

Moment-Based Density Approximation Techniques as Applied to Heavy-Tailed Distributions

John Sang Jin Kang¹, Serge B. Provost¹ & Jiandong Ren¹

¹ Department of Statistical and Actuarial Sciences, University of Western Ontario, London, Ontario, Canada

Correspondence: John Sang Jin Kang, Department of Statistical and Actuarial Sciences, University of Western Ontario, London, ON., N6A 5B7, Canada. E-mail: skang42@uwo.ca

Received: February 17, 2019 Accepted: March 22, 2019 Online Published: April 3, 2019

doi:10.5539/ijsp.v8n3p1 URL: <https://doi.org/10.5539/ijsp.v8n3p1>

Abstract

Several advances are made in connection with the approximation and estimation of heavy-tailed distributions. It is first explained that on initially applying the Esscher transform to heavy-tailed density functions such as the Pareto, Student- t and Cauchy, said densities can be approximated by employing a certain moment-based methodology. Alternatively, density approximants can be obtained by appropriately truncating such distributions or mapping them onto finite supports. These techniques are then extended to the context of density estimation, their validity being demonstrated by means of simulation studies. As well, illustrative actuarial examples are presented.

Keywords: density approximation, heavy-tailed distributions, exponential tilting, integrated squared error, symmetrization technique, risk measures

1. Introduction

Various density approximation techniques that are associated with the moments or the cumulants of a distribution have been proposed in the statistical literature. For instance, Pearson's curves, which are discussed in Solomon and Stephens (1978), rely only on the first few moments of a distribution. In that case, the density function, $f(x)$, is assumed to satisfy the following differential equation:

$$\frac{1}{f(x)} \frac{df(x)}{dx} = \frac{x - a}{C_0 + C_1 x + C_2 x^2} \quad (1)$$

where a , C_0 , C_1 and C_2 are real constants.

Johnson curves, which are described in Elderton and Johnson (1969), make use of the first four moments of a distribution to approximate a density on the basis of a system of frequency curves whose support can be finite, infinite or semi-infinite. Edgeworth expansions as proposed by Edgeworth (1905) and further developed by Fisher and Cornish (1960) and Hill and Davis (1968), aim at increasing the accuracy of an approximation by making use of Taylor series expansions and appealing to the central limit theorem. The resulting approximants are expressed in terms of Hermite polynomials whose coefficients are determined from the cumulants of the target distribution.

The saddlepoint approximation, which was pioneered by Daniels (1954), has been extensively investigated for several decades. Goutis and Casella (1999) discussed the motivation for applying this technique. Jensen (1995) extensively covered the application of this method to various types of random variables, including i.i.d. sums, compound sums, Markov chains, and sums of independent but not necessarily identically distributed variables. In this instance, the density approximant is expressed as

$$f(x) = \frac{1}{\sqrt{2\pi K''(\hat{s})}} \exp [K(\hat{s}) - \hat{s}x] \quad (2)$$

where $K(t)$ is the cumulant generating function, that is, $K(t) = \ln E(e^{tX})$ and \hat{s} is the solution to $K'(\hat{s}) = x$. The main advantage of the saddlepoint approximation is its accuracy in the tails of the target density. However, this approach may leave something to be desired in the case of multimodal distributions.

In the following section, we outline the so-called polynomially adjusted density approximation technique, which, as explained in Provost (2005) and Ha and Provost (2007), expresses the approximant as the product of an initial base

density and a polynomial adjustment. In Section 3, we propose to adapt such an approximant to heavy-tailed density functions to which the exponential tilting technique (the Esscher transform) is initially applied. The following two sections discuss the approximation of such distributions by truncating their support and applying the transformation variables technique, respectively. Several illustrative examples involving well-known heavy-tailed distributions such as the Type II Pareto (Lomax), Cauchy and Student- t are presented in the last section. Special consideration is also paid to symmetric distributions.

2. Polynomially Adjusted Density Approximation

Provost (2005) proposed a unified methodology for approximating density functions whereby the approximant is expressed as the product of a base density function and a polynomial adjustment. This methodology applies to various types of distributions, including those that are multimodal.

Let $f_X(x)$ be the density function of a continuous random variable X defined on the interval (a, b) and $E(X^\ell) \equiv \mu_X(\ell)$. Letting $Y = (X - u)/s$, $u \in \mathbb{R}$, $s \in \mathbb{R}^+$, the density function of Y , $f_Y(y) = sf_X(u + sy)$, is defined on the interval (a_0, b_0) , where $a_0 = \frac{a-u}{s}$ and $b_0 = \frac{b-u}{s}$. The moments of Y can be evaluated as follows:

$$\mu_Y(\ell) = E\left(\frac{X-u}{s}\right)^\ell = \frac{1}{s^\ell} \sum_{k=0}^{\ell} \binom{\ell}{k} \mu_X(k) (-u)^{\ell-k}. \quad (3)$$

Let $\psi_Y(y)$ denote a base density function, that is, an initial approximation to $f_Y(y)$ and $m_Y(\ell)$ be its associated ℓ^{th} moment, that is, $\int_{a_0}^{b_0} y^\ell \psi_Y(y) dy$. Assuming that $\mu_Y(i)$, $i = 1, 2, \dots$ and $m_Y(j)$, $j = 1, 2, \dots$ exist and are uniquely defined, the approximant to $f_Y(y)$ can be expressed as

$$f_{Y_d}(y) = \psi_Y(y) \sum_{k=0}^d \xi_k y^k \quad (4)$$

where d is a suitable degree for the polynomial adjustment and the ξ_k 's are its coefficients.

The choice of the base density function, $\psi_Y(y)$, depends on the support of the random variable Y and the features of the distribution, as revealed by a histogram of the data. Such approximants are mathematically equivalent to those that are adjusted by certain linear combinations of orthogonal polynomials. For example, when beta, gamma and Gaussian distributions are taken as base densities, the adjustments will respectively involve Jacobi, Laguerre and Hermite polynomials, which can be generated from their associated weight functions, as discussed in Provost and Jiang (2012) and Ha and Provost (2008). The r parameters of $\psi_Y(y)$ can be estimated by making use of the method of moments, that is, by equating $\mu_Y(\ell)$ to $m_Y(\ell)$, $\ell = 1, 2, \dots, r$.

The coefficients, ξ_k , of the polynomial of degree k are determined by solving the system of equations resulting from equating the h^{th} moment of the target distribution to the h^{th} moment associated with the density approximant for $h = 0, 1, \dots, d$, that is,

$$\int_{a_0}^{b_0} y^h \psi_Y(y) \sum_{k=0}^d \xi_k y^k dy = \int_{a_0}^{b_0} y^h f_Y(y) dy, \quad h = 0, 1, \dots, d. \quad (5)$$

This leads to a system of linear equations that yields the following solution in matrix notation:

$$\begin{pmatrix} \xi_0 \\ \xi_1 \\ \vdots \\ \xi_d \end{pmatrix} = \begin{pmatrix} m_Y(0) & m_Y(1) & \cdots & m_Y(d-1) & m_Y(d) \\ m_Y(1) & m_Y(2) & \cdots & m_Y(d) & m_Y(d+1) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ m_Y(d) & m_Y(d+1) & \cdots & m_Y(2d-1) & m_Y(2d) \end{pmatrix}^{-1} \begin{pmatrix} \mu_Y(0) \\ \mu_Y(1) \\ \vdots \\ \mu_Y(d) \end{pmatrix}. \quad (6)$$

The degree of the polynomial adjustment is chosen to be that at which the integrated squared difference (ISD) as defined in Equation (7) or the integrated squared error (ISE) as specified by Equation (8) do not decrease noticeably or satisfy a predetermined tolerance level. The ISD between approximants of successive degrees is defined as

$$ISD_d^\Delta = \int_{a_0}^{b_0} (f_{Y_d}(x) - f_{Y_{d+1}}(x))^2 dx, \quad d = r+1, r+2, \dots \quad (7)$$

If the target density is known, the *ISE* between the exact and the approximated densities is evaluated as follows:

$$ISE_d = \int_{a_0}^{b_0} (f_{Y_d}(x) - f_Y(x))^2 dx, \quad d = r + 1, r + 2, \dots \quad (8)$$

The minimum degree of the polynomial adjustment is set to be $r + 1$ since the base density already matches the first r moments.

Finally, the approximate density of X obtained by applying the inverse transformation is given by

$$f_{X_d}(x) = \psi_Y\left(\frac{x-u}{s}\right) \sum_{k=0}^d \frac{\xi_k}{s} \left(\frac{x-u}{s}\right)^k. \quad (9)$$

It should be noted that the resulting density approximants may occasionally happen to be slightly negative on certain subranges of their support. In order to obtain a truly *bona fide* density function, one can set the approximant to be zero wherever it becomes negative and normalize the resulting function so that it integrates to one. Alternatively, an algorithm proposed by Gajek (1986) can be applied.

3. Density Approximation via Exponential Tilting

3.1 Exponentially Tilted Distributions

An exponentially tilted distribution is obtained by transforming the target probability density function, $f_X(x)$, to the following probability density function:

$$f_{X_\theta}(x) = \frac{e^{-\theta x} f_X(x)}{\mathcal{L}_{f_X(x)}(\theta)}, \quad \theta \in \mathbb{R} \quad (10)$$

where $\mathcal{L}_{f_X(x)}(\theta) = \int_{-\infty}^{\infty} e^{-\theta x} f_X(x) dx$ is the bilateral Laplace transform of $f_X(x)$. The original definition of the Laplace transform, that is, the unilateral Laplace transform, is utilized when the support of X is the positive half-line.

This technique has previously been utilized in actuarial science and financial mathematics. Esscher (1932) initially used this approach to approximate the distribution of aggregate claims at a point of interest, x_0 , the tilting parameter θ being such that the mean of the transformed distribution equals x_0 . Gerber and Shiu (1994) and Elliott, Chan, & Siu (2005) utilized the exponential tilting method to value derivative securities whose prices were governed by certain stochastic processes with stationary and independent increments. Bühlmann (1980) introduced the Esscher premium principle as a risk measure within the framework of utility theory and risk exchange. Cox, Lin, & Wang (2006) demonstrated how to utilize multivariate exponential tilting for pricing certain securities. This technique is also used in importance sampling, as discussed for instance in Siegmund (1976).

In order to apply the polynomially adjusted density approximation methodology described in Section 2, all the moments associated with the target density and the base density must be finite. Accordingly, this approximation approach cannot be directly applied to heavy-tailed distributions, which often have only a limited number of moments. Examples of such distributions include the type II Pareto and the Cauchy distributions. Moreover, as was shown in Provost (2005), the resulting approximant may prove inaccurate when the tail behavior of a distribution under consideration does not match that of the assumed base density.

To mitigate those issues, we propose to exponentially tilt the target density function before attempting to determine a polynomially adjusted approximant.

3.2 The General Algorithm

The following algorithm is utilized to approximate density functions that are transformed via the exponential tilting technique on the basis of their theoretical moments.

1. Given the density function, $f_X(x)$, we obtain the corresponding exponentially tilted density function, $f_{X_\theta}(x)$, as specified by Equation (10).
2. The ℓ^{th} moment associated with the tilted density is determined as follows:

$$\mu_{X_\theta}(\ell) = \int_{-\infty}^{\infty} x^\ell f_{X_\theta}(x) dx. \quad (11)$$

3. The approximant of the tilted density, $\tilde{f}_{X_{d,\theta}}(x)$, is expressed as the product of the base density, $\psi_{X_\theta}(x)$, and a polynomial adjustment of degree d :

$$\tilde{f}_{X_{d,\theta}}(x) = \psi_{X_\theta}(x) \sum_{k=0}^d \xi_k x^k. \quad (12)$$

4. The methodology described in Section 2 is applied to estimate the parameter(s) of the base density and to obtain the coefficients of the polynomial adjustment, that is, by equating $\int_{-\infty}^{\infty} x^h f_{X_\theta}(x) dx$ to $\int_{-\infty}^{\infty} x^h \tilde{f}_{X_{d,\theta}}(x) dx$ for $h = 0, 1, \dots, d$.
5. The approximant of the target density is then recovered by applying the inverse tilting transformation, that is,

$$f_{X_{d,\theta}}(x) = \exp(\theta x) \cdot \mathcal{L}_{f_X(x)}(\theta) \cdot \tilde{f}_{X_{d,\theta}}(x). \quad (13)$$

6. Using the integrated squared error (ISE) formula,

$$ISE_{d,\theta} = \int_{-\infty}^{\infty} (f_{X_{d,\theta}}(x) - f_X(x))^2 dx, \quad \theta \in \mathbb{R}, \quad d = 3, 4, \dots, \quad (14)$$

we select the ‘optimal’ polynomial degree d as well as the appropriate tilting parameter, θ , according to the criteria specified in Section 2.

