Workshop on Statistical Computing and Robust Inference for High Dimensional Data
Taiwan, 11 - 14 December 2023

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

Luis Firinguetti joint work with Diego Gallardo, Yolanda Gómez and Héctor Gómez

Universidad del Bío-Bío

December 2023

Outline

- Introduction
- Present a new distribution: The Slash Akash Distribution
- Properties of the distribution
- Estimation
- Simulations
- Application: Plasma beta-carotene
- Some conclusions

Introduction

Data sets with pronounced fluctuations are commonly encountered in such diverse disciplines as economic and actuarial sciences, environmental and earth sciences, among others. Thus, heavy-tailed models are necessary to perform better modelling in the presence of extreme values. For example, the normal distribution does not perform well in modelling data sets with extreme observations. We must therefore resort to heavy-tailed distributions. For example, in problems in which the r. v. involved have high kurtosis, the probability that a rare event occurs can be highly underestimated if a model without heavy tails is used. In the economy, practical examples of rare events are pandemics, and the 2008–09 financial crisis, to name a few. In geology, a rare event might be a mega earthquake or a sudden eruption of a volcano that has been dormant for centuries.

Introduction

The slash distribution is an extended version of the normal distribution. It is characterized by the ratio of two separate random variables: one following a normal distribution and the other following a power of the uniform distribution. Therefore, we define a slash distribution for variable S as:

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

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. [5]. The distribution in question exhibits heavier tails compared to the normal distribution, indicating a higher level of kurtosis.

Introduction

The characteristics of this particular distribution are explored in detail in the works of Rogers and Tukey [12] and Mosteller and Tukey [7]. Kafadar [6] delves into the topic of maximum likelihood estimation for the location and scale parameters. Wang and Genton [18] present a multivariate version of the slash distribution as well as a multivariate skew version. The slash distribution is further extended by Gomez and Venegas [4] through the incorporation of the multivariate elliptic distributions.

Introduction

This methodology to increase the weight of the queues has also been used in distributions with positive support. To name a few, we mention the works of Olmos et al. [9] in the half-normal and Rivera et al. [10] in the Rayleigh model, among others. Based on the work of Rivera et al. [10], the scale mixture of Rayleigh (SMR) model is proposed. We say that $Y \sim SMR(\theta, q)$ with $\theta > 0$ and $q > 0$ if the probability density function (pdf) of Y is

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

Introduction

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

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

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

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

Akash Distribution

Shanker [15] introduced the Akash distribution and applied it to real lifetime data sets from medical science and engineering. Thus, we say that a random variable (r.v.) Y has an Akash model (AK) with shape parameter θ if its pdf is

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

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

Extensions of the Akash Distribution

Extensions of the AK distribution are carried out by Shanker and Shukla [16, 17], 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 [16] 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 [17], 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 $Y \sim PAD(\theta,\alpha)$.

Alternative Distribution

The main motivation of this work is to introduce an extended version of the AK distribution given in Equation (6), making use of the slash methodology, in order to obtain a new distribution with greater kurtosis to be able to accommodate outliers.

The representation of this new distribution is given by

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

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

Density Function

Proposition

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

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

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

Table 1 and Figure 1 illustrate that the weight of the right tail increases. In particular, Table 1 shows $P(X > x)$ for different values of x in the mentioned distribution.

Tail Probability Comparison

Table: Tails comparison

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

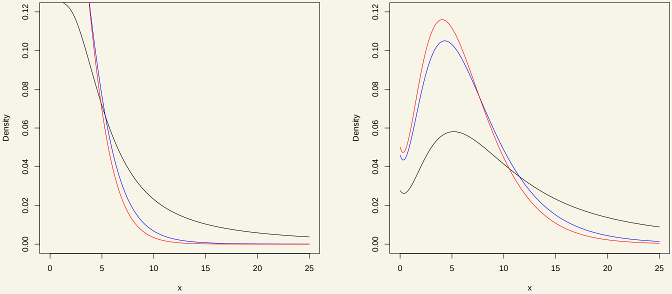

Figure: Left side: examples of the SAK$(1,1)$ (in black), SAK$(1,5)$ (in blue), SAK$(1,10)$ (in red). Right side: examples of the SAK$(0.5,1)$ (in black), SAK$(0.5,5)$ (in blue), SAK$(0.5,10)$ (in red).

