez [1], Luis Firinguetti-Simone [1,*], Diego I. Gallardo [1] and Héctor W. Gómez [2]**

1 Departamento de Estadística, Facultad de Ciencias, Universidad del Bío-Bío, Concepción 4081112, Chile; ygomez@ubiobio.cl (Y.M.G.); lfiringu@ubiobio.cl (L.F.-S.); dgallardo@ubiobio.cl (D.I.G)
2 Departamento de Estadística y Ciencias de Datos, Facultad de Ciencias Básicas, Universidad de Antofagasta, Antofagasta 1240000, Chile; hector.gomez@uantof.cl (H.W.G.)
* Correspondence: lfiringu@ubiobio.cl

**Abstract:** In this article we introduce an extension of the Akash distribution. We use the slash methodology to make the kurtosis of the Akash distribution more flexible. We study the general density of this new distribution, some properties, moments, coefficients of asymmetry and kurtosis. Statistical inference is performed using the methods of moments and maximum likelihood via the EM algorithm. A simulation study is carried out to observe the behavior of the maximum likelihood estimator. An application to a real data set with high kurtosis is considered, where it is shown that the new distribution fits better than other extensions of the Akash distribution.

**Keywords:** Akash distribution; kurtosis; maximum likelihood estimation; slash distribution

## 1. Introduction

The slash distribution is an extension of the normal distribution. Its representation is the quotient between two independent random variables, one normal and the other a power of the uniform distribution. In this way we say that $S$ has a slash distribution if:

$$S = U_1/U_2, \quad (1)$$

where $U_1 \sim N(0,1)$, $U_2 \sim Beta(q,1)$, $U_1$ is independent of $U_2$ and $q > 0$, its representation can be seen in Johnson et al. [1]. This distribution has heavier tails than the normal distribution, that is, it has greater kurtosis. Properties of this family are discussed in Rogers and Tukey [2] and Mosteller and Tukey [3]. The maximum likelihood estimation of the location and scale parameters are discussed in Kafadar [4]. Wang and Genton [5] provide a multivariate version of the slash distribution and a multivariate skew version. Gomez et al. [6] and Gómez and Venegas [7] extend the slash distribution using the family of univariate and multivariate elliptic distributions. This methodology to increase the weight of the queues has also been used in distributions with positive support, for example, Gómez et al. [8] in the Birnbaum-Saunders distribution, Olmos et al. [9,10] in the half-normal and generalized half-normal distributions, Astorga et al. [11] in the Muth power distribution and Rivera et al. [12] in the Rayleigh distribution, among others.

Based on the work of Rivera et al. [12], the scale mixture of Rayleigh (SMR) distribution is introduced. We say that $Y \sim SMR(\theta, q)$ with $\theta > 0$ and $q > 0$. Then the probability density function (pdf) of $Y$ is

$$f_Y(y;\theta,q) = \frac{q\,y}{2\theta\left(\frac{y^2}{2\theta}+1\right)^{\frac{q}{2}+1}}, \quad y > 0\,. \quad (2)$$

Also, a necessary distribution in the development of this paper is the gamma distribution, whose pdf is given by

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 October 2023 doi:10.20944/preprints202310.1196.v1



$$g(t;a,b) = \frac{b^a}{\Gamma(a)} t^{a-1} e^{-bt}, \tag{3}$$

where $a, b, t > 0$. Its corresponding cumulative distribution function (cdf) is denoted by:

$$G(z;a,b) = \int_0^z g(t;a,b)dt \tag{4}$$

Shanker [13] introduced the Akash distribution and applied it to real lifetime data sets from medical science and engineering. Thus, we say that a random variable $Y$ has an Akash distribution (AK) with shape parameter $\theta$ if its pdf is given by

$$f_Y(y;\theta) = \frac{\theta^3}{\theta^2+2}(1+y^2)\exp(-\theta y), \tag{5}$$

where $\theta, y > 0$ and we denote it by $Y \sim AK(\theta)$. The parameter $\theta$ is a shape parameter, and if we add a scale parameter the pdf is given by

$$f_Y(y;\sigma,\theta) = \frac{\theta^3}{\sigma(\theta^2+2)}(1+y^2/\sigma^2)\exp(-\theta y/\sigma), \tag{6}$$

where $\sigma > 0$ is a scale parameter, $\theta > 0$ is a shape parameter and we denote it by $Y \sim AK(\sigma, \theta)$.

Extensions of the AK distribution are carried out by Shanker and Shukla [14,15], among others. Both extensions consider adding a parameter and we will compare them with the new distribution. The two-parameter Akash distribution (TPAD) introduced by Shanker and Shukla [14] has the following pdf:

$$f_Y(y;\theta,\alpha) = \frac{\theta^3}{\alpha\theta^2+2}(\alpha+y^2)\exp(-\theta y), \tag{7}$$

where $\theta, \alpha, y > 0$ and we denote it by $Y \sim TPAD(\theta, \alpha)$.
The power Akash distribution (PAD), introduced by Shanker and Shukla [15], has the following pdf:

$$f_Y(y;\theta,\alpha) = \frac{\alpha\theta^3}{\theta^2+2}(1+\alpha y^{2\alpha})y^{\alpha-1}\exp(-\theta y^\alpha), \tag{8}$$

where $\theta, \alpha, y > 0$ and we denote it by $Y \sim PAD(\theta, \alpha)$.

The main objective of this paper is to introduce an extension of the AK distribution given in (6), making use of the slash methodology, in order to obtain a new distribution with greater kurtosis to be able to accomcodate outliers.

The paper evolves as follows: In Section 2 we deliver the new distribution and present its properties. In Section 3 we perform inference using the method of moments and maximum likelihood via the EM algorithm, a simulation study is also carried out. In Section 4 we apply the distribution to a real data set and compare it with other extensions of the AK distribution. In Section 5 we provide some conclusions.

## 2. New density and its properties

In this section we introduce the representation, density and properties of the new distribution.

### 2.1. Representation

The representation of this new distribution is given by

$$X = \frac{Y}{Z}, \tag{9}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 October 2023 doi:10.20944/preprints202310.1196.v1



where $Y \sim AK(\theta)$, $Z \sim Beta(q,1)$, $Y$ and $Z$ are independent random variables with $\theta, q > 0$. We name the distribution of $X$ slash AK (SAK) and denote it by $X \sim SAK(\theta, q)$.

### 2.2. Density function

The following Proposition shows the pdf of the SAK distribution is generated using the representation given in (9).

**Proposition 1.** *Let $X \sim SAK(\theta, q)$. Then, the pdf of $X$ is given by*

$$f_X(x;\theta,q) = \frac{q^2\Gamma(q)x^{-(q+1)}}{(\theta^2+2)\theta^q}\left\{\theta^2 G(\theta x; q+1,1) + (q+1)(q+2)G(\theta x; q+3,1)\right\}, \tag{10}$$

*where $\theta, q, x > 0$ and $G$ is the cdf of the gamma distribution given in (4).*

**Proof.** Using the representation given in (9) and procedures based on the Jacobian method, we get the result. □

**Observation 1.** *Table 1 and Figure 1 show that as the value of the parameter q diminishes, the weight of the right tail increases.*

In particular, Table 1 compares $P(X > x)$ in the AK and SAK distributions for different values of x.