4. Truncating the Support of a Distribution for Approximation Purposes

Let X be a random variable having a finite number of moments, whose support is the interval $(-\infty, \infty)$. Let $g_{X_{(a,b)}}(x)$ denote the corresponding truncated distribution on the interval (a, b) , whose density function is given by

$$g_{X_{(a,b)}}(x) = \frac{f_X(x)}{\int_a^b f_X(x) dx}, \quad a < x < b \quad (15)$$

where $f_X(x)$ is the density function of X . For the case of a target density function defined on the interval $[0, \infty)$, one can set the right-truncation point b as a certain percentile of the target density, leaving $a = 0$ since only the right-tail behaviour is of interest. If both tails are present and the distribution is symmetric around zero, one can choose $a = -b$ as the left truncation point so that a and b represent the p^{th} and $(100 - p)^{\text{th}}$ percentiles, respectively.

As the support of $X_{(a,b)}$ is now bounded, all the moments of the truncated random variable exists. Thus, the approximation methodology described in Section 2 can be applied. Accordingly, the target density, $f_X(x)$, can be approximated as follows:

1. The ℓ^{th} moment associated with $g_{X_{(a,b)}}(x)$ as specified in Equation (15) is given by

$$\mu_{X_{(a,b)}}(\ell) = \frac{\int_a^b x^\ell f_X(x) dx}{\int_a^b f_X(x) dx}. \quad (16)$$

2. A uniform density on the interval (a, b) is used as base density and the approximant is then

$$f_{X_{d,(a,b)}}(x) = \frac{1}{b-a} \sum_{k=0}^d \xi_k x^k. \quad (17)$$

3. The coefficients of the polynomial adjustments, ξ_k , are determined by solving the linear system resulting from equating $\int_a^b x^h g_{X_{(a,b)}}(x) dx$ to $\int_a^b x^h f_{X_{d,(a,b)}}(x) dx$ for $h = 0, 1, \dots, d$.
4. The ‘optimal’ polynomial degree d is selected on the basis of the ISE between $f_{X_{d,(a,b)}}(x)$ and $g_{X_{(a,b)}}(x)$, denoted by $ISE_{X_{d,(a,b)}}$.

5. Density Approximation via the Transformation of Variables Technique

Consider a random variable X defined on the interval $[0, \infty)$ and the transformed variable, $Y = \frac{X}{X+\delta}$, whose support is the interval $[0, 1)$. Since all the moments of Y exist, the density approximation methodology discussed in Section 2 can be implemented.

More specifically, this methodology involves the following steps:

1. Given $f_X(x)$, the density function of X , the density function of $Y = \frac{X}{X+\delta}$ is

$$f_Y(y) = \frac{\delta}{(1-y)^2} f_X\left(\frac{y\delta}{1-y}\right) \quad (18)$$

and the required associated moments are evaluated.

2. The approximant of the density of Y , $f_{Y_d}(x)$, is expressed as a polynomial as specified in Equation (17), with $a = 0$ and $b = 1$.
3. The coefficients of this polynomial, denoted by ξ_k , are determined by solving the linear systems resulting from equating $\int_0^1 y^h f_Y(y) dy$ to $\int_0^1 y^h f_{Y_d}(y) dy$ for $h = 0, 1, \dots, d$.
4. The approximant of the target density of X is then obtained by applying the inverse transformation:

$$f_{X_{d,\delta}}(x) = \frac{\delta}{(x+\delta)^2} f_{Y_d}\left(\frac{x}{x+\delta}\right). \quad (19)$$

5. The ‘optimal’ polynomial degree, d , and the parameter, δ , are selected on the basis of the *ISE* between $f_{X_{d,\delta}}(x)$ and $f_X(x)$, denoted by $ISE_{d,\delta}$.

6. Applications

In this section, we apply the three proposed density approximation methodologies to certain heavy-tailed distributions. More specifically, we will approximate the density functions of type II Pareto, Cauchy and Student- t distributions. Additionally, the approximation technique relying on exponential tilting is applied in the context of density estimation on the basis of a sample of observed values. Applications involving heavy-tailed actuarial data sets are presented as well.

6.1 Type II Pareto Distribution

In the actuarial context, Pareto distributed individual loss random variables are utilized to model extreme loss and risky types of insurance coverage. Since this distribution is not exponentially bounded, it results in very large values having positive probabilities.

Let X follow a type II Pareto (also called Lomax) distribution, denoted Pareto II (α, β) , with density function

$$f_X(x) = \frac{\alpha\beta^\alpha}{(x+\beta)^{\alpha+1}}, \text{ for } x \geq 0 \text{ and } \alpha, \beta > 0. \quad (20)$$

The ℓ^{th} moment of X is given by

$$\mu_X(\ell) = \begin{cases} \beta^\ell \Gamma(\ell+1) \Gamma(\alpha-\ell) / \Gamma(\alpha) & \text{if } \ell < \alpha \\ \infty & \text{if } \ell \geq \alpha \end{cases}, \quad (21)$$

the number of finite moments available depending on the value of the shape parameter α .

6.1.1 Approximation via Exponential Tilting

According to Nadarajah and Kotz (2006), the analytical solution of the Laplace transform can be expressed either in terms of an incomplete gamma function or the Whittaker function. The solution involving the incomplete gamma function is

$$\mathcal{L}_{f_X(x)}(\theta) = \alpha(\beta\theta)^\alpha e^{\beta\theta} \Gamma(-\alpha, \beta\theta) \quad (22)$$

where $\Gamma(a, x) = \int_x^\infty t^{a-1} e^{-t} dt$, whereas the solution given in terms of Whittaker’s function is

$$\mathcal{L}_{f_X(x)}(\theta) = \alpha(\beta\theta)^{\frac{\alpha-1}{2}} e^{\frac{\beta\theta}{2}} \mathcal{W}_{-\frac{\alpha+1}{2}, -\frac{\alpha}{2}}(\beta\theta), \quad (23)$$

where $\mathcal{W}_{c,d}(z)$ denotes the Whittaker function, which can be expressed as

$$\mathcal{W}_{c,d}(z) = \frac{e^{-\frac{z}{2}} z^{d+\frac{1}{2}}}{\Gamma(d-c+\frac{1}{2})} \int_0^\infty t^{d-c-\frac{1}{2}} (1+t)^{d+c-\frac{1}{2}} e^{-zt} dt, \quad (24)$$

where $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$.

We make use of Equation (23) to obtain the corresponding tilted distribution as Equation (22) can occasionally produce complex numbers.

Consider $X \sim \text{Pareto II}(\alpha, \beta)$, where the symbol “ \sim ” means “is distributed as”. We obtained a new representation of the ℓ^{th} moment of the tilted Type II Pareto distribution, which is

$$\mu_{X_\theta}(\ell) = \left(\frac{\beta}{\theta}\right)^{\ell/2} \Gamma(\ell+1) \frac{\mathcal{W}_{-\frac{\alpha+\ell+1}{2}, -\frac{\alpha-\ell}{2}}(\beta\theta)}{\mathcal{W}_{-\frac{\alpha+1}{2}, -\frac{\alpha}{2}}(\beta\theta)}. \quad (25)$$

The tilted density is then approximated by making use of Equation (12).

No matter what positive real value is being used for the tilting parameter, the tilted density takes on a finite value at $x = 0$. Figure 1 shows a plot of the tilted density function of the Pareto II (3.5, 3.5) where the tilting parameter is $\theta = 1.5$. This suggests taking the exponential density function as base density, that is, $\psi_X(x) = \frac{1}{\lambda} e^{-\frac{x}{\lambda}}$, $x > 0$, where parameter λ is estimated by $\mu_{X_\theta}(1)$.

Other steps that are described in the algorithm included in Section 3, such as determining the coefficients of the polynomial adjustments and recovering the approximant of the target density, are then implemented. The resulting function is finally normalized in order to obtain a *bona fide* density function.

We select the degree of the polynomial adjustment, d , and the tilting parameter, θ , in such a way that the *ISE* with respect to d and θ , as defined in Equation (14), is minimized.

Specifying the value of the tilting parameter can be challenging as it can theoretically take on any positive real values. As suggested by Figure 2, the approximant appears to be robust with respect to the choice of the tilting parameter.

To illustrate how the tilting parameter is chosen, let θ be discretized to two decimals. Figure 3 illustrates how the $ISE_{d,\theta}$ behaves with respect to given values of θ and d , the degree of the polynomial adjustment. Generally, up to polynomial degree 32, the higher the polynomial degree, the lower the $ISE_{d,\theta}$ becomes. For any polynomial degree, d , the $ISE_{d,\theta}$ attains a local minimum at a tilting value lying in the neighbourhood of $\theta = \frac{\alpha}{\beta}$. However, as $\arg \min_{\theta} ISE_{d,\theta}$ gradually

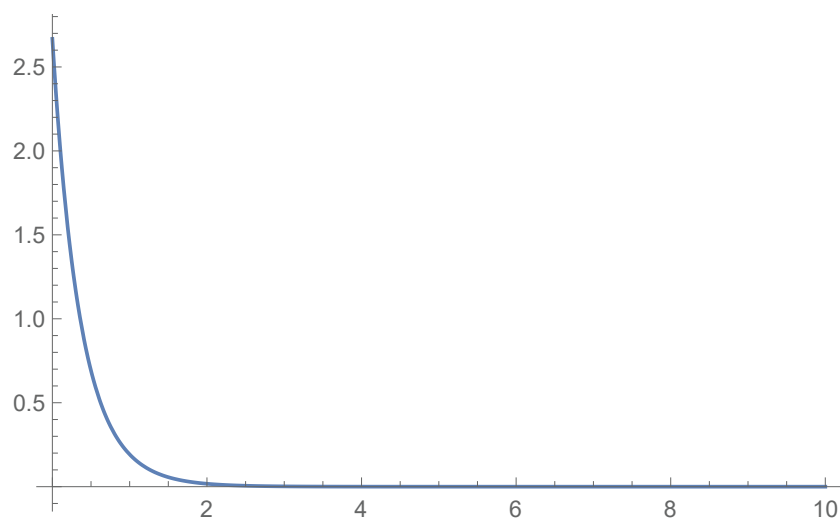


Figure 1. Plot of the tilted Lomax density function ($\alpha = 3.5, \beta = 3.5, \theta = 1.5$)