Cumulative Density Function

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

Proposition

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

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

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

Reliability and Hazard Function

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

Corollary

The reliability and hazard functions of the $SAK(\theta, q)$ model 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 $\theta, q > 0$.

Hazard function

In Figure 2, we present the hazard function of the SAK model for several values of θ and q.

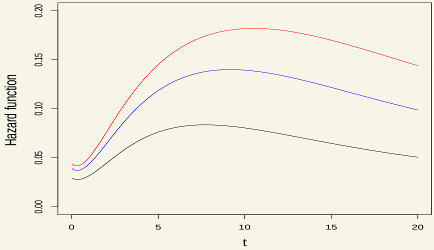

Figure: Hazard function of the SAK$(0.5, 1)$ distribution (in black), SAK$(0.5, 2)$ distribution (in blue), SAK$(0.5, 3)$ distribution (in red).

Right Tail of the SAK Distribution

From Nair et al. [8], a distribution has a heavy right tail if $\forall \mu > 0$,

$$\limsup_{x \to \infty} \left(\frac{r(x)}{e^{-\mu x}} \right) = \infty.$$

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

Proposition

The r.v. $X \sim SAK(\theta, q)$ is heavy-tailed.

Proof.

$$\limsup_{x \to \infty} \left(\frac{r(x)}{e^{-\mu x}} \right) \geq \limsup_{x \to \infty} \frac{\theta^3 q \gamma(\theta x, q) + \gamma(\theta x, q+3)}{e^{-\mu x}(\theta^2 + 2)(\theta x)^q} = \infty$$

Akash as a Special Case of the SAK Distribution

The following proposition illustrates that the AK model is a particular case of the SAK distribution for $q \to \infty$.

Proposition

Let $X \sim SAK(\theta, q)$ and $Y \sim AK(\theta)$. If $q \to \infty$, then X converges in law to Y.

Moments of the SAK Distribution

Proposition

Let $X \sim SAK(\theta, q)$ *with* θ, $q > 0$. *For* $r \in \mathbb{N}$, $\mathbb{E}[X^r]$ *is given by*

$$\mu_r = \mathbb{E}[X^r] = E\left[\left(\frac{Y}{Z}\right)^r\right] = \left[E(Y^r) \times E\left(\frac{1}{Z}\right)^r\right] \tag{12}$$

$$= \frac{q\,(r!\theta^2 + (r+2)!)}{\theta^r(\theta^2+2)(q-r)}, \quad \textit{provided that } q > r. \tag{13}$$

Moments of the SAK Distribution

Corollary

Let $X \sim SAK(\theta, q)$ with θ and $q > 0$. The 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 $\kappa_i = \theta^2 + i$.

Skewness

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

Corollary

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

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

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

Kurtosis

Corollary

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

$$\begin{aligned}\beta_2 &= \frac{\mathbb{E}[(X-\mathbb{E}(X))^4]}{(Var(X))^2} = \frac{\mu_4 - 4\mu_1\mu_3 + 6\mu_1^2\mu_2 - 3\mu_1^4}{(\mu_2-\mu_1^2)^2} \\ &= \frac{3(q-2)\left(8\kappa_2^3\kappa_{30}q_1 - 8q\kappa_6\kappa_{20}\kappa_2^2q_2 + 4q^2\kappa_6^2\kappa_{12}\kappa_2q_3 - q^3\kappa_6^4q_4\right)}{q(q-3)(q-4)\left[2\kappa_{12}\kappa_2(q-1)^2 - q\kappa_6^2(q-2)\right]^2}\end{aligned}$$

.

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 $q_4 = (q-2)(q-3)(q-4)$.