**Table 1.** Tail comparisons of the AK and SAK distributions.

| Distribution | $P(X > 5)$ | $P(X > 10)$ | Distribution | $P(X > 15)$ | $P(X > 20)$ |
|---|---|---|---|---|---|
| SAK(1,1) | 0.443 | 0.233 | SAK(0.5,1) | 0.367 | 0.278 |
| SAK(1,5) | 0.162 | 0.015 | SAK(0.5,5) | 0.063 | 0.020 |
| SAK(1,10) | 0.120 | 0.005 | SAK(0.5,10) | 0.034 | 0.007 |
| AK(1) | 0.085 | 0.002 | AK(0.5) | 0.018 | 0.003 |

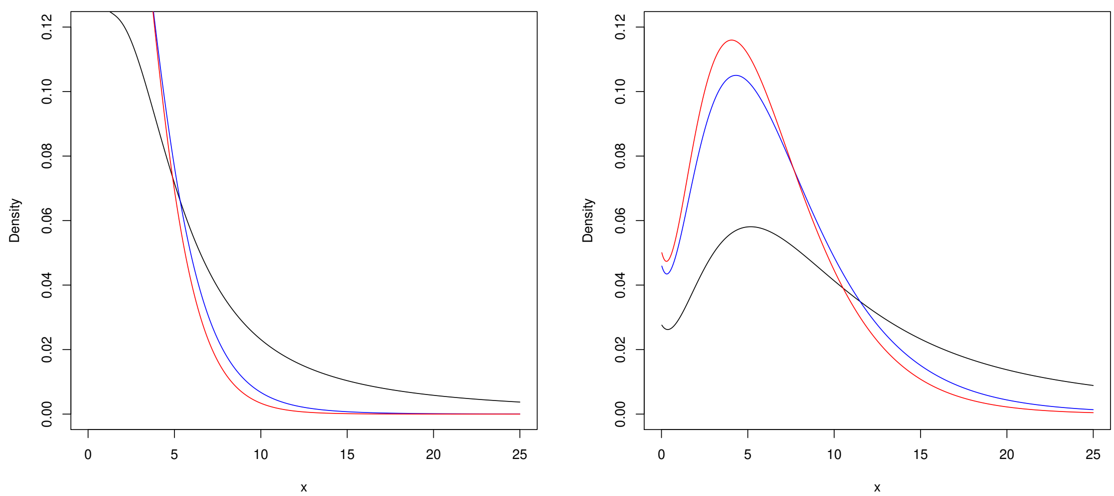


**Figure 1.** Left side: Examples of the SAK(1, 1) (in black), SAK(1, 5) (in blue), SAK(1, 10) (in red). Right side: Examples of the SAK(0.5, 1) (in black), SAK(0.5, 5) (in blue), SAK(0.5, 10) (in red).

### 2.3. Properties

The following Proposition gives the cdf in closed form. It depends on $G$, which is the cdf of the gamma distribution given in (4).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 October 2023 doi:10.20944/preprints202310.1196.v1



**Proposition 2.** *Let $X \sim SAK(\theta, q)$. Then, the cdf of X is given by*

$$F_X(x;\theta,q) = \frac{(\theta^2 + 2G(\theta x;3,1))(\theta x)^q - \theta^3 q\Gamma(q)G(\theta x;q,1) - \Gamma(q+3)G(\theta x;q+3,1)}{(\theta^2+2)(\theta x)^q}, \tag{11}$$

*where $\theta, q, x > 0$ and G is given in (4).*

**Proof.** The result follows from a direct application of the definition of a cdf. □

2.3.1. Reliability analysis

The reliability function $r(t) = 1 - F(t)$ and the hazard function $h(t) = \frac{f(t)}{r(t)}$ of the SAK distribution are given in the following corollary.

**Corollary 1.** *Let $T \sim SAK(\theta, q)$. Then, the $r(t)$ and $h(t)$ of T are given by*

1. $r(t) = 1 - \frac{(\theta^2+2G(\theta t;3,1))(\theta t)^q - \theta^3 q\Gamma(q)G(\theta t;q,1) - \Gamma(q+3)G(\theta t;q+3,1)}{(\theta^2+2)(\theta t)^q}$,
2. $h(t) = \frac{q^2\Gamma(q)\left(\theta^2 G(\theta t;q+1,1)+(q+1)(q+2)G(\theta t;q+3,1)\right)}{t(2(1-G(\theta t;3,1))(\theta t)^q+\theta^3 q\ Gamma(q)G(\theta t;q,1)-\Gamma(q+3)G(\theta t;q+3,1))}$,

*where $\theta, q > 0$.*

In Figure 2, we introduce the Hazard function of the SAK distribution for several values of $\sigma$ and $q$.

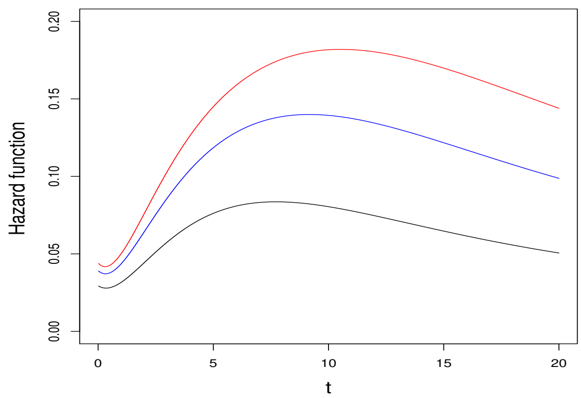


**Figure 2.** Hazard function of the SAK(0.5, 1) distribution (in black), SAK(0.5, 2) distribution (in blue), SAK(0.5, 3) distribution (in red).

2.3.2. Right tail of the SAK distribution

According to Rolski et al. [16] a distribution has a heavy right tail if

$$\limsup_{t\to\infty}\left(-\frac{\log r(t)}{t}\right) = 0.$$

The following result shows that the SAK distribution is heavy-tailed.

**Proposition 3.** *The distribution of the random variable $T \sim SAK(\theta, q)$ is heavy-tailed.*

**Proof.** Applying L'Hospital's rule twice we have,

$$\begin{aligned}\limsup_{t\to\infty}\left(-\frac{\log r(t)}{t}\right) &= \limsup_{t\to\infty}\left(\frac{f_T(t;\sigma,q)}{1-F_T(t;\sigma,q)}\right)\\ &= \limsup_{t\to\infty}\left(\frac{q+1}{t}-\frac{\theta^3 g(\theta t;q+1,1)+(q+1)(q+2)\theta g(\theta t;q+3,1)}{\theta^2 G(\theta t,q+1,1)+(q+1)(q+2)G(\theta t,q+3,1)}\right)\\ &= 0.\end{aligned}$$

□

The following Proposition shows that the SAK distribution can be represented as a scale mixture between the AK and Beta distributions.