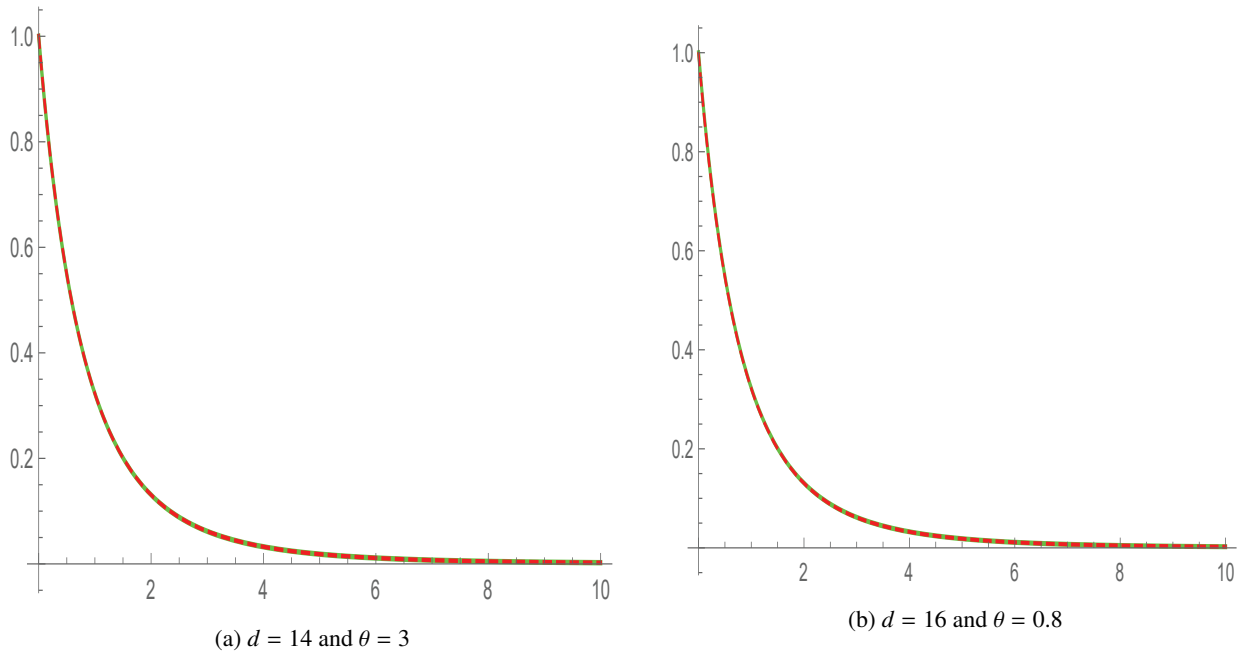


Figure 2. The exact (solid line) and approximated densities (dashed line) of the Lomax distribution for different polynomial degrees and tilting parameters

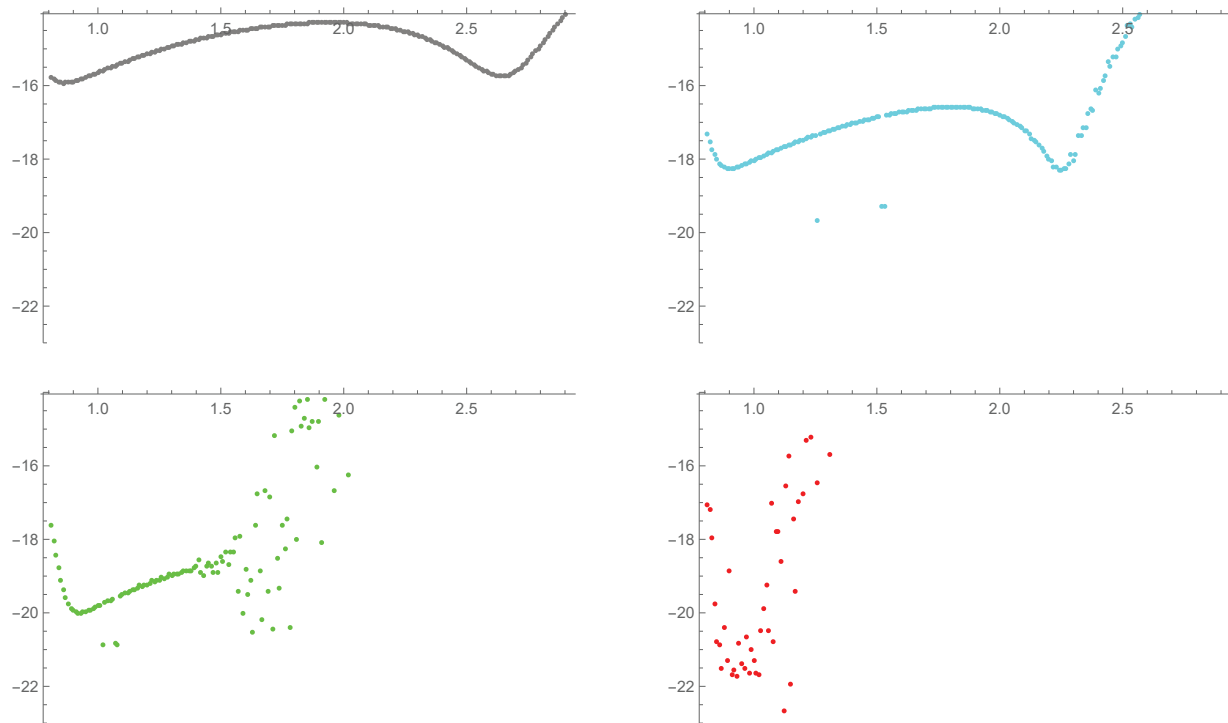


Figure 3. Plots of the log of the integrated squared error (ISE) with respect to θ
 (Top left panel: $d=12$; Top right panel: $d=18$;
 Bottom left panel: $d=24$; Bottom right panel: $d=32$)

Table 1. $ISE_{d,\theta}$'s for the Pareto II ($\alpha = 3.5, \beta = 3.5$) distribution for various values of θ (left column) and d (top numbers)

	26	27	28	29	30	31	32	33	34	35
1	1.49061E-09	1.08512E-09	9.23611E-10	7.52682E-10	5.953E-10	4.57864E-10	5.66844E-10	6.09409E-10	6.84966E-09	2.3436E-08
1.01	1.53571E-09	1.22344E-09	9.49539E-10	7.77049E-10	6.32834E-10	4.34349E-10	3.9275E-10	5.20791E-10	2.26372E-09	1.9667E-09
1.02	1.58918E-09	1.26012E-09	9.82395E-10	8.30968E-10	6.14509E-10	5.5261E-10	3.83891E-10	4.99977E-10	2.75015E-10	1.2383E-08
1.03	1.63385E-09	1.29692E-09	9.72445E-10	9.54928E-10	4.72387E-10	1.45007E-09	1.29628E-09	1.18213E-08	4.39682E-08	1.2267E-06
1.04	1.70458E-09	1.30206E-09	1.08318E-09	8.00901E-10	8.29508E-10	3.30702E-10	2.26703E-09	5.95117E-09	8.85908E-08	2.5198E-07
1.05	1.74003E-09	1.41848E-09	1.00344E-09	1.13852E-09	3.62603E-10	2.93586E-09	4.31604E-09	3.40082E-08	1.15705E-07	2.2295E-06
1.06	1.81514E-09	1.41895E-09	1.1304E-09	9.41529E-10	5.72097E-10	1.40564E-09	1.27666E-09	2.62748E-08	1.52015E-07	2.5022E-06
1.07	1.87428E-09	1.44098E-09	1.26039E-09	6.53731E-10	2.11622E-09	2.07634E-09	4.02014E-08	2.23593E-07	2.03221E-06	7.205E-06
1.08	1.94318E-09	1.4902E-09	1.2294E-09	9.17847E-10	8.37077E-10	5.38434E-10	9.27253E-10	4.3055E-10	1.69338E-08	3.6961E-08
1.09	1.94292E-09	1.66023E-09	1.04757E-09	1.53416E-09	2.37589E-10	6.61543E-09	1.84787E-08	1.79549E-07	1.05305E-06	1.2324E-05
1.1	2.04903E-09	1.59794E-09	1.32109E-09	8.63558E-10	1.45751E-09	3.45979E-10	1.86874E-08	1.16536E-07	1.24969E-06	5.8666E-06
1.11	2.1717E-09	1.52186E-09	1.66558E-09	5.24648E-10	2.86501E-09	9.1443E-10	8.22249E-09	1.39329E-09	3.18481E-07	3.4514E-06
1.12	2.23233E-09	1.60062E-09	1.55925E-09	8.36026E-10	1.17872E-09	7.42794E-10	1.45698E-10	1.002E-08	1.08483E-08	8.0852E-08
1.13	2.29127E-09	1.64248E-09	1.69424E-09	5.95484E-10	3.30866E-09	3.43545E-09	6.47341E-08	4.33466E-07	4.12174E-06	2.4156E-05
1.14	2.25715E-09	1.96971E-09	1.02891E-09	2.85321E-09	1.24387E-09	3.40148E-08	1.49259E-07	1.09429E-06	6.44014E-06	5.7735E-05
1.15	2.39616E-09	1.82269E-09	1.53145E-09	1.19609E-09	7.52665E-10	1.56486E-09	2.90138E-10	1.27062E-08	1.14797E-06	2.2223E-05
1.16	2.41926E-09	1.98698E-09	1.346E-09	1.55629E-09	7.66298E-10	1.79652E-10	2.63977E-08	2.76974E-07	2.92208E-06	1.4001E-05

increases with d , it is difficult to select one particular numerical value of θ as being the 'optimal' tilting parameter. Furthermore, when the polynomial degree gets larger, the $ISE_{d,\theta}$ pattern gradually becomes irregular as it is then subject to numerical evaluation errors.

In light of these considerations, we take $\theta = \frac{\alpha}{\beta} = 1$ as the reference point for the tilting parameter. The resulting $ISE_{d,\theta}$ values appearing in Table 1 indicate that the local minimum of $ISE_{d,\theta}$ occurs at $d = 32$ and $\theta = 1.12$.

After applying the inverse tilting transformation, the approximant obtained with $d = 32$ and $\theta = 1.12$ is in close agreement with the target density, as can be seen from Figure 4.

Figure 5 illustrates how inaccurate an approximant obtained by making use of a gamma density function can be when one resorts to the polynomially adjusted approximation methodology without exponential tilting.

In actuarial applications, risk measures play a major role in pricing, loss reserving and quantitative risk management. Formally, given a random variable X with distribution function $F_X(x)$ and a confidence level $\kappa \in (0, 1)$, the Value-at-Risk (VaR) and Tail Value-at-Risk (TVaR) are respectively defined as

$$\text{VaR}_\kappa(X) = \inf \{ \ell \in \mathbb{R} : F_X(\ell) \geq \kappa \} \quad (26)$$

and

$$\text{TVaR}_\kappa(X) = E(X|X > \text{VaR}_\kappa(X)) = \frac{1}{1-\kappa} \int_\kappa^1 \text{VaR}_t(X) dt. \quad (27)$$

In the actuarial literature, those two risk measures are known as the quantile premium principle and conditional tail expectation, respectively.

Let $X_{d,\theta}$ be a random variable associated with $f_{X_{d,\theta}}(x)$, the density approximant obtained from applying the exponential tilting technique. Letting $X \sim \text{Pareto II}(3.5, 3.5)$ as in the previous example, Tables 2 and 3 provide specific values of VaR and TVaR for the exact and certain approximated distributions. We observe that VaR and TVaR values obtained from the approximated distributions are generally quite close to those determined from the exact distribution.

6.1.2 Approximating Density Functions From Truncated Distributions

We now follow the algorithm presented in Section 4 to obtain an approximant for a type II Pareto density function with parameters $(\alpha = 3.5, \beta = 3.5)$ on the intervals $[0, 95^{th} \text{ percentile}]$ and $[0, 99^{th} \text{ percentile}]$. Figure 6 shows the exact density superimposed on each of the two approximants consisting of polynomials of degrees 25.

The differences between the exact density and its approximants are plotted in Figure 7. We observe that as the support becomes narrower, the error between the exact and approximated density functions and the *ISE* become smaller. In addition, the magnitude of the error increases steeply in the extreme right tail.