Skewness and Kurtosis

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

θ	q	$\sqrt{\beta_1}$	β_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	∞	1.084	4.785
1		1.165	4.834

Inference: Method of Moments

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

Proposition

Given $X_1, \ldots, X_n$ a random sample from $X \sim SAK(\theta, q)$ with $q > 2$, the moment method estimators of θ and q provides the following estimators

$$\widehat{q}_M = \frac{\overline{X}\widehat{\theta}_M(\widehat{\theta}_M^2 + 2)}{\widehat{\theta}_M(\widehat{\theta}_M^2 + 2)\overline{X} - \widehat{\theta}_M^2 - 6}, \tag{14}$$

$$\overline{X^2}\widehat{\theta}_M \left[2(\widehat{\theta}_M^2 + 6) - \widehat{\theta}_M\overline{X}(\widehat{\theta}_M^2 + 2)\right] - 2\overline{X}(\widehat{\theta}_M^2 + 12) = 0 \tag{15}$$

where it is necessary to solve Equation (15) numerically to obtain $\widehat{\theta}_M$. Then $\widehat{\theta}_M$ is replaced in Equation (14) to obtain $\widehat{q}_M$.

Inference: Maximum likelihood

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 $c(\theta, q) = 2n\log(q) + n\log(\Gamma(q)) - n\log(\theta^2 + 2) - nq\log(\theta)$.

Inference: Maximum likelihood

Taking partial derivatives in $l(\theta, q)$ in relation to θ and q and equaling those equations to zero, we obtain

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

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

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
$\eta(\theta, q) = \frac{2n}{q} + n(\psi(q) - \log(\theta))$.

Since solving this system of equations may be a difficult task, we resort to implement an EM algorithm (see Dempster et al. [2]).

EM Algorithm

Using a stochastic approach, the SAK model may be represented as follows:

$$X_i \mid U_i = u_i, Z_i = z_i \sim G(1 + 2u_i, \theta z_i),$$
$$U_i \sim \text{Bernoulli}\left(\frac{2}{\theta^2 + 2}\right),$$
$$Z_i \sim \text{Beta}\,(q, 1)\,.$$

where U_i and Z_i, $i = 1, \ldots, n$, represent non-observable variables.

This representation can be used to implement the EM algorithm (Dempster et al. [2]). In this context, the observed data are given by $D_o = x^\top$, where $x^\top = (x_1, \ldots, x_n)$. The vectors $z^\top = (z_1, \ldots, z_n)$ and $u^\top = (u_1, \ldots, u_n)$ are the latent variables and the vector $D_c = (x^\top, z^\top, u^\top)^\top$ are the complete data.

EM Algorithm

The joint distribution of (X_i, U_i, Z_i) is given by

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

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

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

EM Algorithm

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

EM Algorithm

Note that

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

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
- $\text{TG}_{(0,1)}(a, b)$ denotes the gamma distribution with shape a and rate b truncated in the interval (0,1).

EM Algorithm

Since

$$\mathbb{E}(Z_i \mid x_i) = \mathbb{E}\left(\mathbb{E}(Z_i \mid U_i, x_i) \mid x_i\right) = \frac{\nu_i (q+3) G(\theta x_i, q+4)}{\theta x_i G(\theta x_i, q+3)} + \frac{(1-\nu_i)(q+1) G(\theta x_i, q+2)}{\theta x_i G(\theta x_i, q+1)} \quad (16)$$

$$\mathbb{E}(\log Z_i \mid x_i) = \frac{\nu_i}{\Gamma(q+3) G(1; q+3, \theta x_i)} \int_0^{\theta x_i} \log\left(\frac{w_i}{\theta x_i}\right) w_i^{q+2} e^{-w_i} dw_i + \frac{(1-\nu_i)}{\Gamma(q+1) G(1; q+1, \theta x_i)} \int_0^{\theta x_i} \log\left(\frac{w_i}{\theta x_i}\right) w_i^{q} e^{-w_i} dw_i. \quad (17)$$

EM Algorithm

Therefore, the kth iteration of the algorithm comprises the following steps:

- E-step: given $\widehat{\theta}^{(k-1)}$ and $\widehat{q}^{(k-1)}$, for $i = 1, \ldots, n$ compute $\widehat{z}_i^{(k)}$ and $\widehat{\kappa}_i^{(k)}$ using Equations (16) and (17), 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 iterated until convergence is achieved.