**Proposition 4.** *If $X|Z=z\sim AK(z^{-1},\theta)$ and $Z\sim Beta(q,1)$ then $X\sim SAK(\theta,q)$.*

**Proof.** The marginal pdf of $X$ is given by

$$f_X(x;\theta,q)=\int_0^1 f_{X|Z}(x)f_Z(z)\,dz=\frac{\theta^3}{\theta^2+2}\int_0^1 z(1+z^2x^2)\exp(-\theta zx)qz^{q-1}\,dz,$$

and using (3), (4) and (5) , this result is obtained. □

The following result shows that when the parameter $q$ tends to infinity, the AK distribution is obtained.

**Proposition 5.** *Let $X\sim SAK(\theta,q)$. If $q\to\infty$ then $X$ converges in law to a random variable $Y\sim AK(\theta)$.*

**Proof.** Using its representation $X=\frac{Y}{Z}$ we analyze the convergence of this quotient, where $Y\sim AK(\theta)$ and $Z\sim$Beta$(q,1)$. In the Beta$(q,1)$ distribution we have, $\mathbb{E}[Z]=\frac{q}{1+q}$ and $Var[Z]=\frac{q}{(q+2)(q+1)^2}$. Then, applying Chebychev's inequality for $Z$, we have $\forall\epsilon>0$

$$P\left[|Z-\mathbb{E}[Z]|>\epsilon\right]\leq\frac{Var(Z)}{\epsilon^2}=\frac{q}{(q+2)(q+1)^2\epsilon^2}. \tag{12}$$

If $q\to\infty$ then the right hand side of (12) tends to zero, i.e. $W=Z-\mathbb{E}[Z]$ converges in probability to 0. Also $\mathbb{E}[Z]=\frac{q}{1+q}\longrightarrow 1$, $q\to\infty$, then we have,

$$Z=W+\mathbb{E}[Z]\xrightarrow{\mathcal{P}}1,\ q\to\infty.$$

Since $Y\sim AK(\theta)$, applying Slutsky's Lemma to $X=\frac{Y}{Z}$, we have

$$X\xrightarrow{\mathcal{L}}Y\sim AK(\theta),\qquad q\to\infty.$$

Thus, for increasing values of $q$, $X$ converges in law to a $AK(\theta)$ distribution. □

2.3.3. Moments

In this subsection we obtain the moments of the SAK distribution. To achieve this aim, the next lemma will be useful.

**Lemma 1.** *Let $Y\sim AK(\sigma,\theta)$ with $\sigma,\theta>0$. For $r>0$, $\mathbb{E}[Y^r]$ exists if and only if $q>r$ and in this case*

$$\mathbb{E}[Y^r]\quad=\quad\frac{\sigma^r\left(r!\theta^2+(r+2)!\right)}{\theta^r(\theta^2+2)}. \tag{13}$$

**Proof.** The r-th moment of the random variable $V \sim AK(\theta)$ is given by Shanker [13], which is $\mathbb{E}(V^r) = \frac{r!\theta^2 + (r+2)!}{\theta^r(\theta^2+2)}$, then calculating the r-th moment of the random variable $Y = \sigma V$, where $\sigma$ is a parameter of scale, the result is obtained. □

The moments of a SAK distribution are given in the following Proposition 6,

**Proposition 6.** *Let $X \sim SAK(\theta, q)$ with $\sigma$ and $q > 0$. For $r > 0$, $\mathbb{E}[X^r]$ exists if and only if $q > r$ and in this case*

$$\mu_r = \mathbb{E}[X^r] = \frac{q\left(r!\theta^2 + (r+2)!\right)}{\theta^r(\theta^2+2)(q-r)}. \tag{14}$$

**Proof.** Using the representation given in the Proposition 4 and by Lemma 1, we get

$$\begin{aligned}\mu_r = \mathbb{E}[X^r] &= \mathbb{E}\left[\mathbb{E}\left(X^r|Z\right)\right] = \mathbb{E}\left[\frac{Z^{-r}\left(r!\theta^2 + (r+2)!\right)}{\theta^r(\theta^2+2)}\right] \\ &= \frac{r!\theta^2 + (r+2)!}{\theta^r(\theta^2+2)}\mathbb{E}\left[Z^{-r}\right] = \frac{r!\theta^2 + (r+2)!}{\theta^r(\theta^2+2)}\int_0^1 q z^{q-r-1} dz.\end{aligned}$$

Solving the integral gives the result. □

From Proposition 6, the explicit expression of the noncentral moments, $\mu_r = \mathbb{E}[X^r]$, for $r = 1, 2, 3, 4$ and the variance of $X \sim SAK(\theta, q)$, $Var(X)$ follow.

**Corollary 2.** *Let $X \sim SAK(\theta, q)$ with $\theta$ and $q > 0$. From (14), the following noncentral moments and the variance of $X$, $Var(X)$, are obtained*

$$\mu_1 = \frac{q\kappa_6}{\theta\kappa_2(q-1)}, \quad q > 1, \qquad \mu_2 = \frac{2q\kappa_{12}}{\theta^2\kappa_2(q-2)}, \; q > 2,$$

$$\mu_3 = \frac{6q\kappa_{20}}{\theta^3\kappa_2(q-3)}, \; q > 3, \quad \mu_4 = \frac{24q\kappa_{30}}{\theta^4\kappa_2(q-4)}, \; q > 4,$$

$$Var(X) = \frac{q\left[2\kappa_{12}\kappa_2(q-1)^2 - q\kappa_6^2(q-2)\right]}{\theta^2\kappa_2^2(q-1)^2(q-2)}, \; q > 2.$$

*where $\kappa_i = \theta^2 + i$.*

**Remark 1.** *Note that when $q \to \infty$, $Var(X) \to \frac{\theta^4+16\theta^2+12}{\theta^2(\theta^2+2)^2}$, which is the variance of an $AK(\theta)$ distribution.*

The next Corollary gives the asymmetry coefficient, $\sqrt{\beta_1}$, of a $SAK(\theta, q)$ model.

**Corollary 3.** *Let $X \sim SAK(\theta, q)$ with $\theta > 0$ and $q > 3$. Then the skewness coefficient of $X$ is:*

$$\sqrt{\beta_1} = \frac{2\sqrt{q-2}\left[3\kappa_{20}\kappa_2^2(q-1)^3(q-2) - 3q\kappa_2\kappa_6\kappa_{12}(q-1)^2(q-3) + q^2\kappa_6^3(q-2)(q-3)\right]}{\sqrt{q}(q-3)\left[2\kappa_2\kappa_{12}(q-1)^2 - q(q-2)\kappa_6^2\right]^{3/2}}$$

**Proof.** Recall that

$$\sqrt{\beta_1} = \frac{\mathbb{E}[(X-\mathbb{E}(X))^3]}{(Var(X))^{3/2}} = \frac{\mu_3 - 3\mu_1\mu_2 + 2\mu_1^3}{(\mu_2 - \mu_1^2)^{3/2}},$$

where $\mu_1$, $\mu_2$ and $\mu_3$ were given in Corollary 2. □

Also, the kurtosis coefficient, $\beta_2$, of a $SAK(\theta,q)$ distribution is given in the following Corollary.