6.1.3 Approximation via Transformation of Variables

Consider the type II Pareto density function, $f_X(x) = \frac{\alpha\beta^\alpha}{(x+\beta)^{\alpha+1}}$, and the density function obtained by applying the transformation $Y = \frac{X}{X+\delta}$ with $\delta = \beta$, which turns out to be a beta density function given by $f_Y(y) = \alpha(1-y)^{\alpha-1}$, $0 \leq y \leq 1$. For

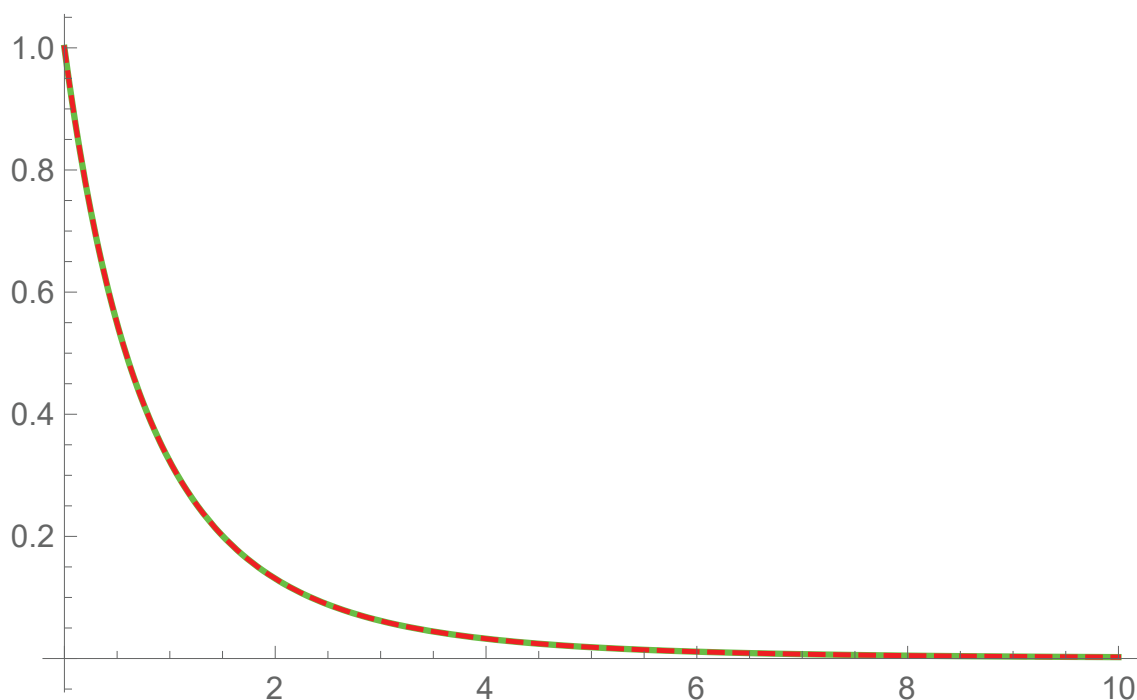


Figure 4. Exact density (solid line) and $f_{X_{d=32, \theta=1.12}}(x)$ (dashed line)
 $ISE = 1.45698 \times 10^{-10}$

Table 2. VaR_κ for X and $X_{d, \theta}$

κ	$VaR_\kappa(X)$	$VaR_\kappa(X_{d=30, \theta=0.88})$	$VaR_\kappa(X_{d=31, \theta=1})$	$VaR_\kappa(X_{d=32, \theta=1.12})$
0.8	2.04337	2.04318	2.04317	2.04335
0.9	3.25744	3.25691	3.2569	3.25739
0.99	9.54658	9.53524	9.53499	9.54544
0.999	21.689	21.4872	21.4679	21.6657
<i>ISE</i>	N/A	7.94042×10^{-10}	4.57864×10^{-10}	1.45698×10^{-10}

Table 3. $TVaR_\kappa$ for X and $X_{d, \theta}$

κ	$TVaR_\kappa(X)$	$TVaR_\kappa(X_{d=30, \theta=0.88})$	$TVaR_\kappa(X_{d=31, \theta=1})$	$TVaR_\kappa(X_{d=32, \theta=1.12})$
0.8	4.26072	4.24757	4.24679	4.25501
0.9	5.96042	5.93443	5.93289	5.94903
0.99	14.7652	14.5228	14.5089	14.6532
0.999	31.7646	29.6872	29.5781	30.6841
<i>ISE</i>	N/A	7.94042×10^{-10}	4.57864×10^{-10}	1.45698×10^{-10}

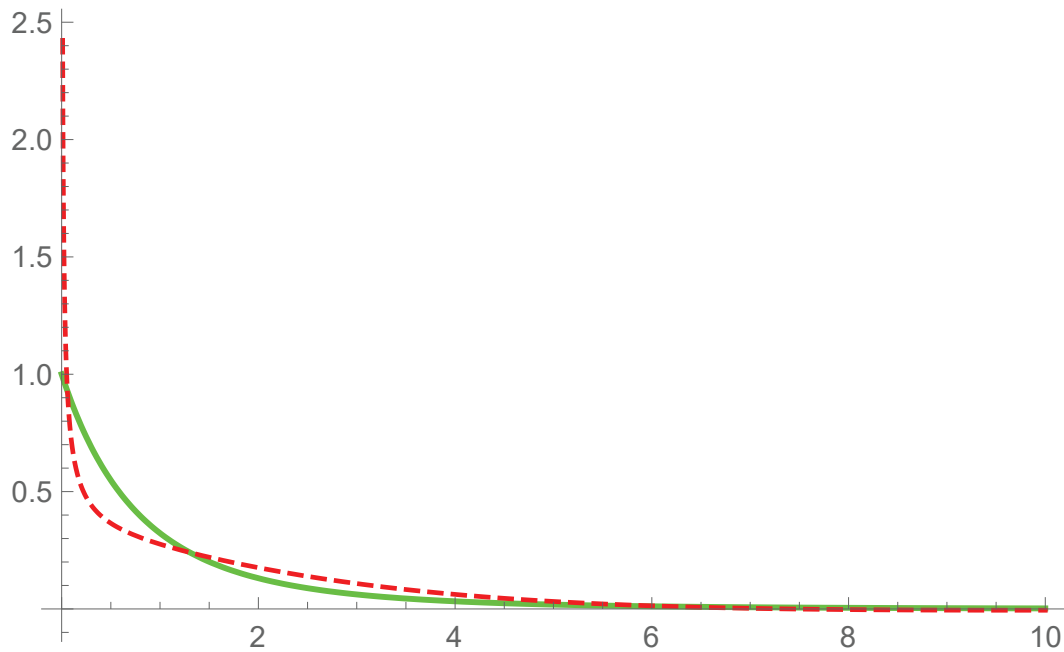


Figure 5. Exact density (solid line) and approximant obtained without applying exponential tilting (dashed line)

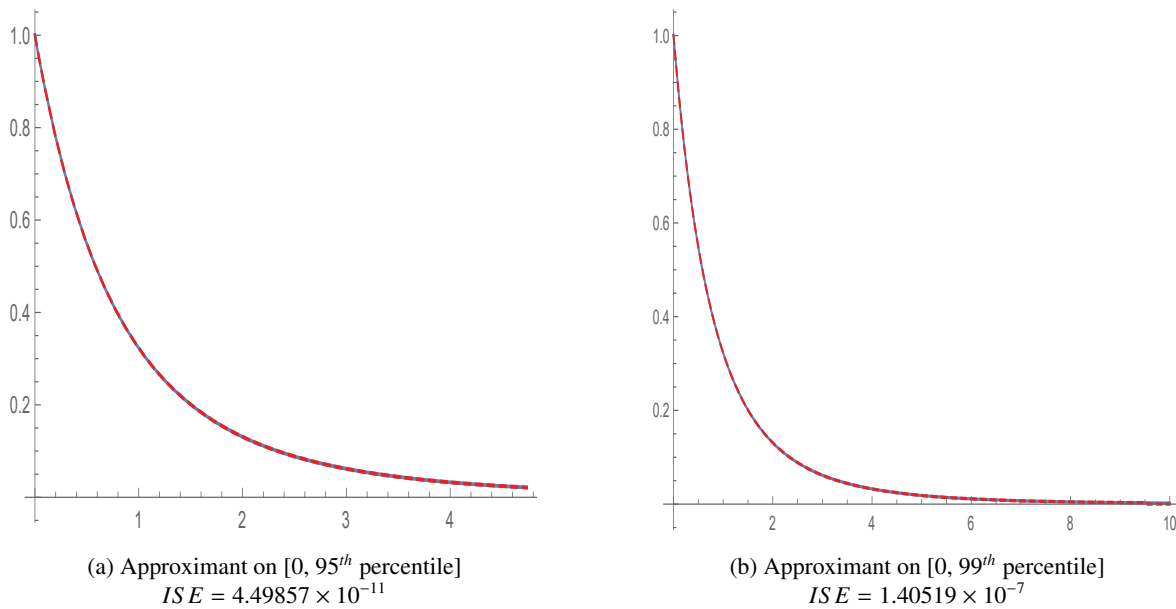


Figure 6. Exact density (solid line) and the approximants obtained by making use of different truncation points (dashed line)

illustrative purposes, we are making use of a Pareto II (3.5, 3.5) distribution. On applying the transformation with $\delta = 3.5$, the resulting 9th degree polynomial approximant is seen to be in close agreement with the target density in Figure 8.

Generally, the ‘optimal’ approximant is obtained when the $ISE_{d,\delta}$ is minimized with respect to the transformation parameter δ and the polynomial degree d . Figure 9 illustrates the patterns followed by the log of $ISE_{d,\delta}$ with respect to changes in δ discretized to two decimal points, while keeping certain polynomial degrees, d , fixed. The discontinuities observed in the $ISE_{d,\delta}$ ’s are due to issues associated with the numerical integration. The $ISE_{d,\delta}$ is minimized at $(d, \delta) = (5.2, 9)$, the corresponding numerical value being 1.31067×10^{-14} . As shown in Figure 8, the resulting approximant is in close agreement with the target density. Additionally, Table 4 provides specific values of the VaR associated the exact density

and the approximant obtained via exponential tilting and transformation of variables.

The TVaR associated with the approximant obtained via the transformation of variables cannot be evaluated as Equation (19) becomes

$$f_{X_{d,\delta}}(x) = \sum_{k=0}^d \xi_k \frac{\delta \cdot x^k}{(x + \delta)^{k+2}}, \quad (28)$$

in which case even the first moment associated with the approximant is unavailable.

6.2 The Student- t Distribution

Let Y be a Student- t random variable with degrees of freedom ν , whose density function is given by

$$f_Y(y; \nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi} \Gamma(\frac{\nu}{2})} \left(1 + \frac{y^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad -\infty < y < \infty, \quad \nu > 0. \quad (29)$$

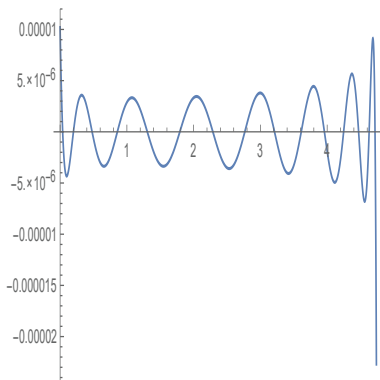
For $\nu > 1$, the ℓ^{th} moment of Y is given by

$$\mu_X(\ell) = \begin{cases} 0 & \ell \text{ odd}, 0 < \ell < \nu \\ \frac{1}{\sqrt{\pi}\Gamma(\frac{\nu}{2})} \left[\Gamma(\frac{\ell+1}{2}) \Gamma(\frac{\nu-\ell}{2}) \nu^{\frac{\ell}{2}} \right] & \ell \text{ even}, 0 < \ell < \nu \end{cases} \quad (30)$$

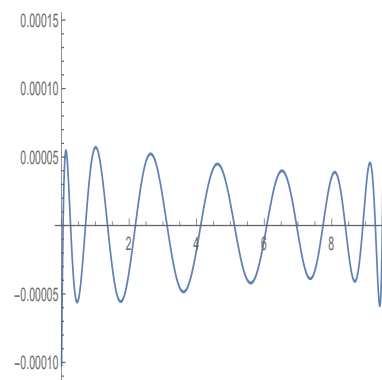
6.2.1 Approximation via Exponential Tilting

As the Laplace transform of the Student- t distribution on ν degrees of freedom does not exist and the number of available moment which depends on ν is finite, we consider the half-Student- t distribution and approximate its distribution by utilizing a symmetrization technique, which was discussed in Ha (2007).