A Simulation Study

To evaluate the performance of the ML estimators of the SAK (θ, q), obtained through the EM algorithm, we considered:

- Three different values for θ (0.5, 3, and 10),
- Three values for q (0.5, 1, and 2),
- Five sample sizes (30, 50, 100, 200, and 500).

For each combination of θ, q, and n, we will draw 1000 replicates and calculate the ML estimators.

The initial values to start the EM algorithm are:

- $\widehat{\theta}^{(0)}$ is obtained from the estimate of θ in the AK model (with scale fixed at 1)
- $\widehat{q}^{(0)} = 1$.

In addition, for each replicate we estimate the standard errors based on the observed information matrix. reports the empirical bias (bias),

A Simulation Study

We report the following statistics:

- The average of the estimated bias (Bias),
- The mean of the standard errors (SE),
- The square root of the mean squared error (RMSE)
- The 95% probability that the estimated parameters fall within the asymptotic distribution (CP).

Simulation Results

Table: Estimated Bias, SE, RMSE and CP of the ML Estimators of the Parameters of the SAK Distribution for $n = 30$.

θ	q	Estimator	Bias	SE	RMSE	CP
0.5	0.5	$\hat{\theta}$	−0.002	0.119	0.124	0.914
		$\hat{q}$	0.036	0.122	0.139	0.961
	1.0	$\hat{\theta}$	−0.004	0.110	0.114	0.918
		$\hat{q}$	−0.159	0.236	0.253	0.924
	2.0	$\hat{\theta}$	−0.003	0.105	0.107	0.931
		$\hat{q}$	−0.137	0.597	0.622	0.904
3.0	0.5	$\hat{\theta}$	0.136	1.063	1.236	0.891
		$\hat{q}$	0.059	0.156	0.206	0.963
	1.0	$\widehat{\theta}$	0.104	0.982	1.112	0.896
		$\hat{q}$	−0.087	0.398	0.446	0.892
	2.0	$\hat{\theta}$	0.145	0.976	1.070	0.922
		$\hat{q}$	−0.105	1.025	1.090	0.915
10.0	0.5	$\hat{\theta}$	0.595	4.688	5.331	0.882
		$\hat{q}$	0.069	0.175	0.184	0.964
	1.0	$\hat{\theta}$	0.559	4.440	4.910	0.904
		$\hat{q}$	−0.097	0.508	0.631	0.899
	2.0	$\hat{\theta}$	0.885	4.575	4.757	0.935
		$\hat{q}$	−0.068	1.224	1.222	0.924

Simulation Results

Table: Estimated Bias, SE, RMSE and CP of the ML Estimators of the Parameters of the SAK Distribution for $n = 50$.

θ	q	Estimator	Bias	SE	RMSE	CP
0.5	0.5	$\widehat{\theta}$	-0.004	0.092	0.094	0.930
		$\widehat{q}$	0.025	0.092	0.100	0.958
	1.0	$\widehat{\theta}$	-0.003	0.085	0.086	0.931
		$\widehat{q}$	-0.112	0.161	0.171	0.929
	2.0	$\widehat{\theta}$	-0.003	0.081	0.082	0.939
		$\widehat{q}$	-0.125	0.395	0.420	0.924
3.0	0.5	$\widehat{\theta}$	0.095	0.794	0.861	0.915
		$\widehat{q}$	0.030	0.110	0.124	0.958
	1.0	$\widehat{\theta}$	0.060	0.729	0.786	0.912
		$\widehat{q}$	-0.057	0.245	0.296	0.925
	2.0	$\widehat{\theta}$	0.068	0.709	0.747	0.929
		$\widehat{q}$	-0.084	0.724	0.790	0.924
10.0	0.5	$\widehat{\theta}$	0.291	3.484	3.709	0.901
		$\widehat{q}$	0.035	0.113	0.128	0.963
	1.0	$\widehat{\theta}$	0.102	2.248	2.328	0.926
		$\widehat{q}$	-0.051	0.284	0.389	0.903
	2.0	$\widehat{\theta}$	0.389	3.286	3.316	0.937
		$\widehat{q}$	-0.057	0.834	0.950	0.931

Simulation Results

Table: Estimated Bias, SE, RMSE and CP of the ML e Estimators of the Parameters of the SAK Distribution for $n = 100$.