**Corollary 4.** *Let $X \sim SAK(\theta,q)$ with $\theta > 0$ and $q > 4$. Then the kurtosis coefficient of X is*

$$\beta_2 = \frac{3(q-2)\left(8\kappa_2^3\kappa_{30}q_1 - 8q\kappa_6\kappa_{20}\kappa_2^2 q_2 + 4q^2\kappa_6^2\kappa_{12}\kappa_2 q_3 - q^3\kappa_6^4 q_4\right)}{q(q-3)(q-4)\left[2\kappa_{12}\kappa_2(q-1)^2 - q\kappa_6^2(q-2)\right]^2}.$$

*where $q_1 = (q-1)^4(q-2)(q-3)$, $q_2 = (q-1)^3(q-2)(q-4)$, $q_3 = (q-1)^2(q-3)(q-4)$ and $q_4 = (q-2)(q-3)(q-4)$.*

**Proof.** Recall that

$$\beta_2 = \frac{\mathbb{E}[(X-\mathbb{E}(X))^4]}{(Var(X))^2} = \frac{\mu_4 - 4\mu_1\mu_3 + 6\mu_1^2\mu_2 - 3\mu_1^4}{(\mu_2-\mu_1^2)^2},$$

where $\mu_1$, $\mu_2$, $\mu_3$, and $\mu_4$ were given in Corollary 2. □

**Remark 2.** *It can be verified that for $q \to \infty$ skewness and kurtosis coefficients converge to $\frac{2(\theta^6+30\theta^4+36\theta^2+24)}{(\theta^4+16\theta^2+12)^{3/2}}$ and $\frac{3(3\theta^8+128\theta^6+408\theta^4+576\theta^2+240)}{(\theta^4+16\theta^2+12)^2}$ respectively, which coincide with the corresponding coefficients for the $AK(\theta)$ distribution (see Shanker, 2015).*

The results of Table 2 show that the values of the skewness and kurtosis coefficients depend on the parameters $\theta$ and $q$ and that as $q$ decreases, the skewness and kurtosis coefficients increase. On the other hand, as $q$ increases, the skewness and kurtosis coefficients are those of the AK($\theta$) distribution (Proposition 5).

**Table 2.** Skewness and kurtosis of the SAK distribution for various values of the shape parameters.

| $\theta$ | q | $\sqrt{\beta_1}$ | $\beta_2$ |
|---|---|---|---|
| 0.5 | 5 | 1.974 | 16.574 |
| 1 | | 1.952 | 15.180 |
| 0.5 | 6 | 1.570 | 9.039 |
| 1 | | 1.596 | 8.650 |
| 0.5 | 7 | 1.391 | 7.009 |
| 1 | | 1.438 | 6.863 |
| 0.5 | 10 | 1.201 | 5.460 |
| 1 | | 1.271 | 5.470 |
| 0.5 | 100 | 1.085 | 4.788 |
| 1 | | 1.166 | 4.837 |
| 0.5 | $\infty$ | 1.084 | 4.785 |
| 1 | | 1.165 | 4.834 |

## 3. Inference

In this section we study the estimation the parameters by the method of moments and ML via the EM algorithm. We also carry out some simulations to study the behavior of the ML estimators.

### 3.1. *Method of moment estimators*

Let $X_1, ..., X_n$ be a random sample from $X \sim SAK(\theta,q)$. Consider the first two sample moments, denoted by $\overline{X} = \frac{\sum_{i=1}^n X_i}{n}$ and $\overline{X^2} = \frac{\sum_{i=1}^n X_i^2}{n}$, respectively.

*Preprints* (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 October 2023 doi:10.20944/preprints202310.1196.v1



**Proposition 7.** *Given $X_1, \ldots, X_n$ a random sample from $X \sim SAK(\theta, q)$ with $q > 2$, the moment method estimators of $\theta$ and $q$ are*

$$\hat{q}_M = \frac{\overline{X}\hat{\theta}_M(\hat{\theta}_M^2 + 2)}{\hat{\theta}_M(\hat{\theta}_M^2 + 2)\overline{X} - \hat{\theta}_M^2 - 6}, \tag{15}$$

$$\overline{X^2}\hat{\theta}_M \left[2(\hat{\theta}_M^2 + 6) - \hat{\theta}_M\overline{X}(\hat{\theta}_M^2 + 2)\right] - 2\overline{X}(\hat{\theta}_M^2 + 12) = 0 \ , \tag{16}$$

*where (16) must be solved numerically to obtain $\hat{\theta}_M$. Then $\hat{\theta}_M$ must be replaced in (15) to get $\hat{q}_M$.*

**Proof.** Consider the method of moment equations

$$\mathbb{E}[X] = \frac{q(\theta^2 + 6)}{\theta(\theta^2 + 2)(q - 1)} = \overline{X} \tag{17}$$

$$\mathbb{E}[X^2] = \frac{2q(\theta^2 + 12)}{\theta^2(\theta^2 + 2)(q - 2)} = \overline{X^2} \tag{18}$$

Solving the equation (17) for the parameter $q$ we obtain (15). Then the value of $\hat{q}_M$ is substituted into the equation (18) and the equation given in (16) is obtained. □

*3.2. ML estimation*

Let $X_1, \ldots, X_n$ be a random sample from $X \sim SAK(\theta, q)$. Then the log-likelihood function is

$$l(\theta, q) = c(\theta, q) - (q + 1)\sum_{i=1}^{n}\log(x_i) + \sum_{i=1}^{n}\log\left[\theta^2 G(\theta x_i; q + 1, 1) + (q + 1)(q + 2)G(\theta x_i; q + 3, 1)\right]$$

where $c(\theta, q) = 2n\log(q) + n\log(\Gamma(q)) - n\log(\theta^2 + 2) - nq\log(\theta)$. Taking partial derivatives in $l(\theta, q)$ with respect to $\theta$ and $q$ and setting them equal to zero, we get

$$\sum_{i=1}^{n}\frac{2\theta G(\theta x_i; q + 1, 1) + \theta^2 J(x_i, q + 1) + (q + 1)(q + 2)J(x_i, q + 3)}{\theta^2 G(\theta x_i; q + 1, 1) + (q + 1)(q + 2)G(\theta x_i; q + 3, 1)} = \frac{2n\theta}{\theta^2 + 2} + \frac{nq}{\theta},$$

$$\sum_{i=1}^{n}\frac{\theta^2 H(x_i; q + 1) + (2q + 3)G(\theta x_i; q + 3, 1) + (q + 1)(q + 2)H(x_i; q + 3)}{\theta^2 G(\theta x_i; q + 1, 1) + (q + 1)(q + 2)G(\theta x_i; q + 3, 1)} = \eta(\theta, q) - \sum_{i=1}^{n}\log(x_i),$$

where $J(x_i, m) = x_i g(\theta x_i; m, 1)$, $H(x_i; v) = \int_0^{\theta x_i}\log(t)g(t; v, 1)dt - \psi(v)G(\theta x_i; v, 1)$ and $\eta(\theta, q) = \frac{2n}{q} + n(\psi(q) - \log(\theta))$. Solving numerically this system of equations to find the ML estimates may be a difficult task due to the functions it involves. However, an EM algorithm can be implemented (see Dempster et al. [17]) to obtain the ML estimates. The following subsection is dedicated to achieving this goal.