As the density function of Y is symmetric about zero, the transformation $X = |Y|$ yields the half-Student- t density function: $f_X(x) = 2f_Y(x)$, $x > 0$. For instance, when $\nu = 1$ and 3, the analytic solutions of the Laplace transform of the half-Student- t distribution are respectively



(a) Approximant on $[0, 95^{\text{th}}$ percentile]



(b) Approximant on $[0, 99^{\text{th}}$ percentile]

Figure 7. Differences between exact density and the approximants obtained by making use of different truncation points

Table 4. VaR_κ for X and $X_{d,\theta}$

κ	$\text{VaR}_\kappa(X)$	$\text{VaR}_\kappa(X_{d=32,\theta=1.12})$	$\text{VaR}_\kappa(X_{d=9,\delta=5.2})$
0.8	2.04337	2.04335	2.04337
0.9	3.25744	3.25739	3.25744
0.99	9.54658	9.54546	9.54659
0.999	21.689	21.662	21.6888
<i>ISE</i>	N/A	1.45698×10^{-10}	1.31067×10^{-14}

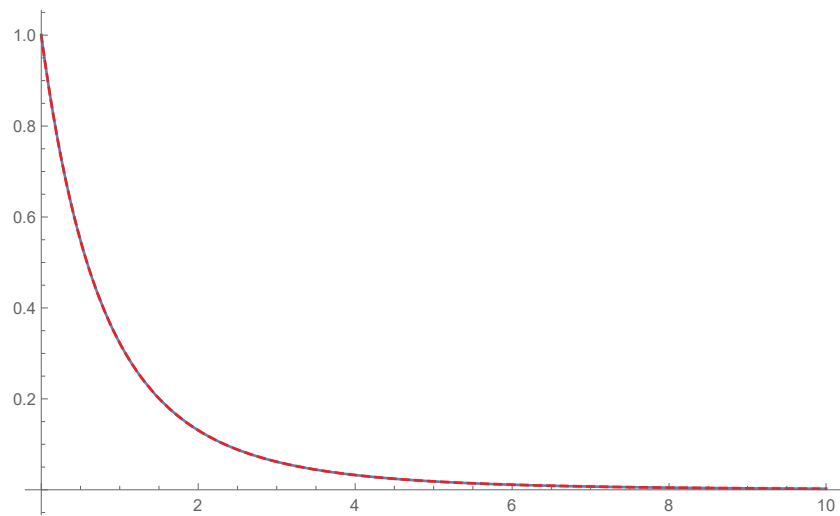


Figure 8. Exact density (solid line) and $f_{X_{d=9, \delta=3.5}}(x)$ (dashed line).
 $ISE = 1.30606 \times 10^{-12}$

$$\mathcal{L}_{f_Y(y; \nu=1)}(\theta) = \frac{2 \text{Ci}(\theta) \sin(\theta) + (\pi - 2 \text{Si}(\theta)) \cos(\theta)}{\pi} \quad (31)$$

and

$$\mathcal{L}_{f_Y(y; \nu=3)}(\theta) = \frac{2 G_{1,3}^{3,1} \left(\frac{3}{4} \theta^2 \middle| \begin{matrix} \frac{1}{2} \\ 0, \frac{1}{2}, \frac{3}{2} \end{matrix} \right)}{\pi^{\frac{3}{2}}} \quad (32)$$

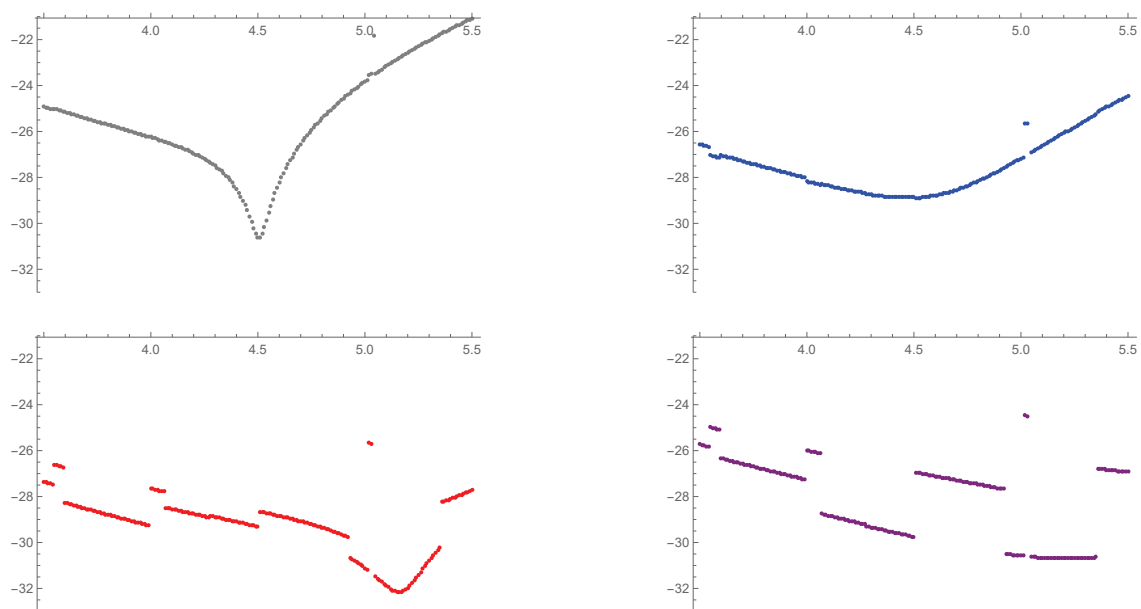


Figure 9. Plots of the log of the integrated squared error (ISE) with respect to δ
 (Top left panel: $d=7$; Top right panel: $d=8$;
 Bottom left panel: $d=9$; Bottom right panel: $d=10$)

where $\text{Ci}(\theta) = -\int_0^\infty \frac{\cos(t)}{t} dt$ and $\text{Si}(\theta) = \int_0^\theta \frac{\sin(t)}{t} dt$ represent the cosine integral, and the sine integral respectively, and

$$G_{p,q}^{m,n} \left(z \left| \begin{matrix} a_1, \dots, a_p \\ b_1, \dots, b_q \end{matrix} \right. \right) = \frac{1}{2\pi i} \int_{\mathcal{L}} \frac{(\prod_{j=1}^m \Gamma(b_j - s)) (\prod_{j=1}^n \Gamma(1 - a_j + s))}{(\prod_{j=n+1}^p \Gamma(a_j - s)) (\prod_{j=m+1}^q \Gamma(1 - b_j + s))} z^s ds \quad (33)$$

denotes Meijer's G -function, where m, n, p , and q are integers with $0 \leq m \leq q$ and $0 \leq n \leq p$, $i = \sqrt{-1}$ and the infinite contour of integration \mathcal{L} , which represents a suitable closed contour in the complex plane, separates the poles of $\Gamma(b_j - s)$, $j = 1, 2, \dots, m$ from the poles of $\Gamma(1 - a_j + s)$, $j = 1, 2, \dots, n$.

Here are new representations of the ℓ^{th} moment of the tilted half-Student- t distribution. When $\nu = 1$ (corresponding to the half-Cauchy distribution), one has

$$\mu_{X_\theta}(\ell; \theta; \nu = 1) = \begin{cases} 1 & \text{if } \ell = 0 \\ \frac{(\pi - 2 \text{Si}(\theta)) \sin(\theta) - 2 \text{Ci}(\theta) \cos(\theta)}{2 \text{Ci}(\theta) \sin(\theta) + (\pi - 2 \text{Si}(\theta)) \cos(\theta)} & \text{if } \ell = 1 \\ \frac{-2 \text{Ci}(\theta) \sin(\theta) + 2 \text{Si}(\theta) \cos(\theta) + \frac{2}{\theta} - \pi \cos(\theta)}{2 \text{Ci}(\theta) \sin(\theta) + (\pi - 2 \text{Si}(\theta)) \cos(\theta)} & \text{if } \ell = 2 \\ \frac{G_{1,3}^{3,1} \left(\frac{\theta^2}{4} \left| \begin{matrix} \frac{1-\ell}{2} \\ \frac{1-\ell}{2}, 0, \frac{1}{2} \end{matrix} \right. \right)}{2 \text{Ci}(\theta) \sin(\theta) + (\pi - 2 \text{Si}(\theta)) \cos(\theta)} \frac{1}{\sqrt{\pi}} & \text{if } \ell = 3, 4, \dots \end{cases} \quad (34)$$

and when $\nu = 3$, one has

$$\mu_{X_\theta}(\ell; \theta; \nu = 3) = \begin{cases} 1 & \text{if } \ell = 0 \\ \frac{\sqrt{3}^\ell G_{1,3}^{3,1} \left(\frac{3}{4} \theta^2 \left| \begin{matrix} \frac{1-\ell}{2} \\ 0, \frac{1}{2}, \frac{3-\ell}{2} \end{matrix} \right. \right)}{G_{1,3}^{3,1} \left(\frac{3}{4} \theta^2 \left| \begin{matrix} \frac{1}{2} \\ 0, \frac{1}{2}, \frac{3}{2} \end{matrix} \right. \right)} & \text{if } \ell = 1, 2 \\ \frac{\sqrt{3}^\ell G_{1,3}^{3,1} \left(\frac{3}{4} \theta^2 \left| \begin{matrix} \frac{1-\ell}{2} \\ \frac{3-\ell}{2}, 0, \frac{1}{2} \end{matrix} \right. \right)}{G_{1,3}^{3,1} \left(\frac{3}{4} \theta^2 \left| \begin{matrix} \frac{1}{2} \\ 0, \frac{1}{2}, \frac{3}{2} \end{matrix} \right. \right)} & \text{if } \ell = 3, 4, \dots \end{cases} \quad (35)$$

If the degree of freedom is odd and greater than or equal to three, the ℓ^{th} tilted moment is represented by a ratio of Meijer's G -function. However, when the degrees of freedom are even, an explicit solution does not exist so that each tilted moment needs to be evaluated via numerical integration.

As the plot of the tilted half-Student- t density function is decreasing and takes a finite value at $x = 0$, the exponential density function is used as base density, its parameter being estimated from $\mu_{X_\theta}(1)$.

On observing the behavior of $IS E_{d,\theta}$ with respect to the tilting parameter, θ , we determined that setting $\theta = \frac{\nu+1}{\nu+2}$ as the reference point for the tilting parameter generally yields accurate approximants. Exceptionally, when we attempt to approximate the half-Cauchy distribution, for which $\nu = 1$, setting the tilting parameter θ equal to one significantly lowers the $IS E_{d,\theta}$. The alternating behavior of the log of $IS E_{d,\theta}$ with $\theta = 1$, $\nu = 1$ is displayed in Figure 10. Once the approximant of half-Cauchy density, $f_{X_{d,\theta}}(x)$ as specified in Equation (13), is obtained after applying the inverse tilting transformation, the approximant of the Cauchy density, which is symmetric, is obtained as follows:

$$f_{Y_{d,\theta}}(x) = \frac{1}{2} f_{X_{d,\theta}}(-x) \mathbb{I}_{(-\infty, 0)}(x) + \frac{1}{2} f_{X_{d,\theta}}(x) \mathbb{I}_{(0, \infty)}(x) \quad (36)$$

$$\text{where } \mathbb{I}_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}.$$

Figure 11 shows the target density superimposed on the approximant consisting of a 37^{th} degree polynomial.

6.2.2 Approximating Density Functions From Truncated Distributions

We now attempt to approximate the truncated density function of a Student- t random variable by applying the methodology discussed in Section 4. Fixing the polynomial degree to 19, plots of the resulting approximants of the Student- t density

functions with degrees of freedom 1, 2, 3, and 4 within two-sided 99% probability intervals are presented in Figure 12. For lower numbers of degrees of freedom, the approximants show oscillating patterns, which is not the case (except when $\nu = 1$) when the symmetrization technique is applied as can be seen from Figure 13.

6.2.3 Approximation via Transformation of Variables

Once the symmetrization technique is applied, the half-Student- t density functions with degrees of freedom, ν , are approximated by making use of the methodology described in Section 5. The approximants of the Student- t density functions are recovered by making use of Equation (36).

The approximants of the Student- t density functions having degrees of freedom 1, 2, 3, and 4 are plotted in Figure 14 whose caption specifies the corresponding $ISE_{d,\delta}$'s.

6.3 Estimating Densities From Data by Means of Exponential Tilting

This section explains how the density approximation methodology that is based on exponential tilting, which is discussed in Section 3, is applied when a sample of observations is available. As the counterpart of density approximants, which make use of the exact tilted moments, the density estimates rely on the sample tilted moments.