θ	q	Estimator	Bias	SE	RMSE	CP
0.5	0.5	$\widehat{\theta}$	-0.001	0.065	0.066	0.937
		$\widehat{q}$	0.012	0.063	0.065	0.952
	1.0	$\widehat{\theta}$	-0.002	0.060	0.061	0.940
		$\widehat{q}$	-0.087	0.108	0.115	0.939
	2.0	$\widehat{\theta}$	-0.002	0.057	0.058	0.940
		$\widehat{q}$	-0.077	0.233	0.250	0.932
3.0	0.5	$\widehat{\theta}$	0.035	0.537	0.556	0.927
		$\widehat{q}$	0.015	0.075	0.079	0.955
	1.0	$\widehat{\theta}$	0.028	0.499	0.517	0.929
		$\widehat{q}$	-0.021	0.145	0.188	0.938
	2.0	$\widehat{\theta}$	0.018	0.478	0.491	0.934
		$\widehat{q}$	-0.069	0.440	0.485	0.935
10.0	0.5	$\widehat{\theta}$	0.126	2.400	2.470	0.925
		$\widehat{q}$	0.016	0.075	0.080	0.957
	1.0	$\widehat{\theta}$	0.102	2.248	2.328	0.926
		$\widehat{q}$	-0.031	0.152	0.199	0.939
	2.0	$\widehat{\theta}$	0.172	2.209	2.217	0.944
		$\widehat{q}$	-0.037	0.440	0.483	0.935

Simulation Results

Table: Estimated Bias, SE, RMSE and CP of the ML Estimators of the Parameters of the SAK Distribution for $n = 200$.

θ	q	Estimator	Bias	SE	RMSE	CP
0.5	0.5	$\widehat{\theta}$	0.000	0.046	0.046	0.946
		$\widehat{q}$	0.005	0.043	0.044	0.952
	1.0	$\widehat{\theta}$	-0.001	0.043	0.043	0.946
		$\widehat{q}$	-0.059	0.074	0.081	0.948
	2.0	$\widehat{\theta}$	-0.001	0.040	0.041	0.945
		$\widehat{q}$	-0.041	0.151	0.162	0.942
3.0	0.5	$\widehat{\theta}$	0.013	0.373	0.380	0.940
		$\widehat{q}$	0.009	0.052	0.054	0.953
	1.0	$\widehat{\theta}$	0.012	0.347	0.354	0.941
		$\widehat{q}$	-0.012	0.097	0.117	0.948
	2.0	$\widehat{\theta}$	0.006	0.332	0.339	0.941
		$\widehat{q}$	-0.048	0.255	0.282	0.942
10.0	0.5	$\widehat{\theta}$	0.088	1.684	1.706	0.942
		$\widehat{q}$	0.007	0.052	0.053	0.951
	1.0	$\widehat{\theta}$	0.059	1.574	1.600	0.941
		$\widehat{q}$	-0.023	0.098	0.117	0.948
	2.0	$\widehat{\theta}$	0.035	1.533	1.546	0.947
		$\widehat{q}$	-0.027	0.305	0.313	0.942

Simulation Results

Table: Estimated Bias, SE, RMSE and CP of the ML Estimators of the Parameters of the SAK Distribution for $n = 500$.