*3.3. EM Algorithm*

An alternative stochastic representation for the SAK model is given by

$$X_i \mid U_i = u_i, Z_i = z_i \sim G(1 + 2u_i, \theta z_i),$$

$$U_i \sim \text{Bern}\left(\frac{2}{\theta^2 + 2}\right),$$

$$Z_i \sim \text{Beta}\,(q, 1)\,.$$

where $U_i$ and $Z_i$, $i = 1, \ldots, n$, represent non-observable variables. This representation can be used for an alternative estimation procedure based on the EM algorithm (Dempster et al. [17]). In this context, the observed data are given by $\boldsymbol{D}_o = \boldsymbol{x}^\top$, where $\boldsymbol{x}^\top = (x_1, \ldots, x_n)$. The vectors $\boldsymbol{z}^\top = (z_1, \ldots, z_n)$ and

 doi:10.20944/preprints202310.1196.v1



$\boldsymbol{u}^\top = (u_1, \dots, u_n)$ are the latent variables and the vector $\boldsymbol{D}_c = (\boldsymbol{x}^\top, \boldsymbol{z}^\top, \boldsymbol{u}^\top)^\top$ are the complete data. Note that the joint distribution of $(X_i, U_i, Z_i)$ is given by

$$\begin{aligned} f(x_i, u_i, z_i) &= f(x_i \mid u_i, z_i) \times f(u_i) \times f(z_i) \\ &= \frac{(\theta z_i)^{1+2u_i}}{\Gamma(1+2u_i)} x_i^{2u_i} e^{-\theta z_i x_i} \times \left(\frac{2}{\theta^2+2}\right)^{u_i} \left(\frac{\theta^2}{\theta^2+2}\right)^{1-u_i} \times q z_i^{q-1} \\ &= \frac{q\theta^3 z_i^{2u_i+q} 2^{u_i}}{(\theta^2+2)\Gamma(1+2u_i)} x_i^{2u_i} e^{-\theta z_i x_i}. \end{aligned}$$

Therefore, up to a constant that does not depend on the vector of parameters $\psi = (\theta, q)$, the complete log-likelihood function for the model is given by

$$\ell_c(\psi; D_c) = n\left[\log q + 3\log\theta - \log(\theta^2+2)\right] + \sum_{i=1}^n \left[q \log z_i - \theta x_i z_i\right].$$

With this, the expected value of $\ell_c(\psi; D_c)$, given the observed data, is

$$Q(\psi \mid \psi^{(k)}) = n\left[\log q + 3\log\theta - \log(\theta^2+2)\right] + \sum_{i=1}^n \left[q\widehat{\kappa}_i^{(k)} - \theta x_i \widehat{z}_i^{(k)}\right],$$

where $\widehat{z}_i^{(k)} = \mathbb{E}(Z_i \mid x_i, \psi = \widehat{\psi}^{(k)})$ and $\widehat{\kappa}_i^{(k)} = \mathbb{E}(\log Z_i \mid x_i, \psi = \widehat{\psi}^{(k)})$. Note that

$$f(z_i, u_i \mid x_i) \propto \underbrace{\frac{(\theta x_i)^{2u_i+q+1}}{\Gamma(2u_i+q+1)} \frac{z_i^{(2u_i+q+1)-1} e^{-\theta x_i z_i}}{G(1; 2u_i+q+1, \theta x_i)}}_{Z_i \mid u_i, x_i \sim TG_{(0,1)}(2u_i+q+1, \theta x_i)} \times \underbrace{\frac{\Gamma(2u_i+q+1)}{\Gamma(2u_i+1)} \left(\frac{2}{\theta^2}\right)^{u_i} G(1; 2u_i+q+1, \theta x_i)}_{U_i \mid x_i \sim Bern(\nu_i)}, \quad (19)$$

where $\nu_i = \Gamma(q+3)G(\theta x_i; q+3)/[\theta^2\Gamma(q+1)G(\theta x_i; q+1) + \Gamma(q+3)G(\theta x_i; q+3)]$, $G(x; a) = \int_0^x \frac{1}{\Gamma(a)} t^{a-1} e^{-t} dt$ is the cdf for the gamma model and $\mathrm{TG}_{(0,1)}(a, b)$ denotes de gamma distribution with shape $a$ and rate $b$ truncated in the interval (0,1). Therefore, using properties of conditional expectations, we have $\mathbb{E}(Z_i \mid x_i) = \mathbb{E}\left(\mathbb{E}(Z_i \mid U_i, x_i) \mid x_i\right)$ and by (19) such expectations are simple to be computed. In a similar manner, we can compute $\mathbb{E}(\log Z_i \mid x_i)$, obtaining as results

$$\mathbb{E}(Z_i \mid x_i) = \frac{\nu_i (q+3) G(\theta x_i, q+4)}{\theta x_i G(\theta x_i, q+3)} + \frac{(1-\nu_i)(q+1) G(\theta x_i, q+2)}{\theta x_i G(\theta x_i, q+1)}, \quad (20)$$

$$\begin{aligned} \mathbb{E}(\log Z_i \mid x_i) &= \frac{\nu_i}{\Gamma(q+3) G(1; q+3, \theta x_i)} \int_0^{\theta x_i} \log\left(\frac{w_i}{\theta x_i}\right) w_i^{q+2} e^{-w_i} dw_i \\ &\quad + \frac{(1-\nu_i)}{\Gamma(q+1) G(1; q+1, \theta x_i)} \int_0^{\theta x_i} \log\left(\frac{w_i}{\theta x_i}\right) w_i^{q} e^{-w_i} dw_i. \end{aligned} \quad (21)$$

Therefore, the $k$th iteration of the algorithm comprises the following steps:

- E-step: Given $\widehat{\theta}^{(k-1)}$ and $\widehat{q}^{(k-1)}$, for $i = 1, \dots, n$ compute $\widehat{z}_i^{(k)}$ and $\widehat{\kappa}_i^{(k)}$ using equations (20) and (21), respectively.
- M1-step: Update $\widehat{q}^{(k)}$ as

$$\widehat{q}^{(k)} = \frac{-n}{\sum_{i=1}^n \widehat{\kappa}_i^{(k)}}.$$

- M2-step: Update $\widehat{\theta}^{(k)}$ as the solution for the non-linear equation

$$\frac{3}{\theta} - \frac{2\theta}{\theta^2+2} = \frac{1}{n} \sum_{i=1}^n x_i \widehat{z}_i^{(k)}.$$

The E, M1 and M2 steps are repeated until convergence is obtained, i.e. until the maximum distance between the estimate obtained in two consecutive iterations is less than a specified value.