Consider an observation vector $\mathbf{x} = (x_1, x_2, \dots, x_n)$. When exponential tilting is utilized in the context of density estimation, the following points ought to be considered:

1. Before the density is estimated, the sample should be rescaled so that the sample standard deviation be equal to one. Heavy-tailed features are detected from the extreme quantiles of the sample or from the positive-sloped mean excess plot, as discussed by Ghosh and Resnick (2010). To justify the application of the exponential tilting technique, the resulting estimate needs to be compared to that obtained without tilting.
2. The ℓ^{th} sample tilted moment with tilting parameter θ , denoted by $m_\theta(\ell)$, is given by

$$m_\theta(\ell) = \frac{\sum_{j=1}^n \exp(-\theta x_j) x_j^\ell}{\sum_{j=1}^n \exp(-\theta x_j)}. \quad (37)$$

3. A histogram of the data can be utilized to identify an appropriate distribution to be used as base density function. For non-negative samples, gamma or exponential density functions are usually suitable as base densities.

For the gamma base density function, that is,

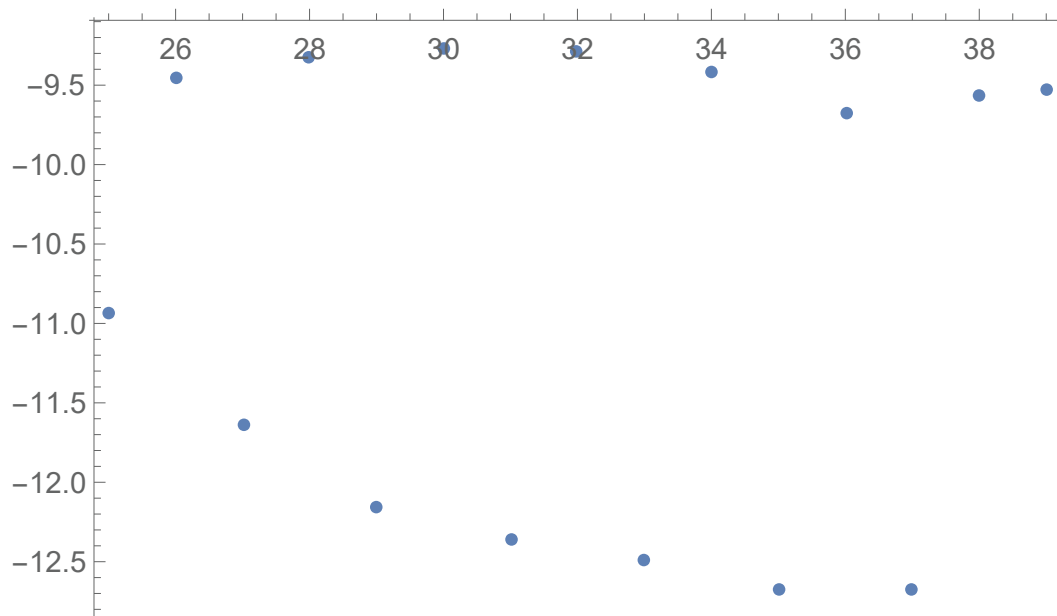


Figure 10. Plot of the log of $ISE_{d,\theta}$ with $\theta = 1$

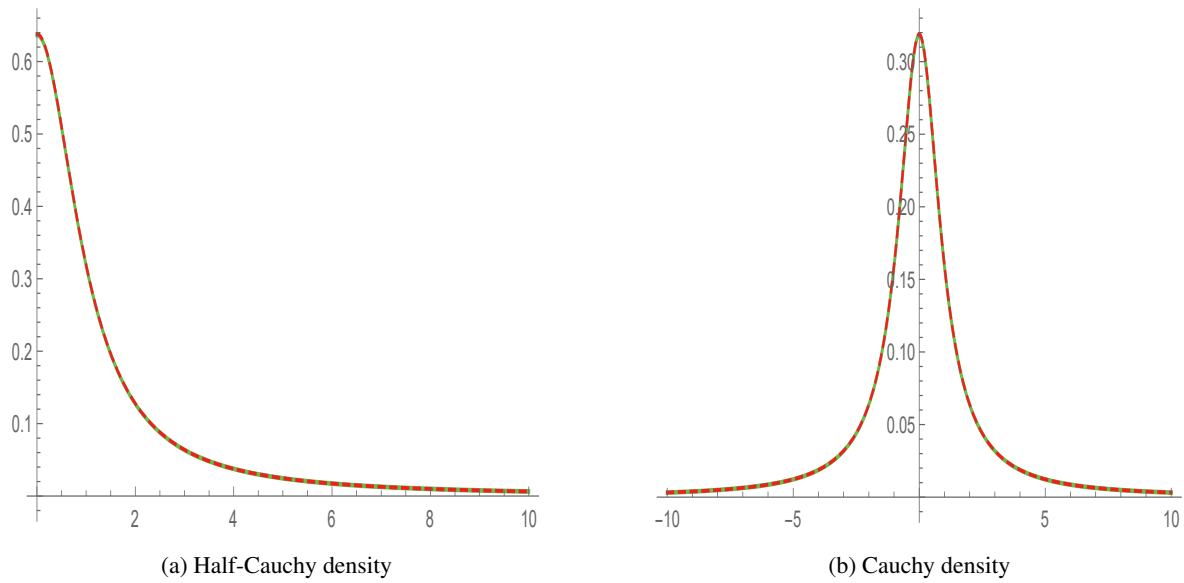


Figure 11. Exact density (green solid line) and $f_{X_{d=37, \theta=1.12}}(x)$ (red dashed line)
 $ISE = 3.10482 \times 10^{-6}$

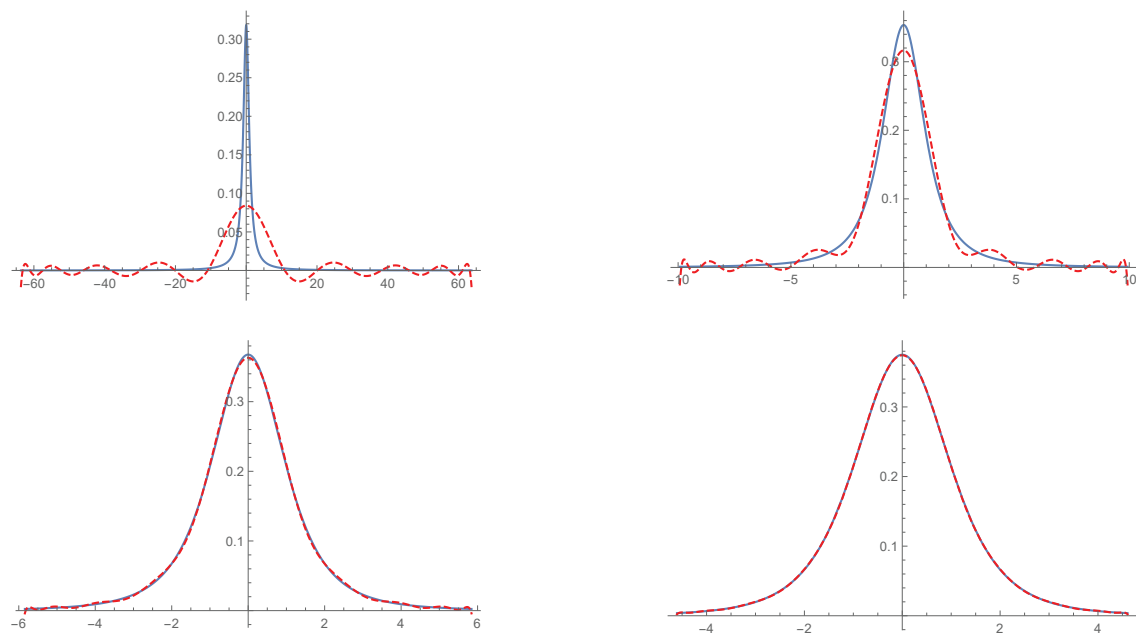


Figure 12. Plots of the exact (solid line) and approximate (dashed line)
 Student- t densities for various degrees of freedom, ν
 (Top left panel: $\nu=1$; Top right panel: $\nu=2$;
 Bottom left panel: $\nu=3$; Bottom right panel: $\nu=4$)

$$\psi_X(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}, \quad x \geq 0, \quad (38)$$

the parameters α and β are estimated by

$$\alpha = \frac{m_{\theta}(1)^2}{m_{\theta}(2) - m_{\theta}(1)^2} \text{ and } \beta = \frac{m_{\theta}(2)}{m_{\theta}(1)} - m_{\theta}(1). \quad (39)$$

If the histogram is more or less symmetric, the symmetrization technique discussed in Section 6.2.1 is applied to the half-distribution; then a gamma or exponential density function can be utilized as base density function, with a shift of location if necessary.

4. The choice of the tilting parameter, θ , as well as the degree of the polynomial adjustment, d , is based on the Anderson-Darling test. Let $AD_{d,\theta}$ denote the Anderson-Darling test statistic evaluated at the resulting cumulative distribution function (CDF) whose associated tilting parameter and polynomial adjustment degree are respectively θ and d , that is,

$$AD_{d,\theta} = -n - \sum_{k=1}^n \frac{2k-1}{n} [\log(1 - F_{d,\theta}(x_{(n-k+1)})) + \log(F_{d,\theta}(x_{(k)}))] \quad (40)$$

where n is the sample size, $F_{d,\theta}(x)$ denotes the CDF determined from the density estimate, $f_{d,\theta}(x)$, and $x_{(1)} < x_{(2)} < \dots < x_{(n)}$ is the ordered sample.

Such a goodness-of-fit measure can be utilized even when $f_{d,\theta}(x)$ is slightly negative at some sample points. However, the log-likelihood statistics or related discrepancy measures, such as the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC), are inappropriate when the adjusted density estimates take on negative values.

5. From the several candidates of $f_{d,\theta}(x)$ having low $AD_{d,\theta}$ values, it is desirable to select one that has no root. If, however, we cannot find such a density estimate, then we choose the one having an even number of roots in the right tail. Such a density estimate can be made *bona fide* by setting it to be zero wherever it becomes negative and normalizing the resulting function.

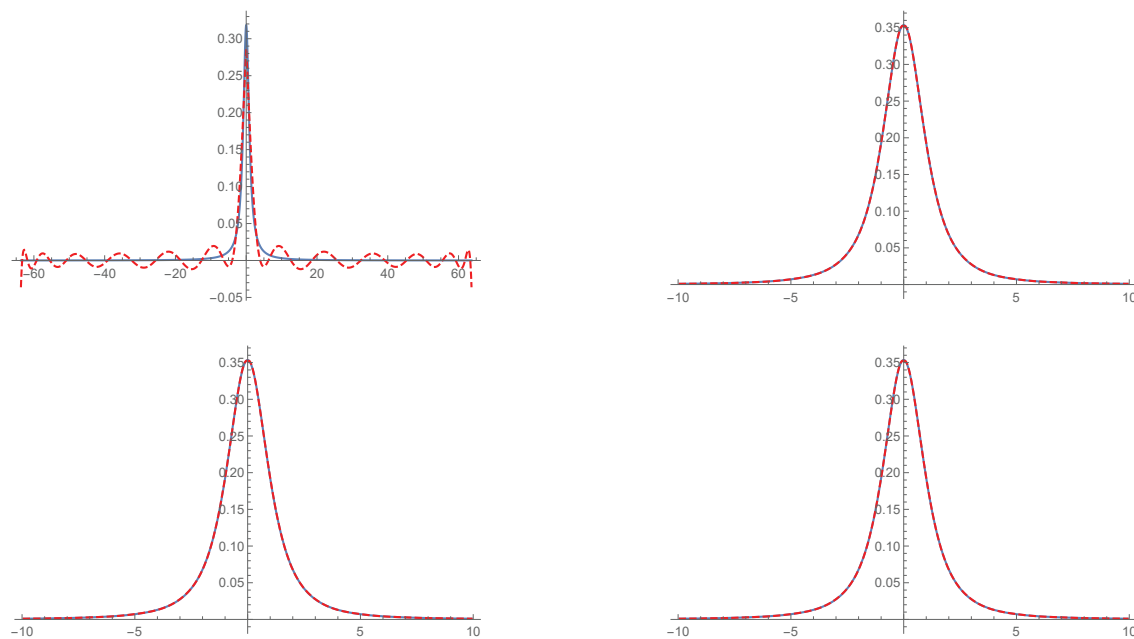


Figure 13. Plots of exact (solid line) and approximate (dashed line) truncated Student- t densities obtained by applying the symmetrization technique
(Top left panel: $\nu=1$; Top right panel: $\nu=2$;
Bottom left panel: $\nu=3$; Bottom right panel: $\nu=4$)