θ	q	Estimator	Bias	SE	RMSE	CP
0.5	0.5	$\widehat{\theta}$	0.000	0.029	0.029	0.947
		$\widehat{q}$	0.001	0.027	0.027	0.951
	1.0	$\widehat{\theta}$	0.000	0.027	0.027	0.946
		$\widehat{q}$	−0.046	0.046	0.051	0.948
	2.0	$\widehat{\theta}$	0.000	0.025	0.026	0.947
		$\widehat{q}$	−0.023	0.092	0.095	0.948
3.0	0.5	$\widehat{\theta}$	0.005	0.234	0.235	0.947
		$\widehat{q}$	0.003	0.032	0.033	0.952
	1.0	$\widehat{\theta}$	0.003	0.218	0.219	0.948
		$\widehat{q}$	−0.002	0.060	0.066	0.947
	2.0	$\widehat{\theta}$	0.000	0.208	0.210	0.946
		$\widehat{q}$	−0.008	0.140	0.155	0.948
10.0	0.5	$\widehat{\theta}$	0.019	1.056	1.049	0.944
		$\widehat{q}$	0.003	0.032	0.033	0.951
	1.0	$\widehat{\theta}$	0.009	0.987	0.980	0.948
		$\widehat{q}$	−0.012	0.060	0.080	0.948
	2.0	$\widehat{\theta}$	−0.006	0.955	0.955	0.947
		$\widehat{q}$	−0.018	0.149	0.159	0.943

An Application

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/PlasmaRetinol` (accessed on 31 October 2023).

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 8, 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: Summary for betaplasma data.

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

An Application

Table: ML estimates for AK, TPAD, PAD, SMR and SAK models (standard errors are in parenthesis).

Parameter Estimates	AK	TPAD	PAD	SMR	SAK
θ	0.387 (0.120)	0.016 (0.004)	0.012 (0.003)	16998.167 (3399.076)	0.027 (0.002)
α	—	1.830 (0.133)	1.052 (0.038)	—	—
q	—	—	—	2.926 (0.385)	2.331 (0.294)
σ	25.767 (8.697)	—	—	—	—
log-likelihood	−1952.939	−1955.297	−1953.632	−1910.472	−1908.147

Table: AIC and BIC criteria for fitted models.

Criterion	AK	TPAD	PAD	SMR	SAK
AIC	3909.878	3914.594	3911.264	3824.944	3820.294
BIC	3917.376	3922.092	3918.763	3832.443	3827.793

An Application

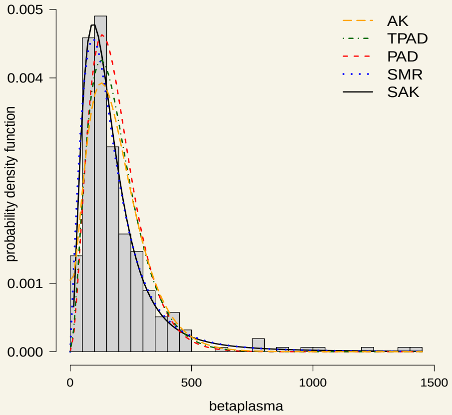

Figure: Betaplasma: histogram and fitted pdf for AK, TPAD, PAD, SMR and SAK distributions.

An Application

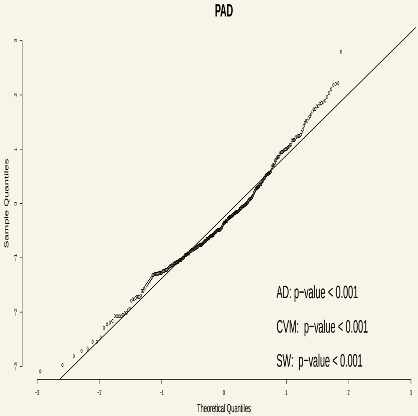

Figure: The qqplots of the quantile residuals for the fitted modelscand *p*-values of the AD, CVM and SW tests.

An Application

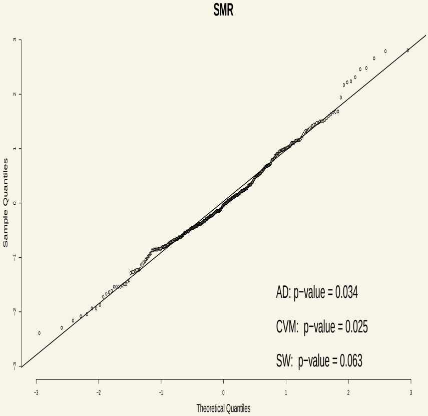

Figure: The qqplots of the quantile residuals for the fitted modelscand p-values of the AD, CVM and SW tests.