In the following subsection we run some simulations to study the behavior of the ML estimators.

### 3.4. Simulation study

Table 3 shows the empirical bias (bias), the average of the standard errors (SE), the root of the empirical mean squared error (RMSE), and the 95% coverage probability (CP) based on the asymptotic distribution for the ML estimators of the parameters. Table 3 shows that the performance of the estimator improves as n increases.

**Table 3.** Estimated bias, SE, RMSE and CP of the ML estimators of the parameters of the SAK model for different values of $n$

| | | | $n=30$ | | | | $n=50$ | | | | $n=100$ | | | | $n=200$ | | | | $n=500$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\theta$ | $q$ | estimator | bias | SE | RMSE | CP | bias | SE | RMSE | CP | bias | SE | RMSE | CP | bias | SE | RMSE | CP | bias | SE | RMSE | CP |
| 0.5 | 0.5 | $\widehat{\theta}$ | -0.002 | 0.119 | 0.124 | 0.914 | -0.004 | 0.092 | 0.094 | 0.930 | -0.001 | 0.065 | 0.066 | 0.937 | 0.000 | 0.046 | 0.046 | 0.946 | 0.000 | 0.029 | 0.029 | 0.947 |
| | | $\widehat{q}$ | 0.036 | 0.122 | 0.139 | 0.961 | 0.025 | 0.092 | 0.100 | 0.958 | 0.012 | 0.063 | 0.065 | 0.952 | 0.005 | 0.043 | 0.044 | 0.952 | 0.001 | 0.027 | 0.027 | 0.951 |
| | 1.0 | $\widehat{\theta}$ | -0.004 | 0.110 | 0.114 | 0.918 | -0.003 | 0.085 | 0.086 | 0.931 | -0.002 | 0.060 | 0.061 | 0.940 | -0.001 | 0.043 | 0.043 | 0.946 | 0.000 | 0.027 | 0.027 | 0.946 |
| | | $\widehat{q}$ | -0.159 | 0.236 | 0.253 | 0.924 | -0.112 | 0.161 | 0.171 | 0.929 | -0.087 | 0.108 | 0.115 | 0.939 | -0.059 | 0.074 | 0.081 | 0.948 | -0.046 | 0.046 | 0.051 | 0.948 |
| | 2.0 | $\widehat{\theta}$ | -0.003 | 0.105 | 0.107 | 0.931 | -0.003 | 0.081 | 0.082 | 0.939 | -0.002 | 0.057 | 0.058 | 0.940 | -0.001 | 0.040 | 0.041 | 0.945 | 0.000 | 0.025 | 0.026 | 0.947 |
| | | $\widehat{q}$ | -0.137 | 0.597 | 0.622 | 0.904 | -0.125 | 0.395 | 0.420 | 0.924 | -0.077 | 0.233 | 0.250 | 0.932 | -0.041 | 0.151 | 0.162 | 0.942 | -0.023 | 0.092 | 0.095 | 0.948 |
| 3.0 | 0.5 | $\widehat{\theta}$ | 0.136 | 1.063 | 1.236 | 0.891 | 0.095 | 0.794 | 0.861 | 0.915 | 0.035 | 0.537 | 0.556 | 0.927 | 0.013 | 0.373 | 0.380 | 0.940 | 0.005 | 0.234 | 0.235 | 0.947 |
| | | $\widehat{q}$ | 0.059 | 0.156 | 0.206 | 0.963 | 0.030 | 0.110 | 0.124 | 0.958 | 0.015 | 0.075 | 0.079 | 0.955 | 0.009 | 0.052 | 0.054 | 0.953 | 0.003 | 0.032 | 0.033 | 0.952 |
| | 1.0 | $\widehat{\theta}$ | 0.104 | 0.982 | 1.112 | 0.896 | 0.060 | 0.729 | 0.786 | 0.912 | 0.028 | 0.499 | 0.517 | 0.929 | 0.012 | 0.347 | 0.354 | 0.941 | 0.003 | 0.218 | 0.219 | 0.948 |
| | | $\widehat{q}$ | -0.087 | 0.398 | 0.446 | 0.892 | -0.057 | 0.245 | 0.296 | 0.925 | -0.021 | 0.145 | 0.188 | 0.938 | -0.012 | 0.097 | 0.117 | 0.948 | -0.002 | 0.060 | 0.066 | 0.947 |
| | 2.0 | $\widehat{\theta}$ | 0.145 | 0.976 | 1.070 | 0.922 | 0.068 | 0.709 | 0.747 | 0.929 | 0.018 | 0.478 | 0.491 | 0.934 | 0.006 | 0.332 | 0.339 | 0.941 | 0.000 | 0.208 | 0.210 | 0.946 |
| | | $\widehat{q}$ | -0.105 | 1.025 | 1.090 | 0.915 | -0.084 | 0.724 | 0.790 | 0.924 | -0.069 | 0.440 | 0.485 | 0.935 | -0.048 | 0.255 | 0.282 | 0.942 | -0.008 | 0.140 | 0.155 | 0.948 |
| 10.0 | 0.5 | $\widehat{\theta}$ | 0.595 | 4.688 | 5.331 | 0.882 | 0.291 | 3.484 | 3.709 | 0.901 | 0.126 | 2.400 | 2.470 | 0.925 | 0.088 | 1.684 | 1.706 | 0.942 | 0.019 | 1.056 | 1.049 | 0.944 |
| | | $\widehat{q}$ | 0.069 | 0.175 | 0.184 | 0.964 | 0.035 | 0.113 | 0.128 | 0.963 | 0.016 | 0.075 | 0.080 | 0.957 | 0.007 | 0.052 | 0.053 | 0.951 | 0.003 | 0.032 | 0.033 | 0.951 |
| | 1.0 | $\widehat{\theta}$ | 0.559 | 4.440 | 4.910 | 0.904 | 0.222 | 3.260 | 3.453 | 0.910 | 0.102 | 2.248 | 2.328 | 0.926 | 0.059 | 1.574 | 1.600 | 0.941 | 0.009 | 0.987 | 0.980 | 0.948 |
| | | $\widehat{q}$ | -0.097 | 0.508 | 0.631 | 0.899 | -0.051 | 0.284 | 0.389 | 0.903 | -0.031 | 0.152 | 0.199 | 0.939 | -0.023 | 0.098 | 0.117 | 0.948 | -0.012 | 0.060 | 0.080 | 0.948 |
| | 2.0 | $\widehat{\theta}$ | 0.885 | 4.575 | 4.757 | 0.935 | 0.389 | 3.286 | 3.316 | 0.937 | 0.172 | 2.209 | 2.217 | 0.944 | 0.035 | 1.533 | 1.546 | 0.947 | -0.006 | 0.955 | 0.955 | 0.947 |
| | | $\widehat{q}$ | -0.068 | 1.224 | 1.222 | 0.924 | -0.057 | 0.834 | 0.950 | 0.931 | -0.037 | 0.440 | 0.483 | 0.935 | -0.027 | 0.305 | 0.313 | 0.942 | -0.018 | 0.149 | 0.159 | 0.943 |