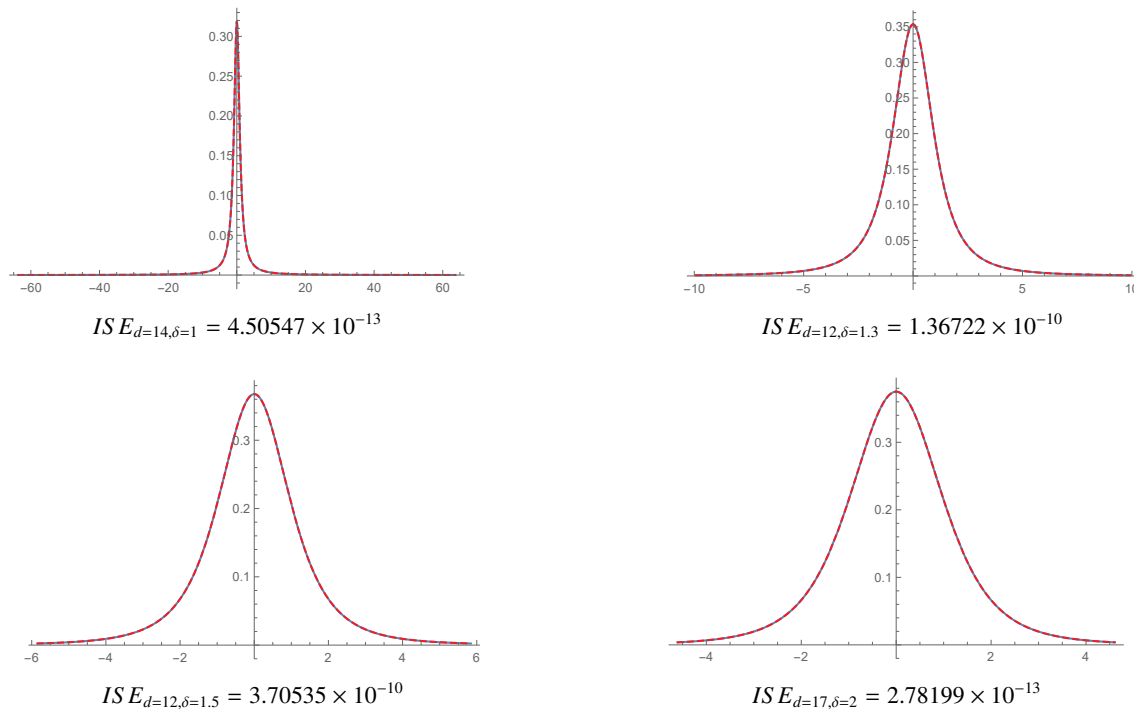


Figure 14. Plots of exact (solid line) and approximate (dashed line) Student- t densities obtained by applying the transformation of variables techniques
(Top left panel: $\nu=1$; Top right panel: $\nu=2$;
Bottom left panel: $\nu=3$; Bottom right panel: $\nu=4$)

6.3.1 Simulated Data From a Type II Pareto Distribution

Consider a simulated sample of size 5,000 from a Pareto II (3.5, 35) distribution. Once the values are rescaled by dividing them by their standard deviation, the methodology described in Section 3 is applied, except that the exact tilted moments are replaced with the sample tilted moments. In this case, an exponential density function is used as base density function.

Plots of $AD_{d,\theta}$ values for polynomial adjustments of various degrees are shown in Figure 15. When taking parsimony into account, the graphs indicate that one could select $d = 9$ and $\theta = 3$, as suitable polynomial degree and tilting parameter, respectively.

The density estimate is obtained after applying the inverse tilting and scale transformations. The resulting density estimate is plotted in Figure 16.

Before comparing the risk measures associated with the density estimates, we introduce the following notations:

1. Tilted Exponential Polynomial (TEP) as the density estimate resulting from the exponential tilting technique, where the exponential density function is used as the base density (with d and θ representing the degree of the polynomial adjustment and tilting parameter, respectively).
2. Tilted Gamma polynomial (TGP) as the density estimate resulting from applying the exponential tilting technique, where the gamma density function is used as the base density (with d and θ representing the degree of the polynomial adjustment and tilting parameter, respectively).
3. Box-Cox Polynomial (BCP) as the density estimate obtained by applying the Box-Cox transform in conjunction with a polynomial adjustment. This transformation, introduced by Box and Cox (1964), is defined as

$$y(\lambda) = \begin{cases} \frac{y^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0 \\ \log y & \text{if } \lambda = 0 \end{cases} \quad (41)$$

Once the sample is transformed to a more or less symmetric shape by applying Box-Cox transform, the normal density function which is used as base density, is estimated and a polynomial adjustment is applied, as discussed in

Section 2 (with d and λ representing the degree of the polynomial adjustment and transform parameter, respectively).

Once those estimates are back-transformed, one can compare the risk measures, which is done in Tables 5 and 6 with respect to VaR and TVaR, respectively.

6.3.2 Simulated Data From a Student- t Distribution

Consider a simulated sample of size 10,000 from a Student- t distribution on 3 degrees of freedom. After rescaling, symmetrizing and applying the exponential tilting technique along with making use of an exponential density function as base density for estimating the corresponding ‘half Student- t ’ density function, one can select $d = 7$ and $\theta = 0.9$ as the degree of the polynomial adjustment and tilting parameter. The plot of the density estimate is superimposed on a histogram of the sample in Figure 17.

6.3.3 Automobile Insurance Claims Data Set

This data set, which was introduced by Frees (2011), consists of 6 773 individual claim amounts for private passenger automobile insurance issued by a large US mid-western property and casualty insurance company. This data set was originally used to examine claim distributions for several risk classification categories.

The summary statistics for the rescaled data whose standard deviation is equal to one are presented in Table 7. The heavy-tailed nature of the sample is clearly indicated by the high kurtosis measure as well as an extreme maximum observed

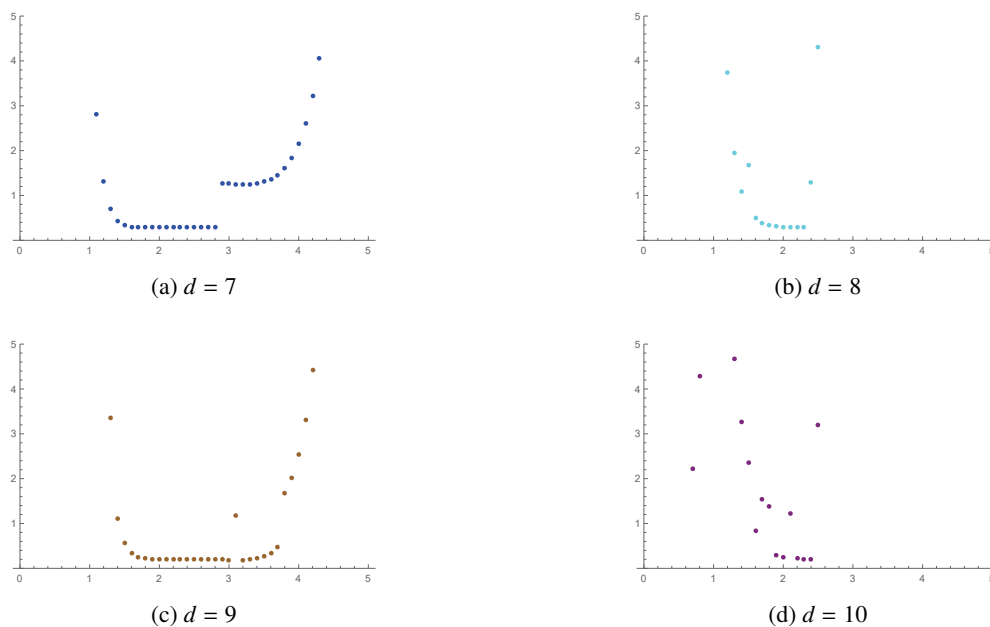


Figure 15. Plots of $AD_{d,\theta}$ vs θ leading to the selection of the optimal d and θ

Table 5. VaR_κ evaluated from the sample, the exact distribution and two density estimates

κ	Empirical	Pareto (3.5,35)	TEP ($d = 9$ and $\theta = 3$)	BCP ($d = 4$ and $\lambda = 0.1$)
0.8	20.3432	20.4337	20.4421	20.2569
0.9	33.0825	32.5774	32.7712	33.0464
0.95	48.7676	47.3741	48.2939	49.1294
0.99	100.004	95.4658	97.6907	104.338

Table 6. $TVaR_\kappa$ evaluated from the sample, the exact distribution and two density estimates

κ	Empirical	Pareto (3.5,35)	TEP ($d = 9$ and $\theta = 3$)	BCP ($d = 4$ and $\lambda = 0.1$)
0.8	44.9521	42.6072	42.6498	44.6046
0.9	64.2477	59.6042	59.7103	63.5354
0.95	89.0128	80.3238	79.9156	87.2058
0.99	182.639	147.652	137.345	166.829

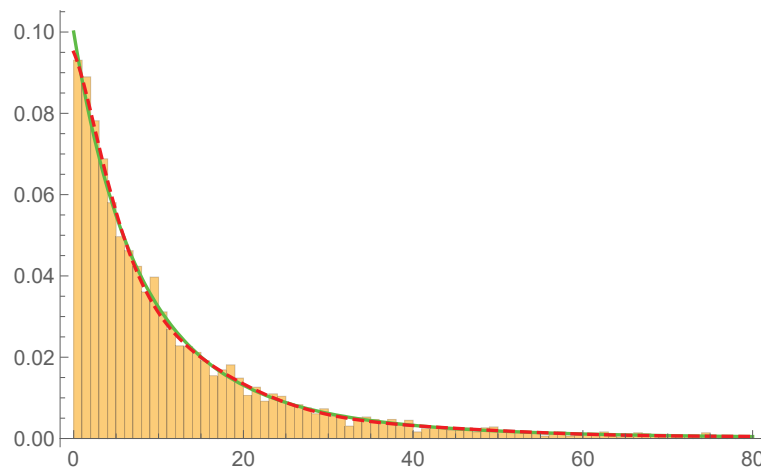


Figure 16. Histogram of the simulated data from a Pareto II (3.5, 35) distribution and plots of the exact density (solid line) and the density estimate (dashed line)

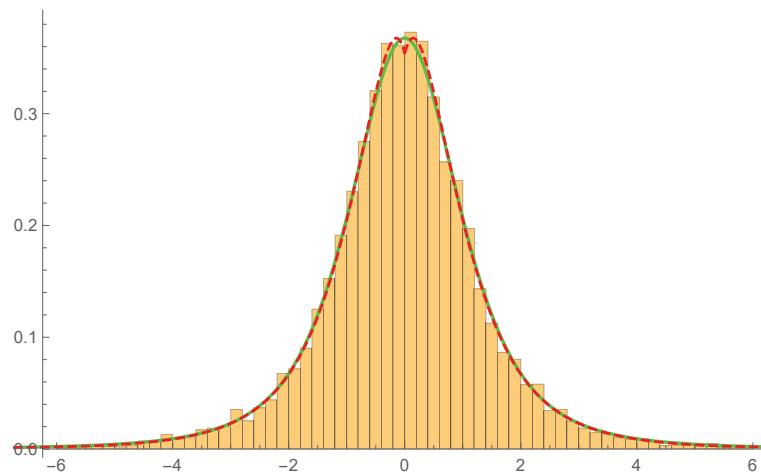


Figure 17. Histogram of the simulated data from a Student- t distribution on 3 degrees of freedom, the exact density (solid line) and the density estimate (dashed line)

value.