An Application

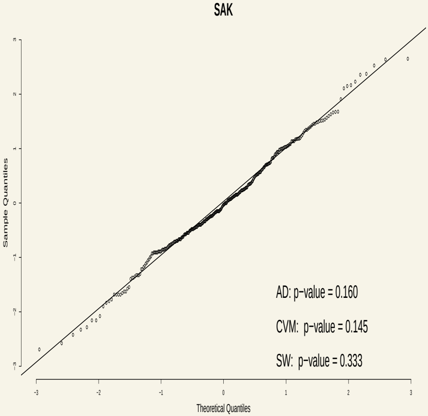

Figure: The qqplots of the quantile residuals for the fitted modelscand p-values of the AD, CVM and SW tests.

Some Final Remarks

- The distribution has two stochastic representations, one based on the quotient of two independent r.v.'s and the other based on a scale mixture between the AK and Beta distributions.
- The pdf, cdf and hazard function of the SAK distribution are represented by the cdf of the gamma model.
- The proposed model has a heavy right tail.
- The model contains the AK distribution as a limit when the parameter q tends to infinity.
- The moments and the skewness and kurtosis coefficient have an explicit form.
- In the application, observing the AIC and BIC and the AD, CVM and SW statistical tests, we may conclude that the SAK distribution fits the Betaplasma data set better than the PAD and SMR distributions, which are also extensions of the AK distribution.

References

- Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control.* **1974**, *19*, 716–723.
- Dempster. A.P.; Laird. N.M.; Rubim. D.B. Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. R. Stat. Soc. Ser. B* **1977**, *39*, 1–38.
- Dunn, P.K.; Smyth, G.K. Randomized Quantile Residuals. *J. Comput. Graph. Stat.* **1996**, *5*, 236–244.
- Gómez, H.W., Venegas, O. Erratum to: A new family of slash-distributions with elliptical contours [Statist. Probab. Lett. 77 (2007) 717–725]. *Stat. Probab. Lett.* **2008**, *78*, 2273–2274.
- Jonhson, N.L.; Kotz, S.; Balakrishnan, N. *Continuous Univariate Distributions*, 2nd ed.; Wiley: New York, NY, USA, 1995; Volume 1.

References

- Kafadar, K. A biweight approach to the one-sample problem. *J. Amer. Statist. Assoc.* **1982**, *77*, 416–424.
- Mosteller, F.; Tukey, J.W. *Data Analysis and Regression*; Addison-Wesley: Reading, MA, USA, 1977.
- Nair, J., Wierman, A. and Zwart, B. (2022). The Fundamentals of Heavy Tails: Properties, Emergence, and Estimation. Cambridge University Press.
- Olmos, N.M.; Varela, H.; Gómez, H.W.; Bolfarine, H. An extension of the half-normal distribution. *Stat. Pap.* **2012**, *53*, 875–886.
- Rivera, P.A.; Barranco-Chamorro, I.; Gallardo, D.I.; Gómez, H.W. Scale Mixture of Rayleigh Distribution. *Mathematics* **2020**, *8*, 1842.
- R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna

References

- Rogers, W.H.; Tukey, J.W. Understanding some long-tailed symmetrical distributions. *Statist. Neerlandica* **1972**, *26*, 211–226.
- Rolski, T.; Schmidli, H.; Schmidt, V.; Teugel, J. *Stochastic Processes for Insurance and Finance*; John Wiley & Sons: Hoboken, NJ, USA, 1999.
- Schwarz, G. Estimating the dimension of a model. *Ann. Statist.* **1978**, *6*, 461–464.
- Shanker, R. Akash Distribution and Its Applications. *Int. J. Probab. Stat.* **2015**, *4*, 65–75.
- Shanker, R.; Shukla, K.K. On Two-Parameter Akash Distribution. *Biom. Biostat. Int. J.* **2017**, *6*, 00178.
- Shanker, R.; Shukla, K.K. Power Akash Distribution and Its Application. *J. Appl. Quant. Methods* **2017**, *12*, 1–10.

References

Wang, J.; Genton, M.G. The multivariate skew-slash distribution. *J. Stat. Plan. Inference* **2006**, *136*, 209–220.