## 4. Application

In this section we analyze a real data set showing that the SAK distribution can be more appropriate than other commongly used distributions to model heavy right-tailed data. The data correspond to plasma beta-carotene levels (ng/ml) of 314 patients. This data set contains 14 variables and is available online at "http://Lib.stat.cmu.edu/datasets/Plasma Retinol". In this study, we consider the variable Betaplasma. The medical interest in this variable comes from the fact that low levels of plasma beta-carotene may be associated with higher risk of developing certain types of cancer. In Table 4 we present some descriptive statistics including the sample skewness, $b_1$, and sample and kurtosis $b_2$. We may observe high kurtosis in this data set.

**Table 4.** Descriptive statistics for the data set.

| $n$ | $\overline{x}$ | $s^2$ | $b_1$ | $b_2$ |
|---|---|---|---|---|
| 314 | 190.4968 | 33480.72 | 3.536562 | 16.8145 |

The moment estimates for the parameters of the SAK distribution are $\hat{\theta}_M = 0.025$ and $\hat{q}_M = 2.810$. These estimates are useful starting values, required to implement maximum likelihood estimation using numerical methods. Table 5 shows the ML estimates for the parameters of the PAD, SMR and SAK models. For each model we report the value of the log-likelihood. It can be seen that the SAK model presented a larger value of log-likelihood than the other models.

**Table 5.** ML estimates for PAD, SMR and SAK models.

| Parameter estimates | PAD (SE) | SMR (SE) | SAK (SE) |
|---|---|---|---|
| $\theta$ | 0.012 (0.003) | 16998.167 (3399.076) | 0.027 (0.002) |
| $\alpha$ | 1.052 (0.038) | – | – |
| $q$ | – | 2.926 (0.385) | 2.331 (0.294) |
| Log-likelihood | −1953.632 | −1910.472 | −1908.147 |

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 October 2023 doi:10.20944/preprints202310.1196.v1



In order to compare the fit of the distributions, we considered the usual Akaike information criterion (AIC), introduced by Akaike [18], and the Bayesian information criterion (BIC), proposed by Schwarz [19]. It is known that AIC=$2k - 2\log lik$ and BIC=$k\log n - 2\log lik$ where $k$ is the number of parameters in the model, $n$ is the sample size and $\log lik$ is the maximized value of the log-likelihood function. Table 6 shows the AIC and BIC for each model, indicating that the SAK distribution leads to a better fit than the other distributions. Figure 3 presents the histogram for the data together with the fitted densities.

**Table 6.** AIC and BIC for fitted models.

| Criterion | PAD | SMR | SAK |
|---|---|---|---|
| AIC | 3911.264 | 3824.944 | 3820.294 |
| BIC | 3918.763 | 3832.443 | 3827.793 |

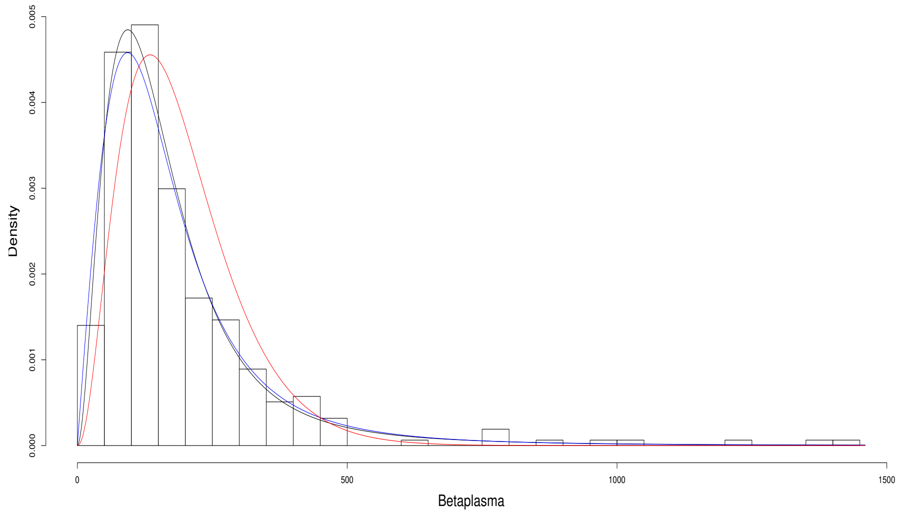


**Figure 3.** Betaplasma: Histogram and fitted PAD pdf (in red), SMR pdf (in blue) and SAK pdf (in black).

We also calculated the quantile residuals (QR). If the model is appropriate for the data, the QR should be a sample of the standard normal distribution (see Dunn and Smyth, [20]). This assumption can be validated with traditional normality tests, such as the Anderson-Darling (AD), Cramér-von Mises (CVM) and Shapiro-Wilkes (SW) tests. Figure 4 shows the qqplot of the quantile residuals of the three fitted distributions. All three tests suggest that the SAK model provides a better fit for this data set.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 October 2023 doi:10.20944/preprints202310.1196.v1



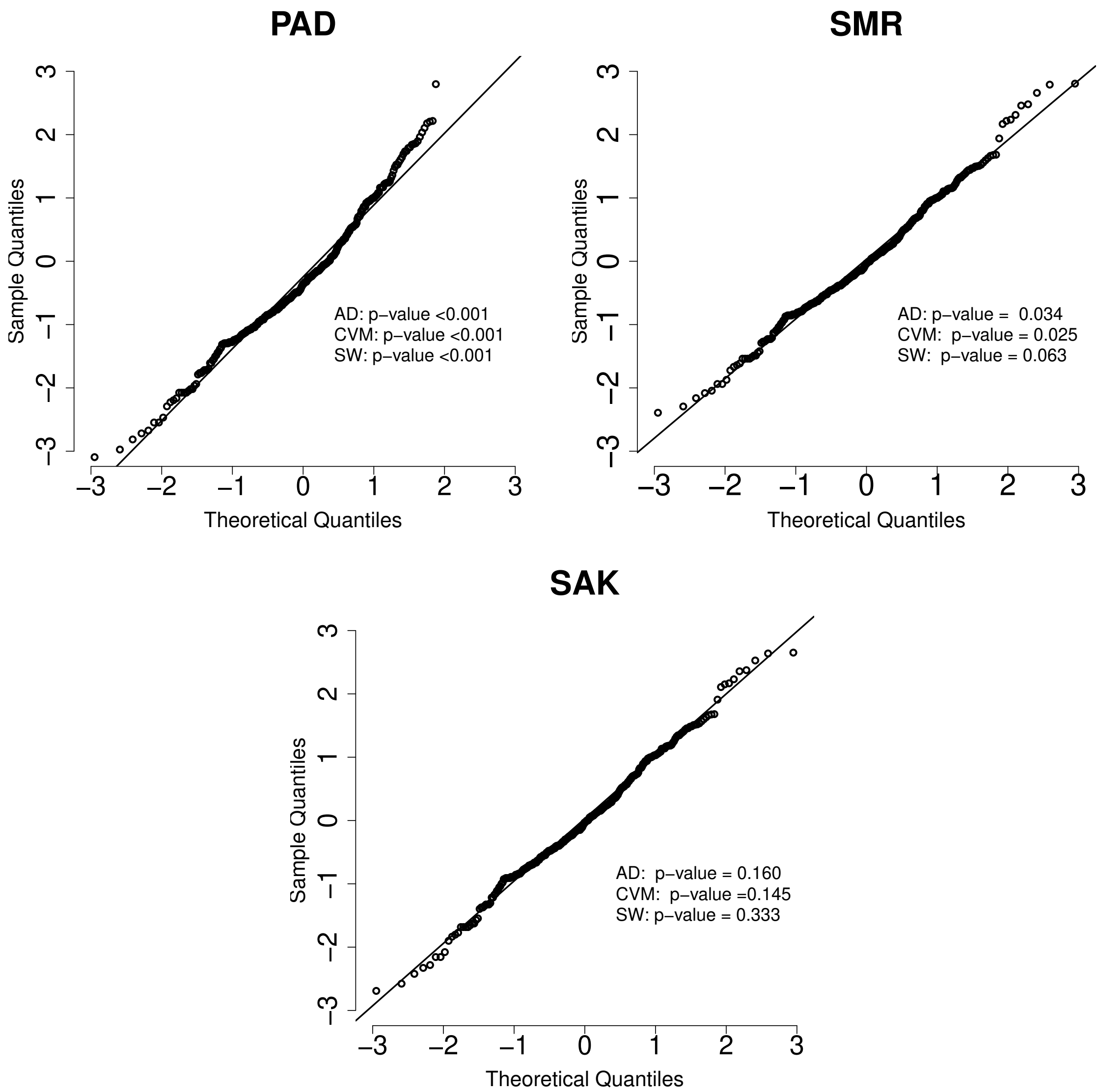


**Figure 4.** qqplots of the quantile residuals for the fitted models, together with the p-values of the AD, CVM and SW normality tests.

## 5. Discussion

This paper presents an extension of the AK distribution based on the slash methodology. Some properties of this new distribution are derived. It is also compared with two other distributions using a real data set. Estimation is done through ML via the EM algorithm. The new SAK distribution is an alternative to fit heavy-tailed right-skewed data. Additional features of the SAK distribution are:

- The distribution has two representations, one based on the quotient of two independent random variables and another based on a scale mixture between the AK and Beta distributions.
- The pdf, cdf and hazard function of the SAK distribution are explicit and are represented by the cdf of the gamma distribution.
- The distribution has a heavy right tail.
- The distribution contains the AK distribution as a limit, that is, when the parameter $q$ tends to infinity in the distribution SAK, the AK distribution is obtained.
- The moments and the coefficients of skewness and kurtosis are explicit.
- In the application, observing the AIC and BIC and the Anderson-Darling, Cramér-von Mises and Shapiro-Wilkes statistical tests, we may conclude that the SAK distribution fits the Betaplasma

 doi:10.20944/preprints202310.1196.v1



data set better than the PAD and SMR distributions, which are also extensions of the AK distribution.

## References


1. Jonhson, N.L., Kotz, S., and Balakrishnan, N.1995.*Continuous univariate distributions.* Vol 1, 2nd edn. New York: Wiley.
2. Rogers, W.H., Tukey, J.W. 1972. Understanding some long-tailed symmetrical distributions. *Statist. Neerlandica* 26 :211-226.
3. Mosteller, F., Tukey, J.W. 1977. Data analysis and regression. Addison-Wesley, Reading, MA.
4. Kafadar, K. 1982. A biweight approach to the one-sample problem. *J. Amer. Statist. Assoc.* 77 :416-424.
5. Wang, J., Genton, M.G. 2006. The multivariate skew-slash distribution. *Journal Statistical Planning and Inference* 136 :209-220.
6. Gómez, H.W., Quintana, F.A., Torres, F. J. 2007. A New Family of Slash-Distributions with Elliptical Contours. *Statistics and Probability Letters* 77(7):717-725.
7. Gómez, H.W., Venegas, O. 2008. Erratum to: A new family of slash-distributions with elliptical contours [Statist. Probab. Lett. 77 (2007) 717-725]. *Statistics and Probability Letters* 78 (14):2273-2274.
8. Gómez, H.W., Olivares-Pacheco, J.F., Bolfarine, H. 2009. An extension of the generalized birnbaun-saunders distribution. *Statistics and Probability Letters* 79 (3):331-338.
9. Olmos, N.M., Varela, H., Gómez, H.W. and Bolfarine, H. 2012. An extension of the half-normal distribution. *Statistical Papers* 53 :875-886.
10. Olmos, N.M., Varela, H., Bolfarine, H., Gómez, H.W. 2014. An extension of the generalized half-normal distribution. *Statistical Papers* 55:967-981.
11. Astorga, J.M., Reyes, J., Santoro, K.I., Venegas, O., Gómez, H.W. 2020. A Reliability Model Based on the Incomplete Generalized Integro-Exponential Function. *Mathematics* 8:1537.
12. Rivera P.A., Barranco-Chamorro I., Gallardo D.I., Gómez H.W. 2020. Scale Mixture of Rayleigh Distribution. *Mathematics* 8(10):1842.
13. Shanker, R. 2015. Akash Distribution and Its Applications. *International Journal of Probability and Statistics* 4(3):65-75.
14. Shanker, R., Shukla, K.K. 2017a. On Two-Parameter Akash Distribution. *Biometrics & Biostatistics International Journal* 6(5):00178.
15. Shanker, R., Shukla, K.K. 2017b. Power Akash Distribution and Its Application. *Journal of Applied Quantitative Methods* 12(3):1-10.
16. Rolski, T., H. Schmidli, V. Schmidt, and J. Teugel. 1999. Stochastic Processes for Insurance and Finance. John Wiley & Sons.
17. Dempster. A.P.; Laird. N.M.; Rubim. D.B. 1977. Maximum likelihood from incomplete data via the EM algorithm (with discussion). J. R. Stat. Soc. Ser. B, 39, 1–38.
18. Akaike, H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19(6)**, 716-723.
19. Schwarz, G. 1978. Estimating the dimension of a model. *Ann. Statist.*, **6**(2), 461-464.
20. Dunn, P.K., Smyth, G.K. 1996. Randomized Quantile Residuals. *Journal of Computational and Graphical Statistics*, 5(3), 236-244.