Interestingly, after the proposed density estimation methodology is applied, including the exponential tilting, and several $AD_{d,\theta}$ values are compared, both a TEP ($d = 11, \theta = 1.1$) estimate and a TGP ($d = 9, \theta = 1.5$) estimate provide suitable models as can be seen from Figure 18. We note that the exponential tilting is also useful in the context of density estimation since otherwise the tail of the resulting density estimate presents undesirable fluctuations with subintervals where the estimates are negative. It is seen that the kernel estimate also included in this figure is outperformed by the proposed estimates.

Additionally, neither density estimate has a root in the right tail. VaR and TVaR are compared in Tables 8 and 9, respec-

Table 7. Summary of the rescaled automobile insurance claims data set (Original standard deviation: 2 646.91)

Sample size	6 773	Min	0.70008
Mean	0.70008	1st quartile	0.19787
Skewness	6.23567	2nd quartile	0.37844
Kurtosis	87.2775	3rd quartile	0.80751
		Max	22.6679

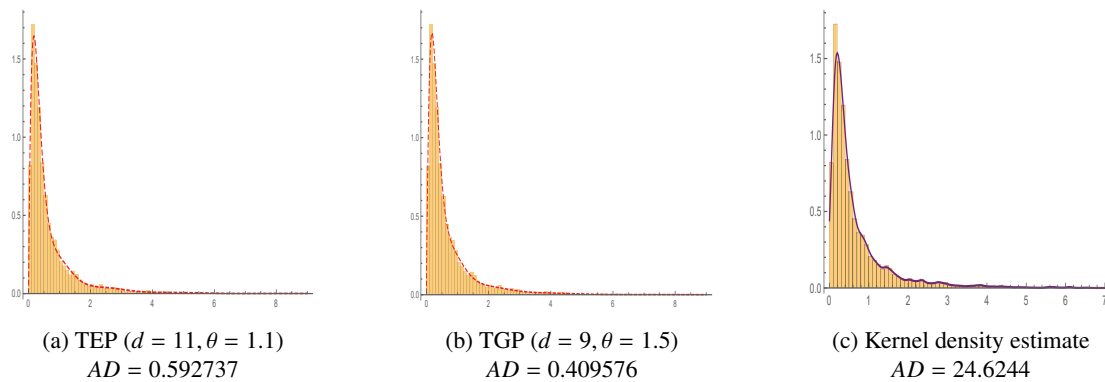


Figure 18. Density estimates and rescaled histogram for the auto insurance claims data set

Table 8. VaR_κ based on the auto insurance claims data

κ	Empirical	TEP ($d = 11$ and $\theta = 1.1$)	TGP ($d = 9$ and $\theta = 1.5$)	BCP ($d = 7$ and $\lambda = 0$)
0.8	2 545	2 591.97	2 576.89	2 534.5
0.9	4 171.01	4 159.98	4 185.58	4 224.53
0.95	6 356.9	6 468.07	6 356.05	6 358.44
0.99	12 091.5	12 413.1	11 971.5	12 668.8

Table 9. TVaR_κ based on the auto insurance claims data

κ	Empirical	TEP ($d = 11$ and $\theta = 1.1$)	TGP ($d = 9$ and $\theta = 1.5$)	BCP ($d = 7$ and $\lambda = 0$)
0.8	5 518.88	5 452.73	5 585.01	5 521.61
0.9	7 772.38	7 641.58	8 028.22	7 800.34
0.95	10 415.8	10 156.1	10 609.9	10 465
0.99	18 263.7	17 027.9	19 404.8	18 070.7

tively.

6.3.4 The Danish Fire Data Set

The Danish fire data set is often used in applications involving extreme value theory; see for instance McNeil (1997) and Resnick (1997). Collected at Copenhagen Reinsurance from 1980 to 1990, the data consists of inflation-adjusted total fire losses, expressed in millions of Danish Krone. Although the sample size of the full data set is 2 492, we confine our attention to the 2 156 losses exceeding one million Krone, taking into account the insurance deductible.

As various density estimate candidates are suggested, we describe each one, along with the determination of the parameters, such as the degree of polynomial adjustment (d), the tilting parameter (θ) and the Box-Cox transform parameter (λ).

1. Density I: TEP obtained from the rescaled data

Table 10 provides summary statistics for the rescaled data, which are clearly indicative of a heavy-tailed distribution. After comparing the density estimates obtained with and without exponential tilting, along with the choice of a gamma or exponential density functions as base densities, a TEP ($d = 9, \theta = 1.15$) estimate which has no root in the right tail is a possible candidate. However, the corresponding Anderson-Darling test statistic, which is equal to 44.29 is large due to the fact that the deductible was not taken into consideration. As Figure 19 (a) reveals, the back-transformed density estimate does not provide a good fit around the mode of the distribution.

2. Density II: Location-shifted TEP obtained from the rescaled data

This shortcoming is circumvented by making use of a shifted density estimate. Given a deductible level, d_0 , and denoting by σ the standard deviation of the sample $\mathbf{x} = (x_1, x_2, \dots, x_n)$, where $x_i > d_0$ for $i = 1, 2, \dots, n$, the linear transformation $y_i = \frac{x_i - d_0}{\sigma}$ ensures that the resulting sample vector has a minimum value that is close to zero. The minimizer of the Anderson-Darling test statistic was determined, no roots being present in the right tail of the density estimate obtained via exponential tilting from an exponential density function as the base density, whose associated degree and tilting parameter are $\hat{d} = 5$ and $\hat{\theta} = 4.8$, which yields an $AD_{\hat{d}, \hat{\theta}}$ value of 11.2854.

3. Density III: Location-shifted TGP obtained from the rescaled data

When the gamma density function is utilized as base density for determining a density estimate via exponential tilting, the fit of the distribution is further improved as the $AD_{\hat{d},\hat{\theta}}$ value is then 0.730527 with $\hat{d} = 9$ and $\hat{\theta} = 3.6$.

4. Density IV: BCP obtained from the rescaled data

On applying the methodology for estimating the BCP that is described in Section 6.3.1, it was determined that a Box-Cox transform parameter value of -1 and a degree 5 for the polynomial adjustment are suitable.

5. Density V: Exponential polynomial obtained from the log-scaled data

As the deductible level, d_0 , is one, other possible density estimates result from the log-scaled data. The summary statistics for the log-scaled data that are included in Table 11 suggest a light-tailed distribution, in which case exponential tilting may not be necessary. Using the exponential density function as base density in conjunction with a polynomial adjustment of degree 9 results in two roots in the right tail. Although the Anderson-Darling test statistics are slightly smaller with adjustments of larger degrees, the number of roots in the tail then increases, which precludes choosing a higher polynomial adjustment degree.

6. Density VI: TEP obtained from the log-scaled data

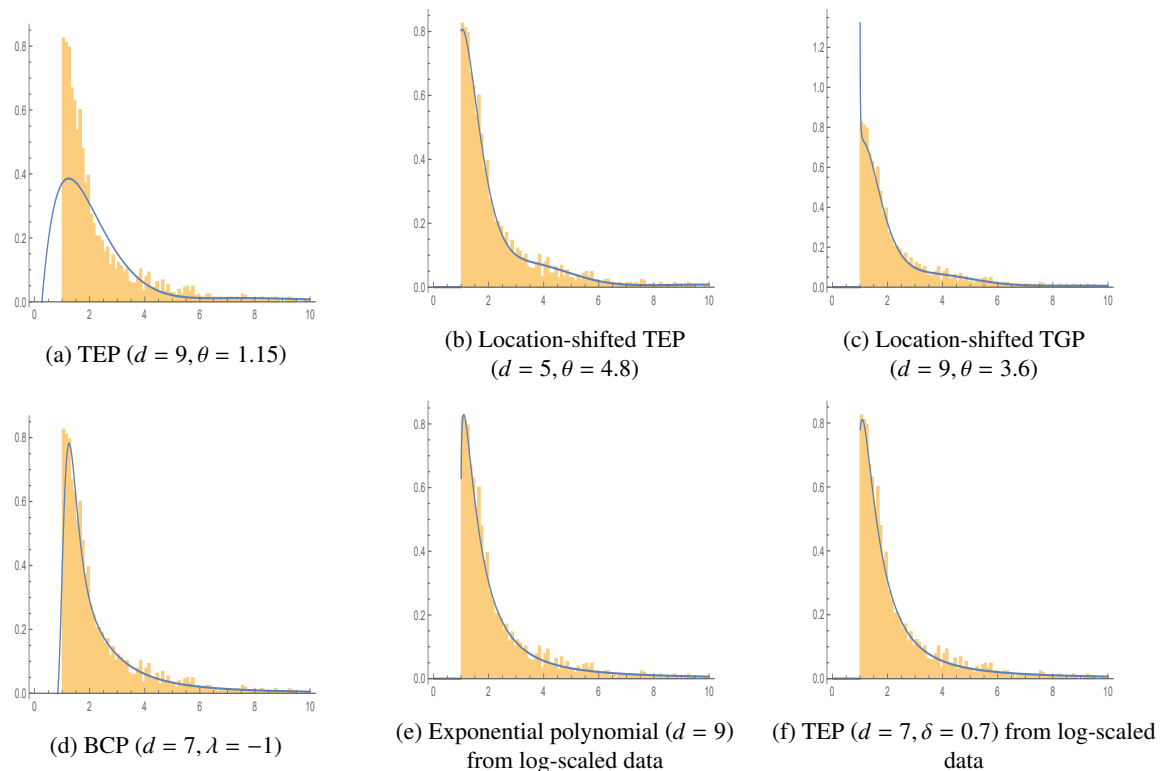


Figure 19. Density estimates and histogram of the original data for the Danish fire data set

A density estimate obtained via exponential tilting for which $d = 7$ and $\theta = 0.7$ was determined and its VaR, compared to those associated with aforementioned density candidates. Table 12 presents specific VaR values associated the various density estimates. Figure 19 includes plots of the density estimates superimposed of a histogram of the original data.

Table 10. Summary of the rescaled Danish fire data (Original standard deviation: 8.52743)

Sample size	2 156	Min	0.117608
Mean	0.398392	1st quartile	0.156057
Skewness	18.7093	2nd quartile	0.208751
Kurtosis	483.462	3rd quartile	0.34858
		Max	30.781

Table 11. Summary of the log-scaled Danish fire data

Sample size	2156	Min	0.00288882
Mean	0.790965	1st quartile	0.285752
S.D.	0.716333	2nd quartile	0.576672
Skewness	1.76413	3rd quartile	1.0894
Kurtosis	7.1908	Max	5.57311

Table 12. VaR_κ values obtained from the sample and the density estimates

κ	Empirical	Density II	Density III	Density IV	Density V	Density VI
0.8	3.48145	3.35407	3.51668	3.32754	3.4303	3.32754
0.9	5.56174	4.87409	5.35254	5.11591	5.70866	5.73878
0.95	10.0111	7.20204	9.32785	7.94532	9.50778	9.7081
0.99	26.2146	12.9562	22.1245	26.3172	27.7974	27.7927

In light of the Anderson-Darling test statistics, the VaR values appearing in Table 12 and parsimony considerations, density estimate VI could be viewed as being the most suitable.

Acknowledgements

The financial support of the Natural Sciences and Engineering Research Council of Canada is gratefully acknowledged. Thanks are also due to the referees.

References

- Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 26(2), 211–252. <https://doi.org/10.1111/j.2517-6161.1964.tb00553.x>
- Bühlmann, H. (1980). An economic premium principle. *ASTIN Bulletin: The Journal of the IAA*, 11(1), 52–60. <https://doi.org/10.1017/s0515036100006619>
- Cox, S. H., Lin, Y., & Wang, S. (2006). Multivariate exponential tilting and pricing implications for mortality securitization. *Journal of Risk and Insurance*, 73(4), 719–736. <https://doi.org/10.1111/j.1539-6975.2006.00196.x>
- Daniels, H. E. (1954). Saddlepoint approximations in statistics. *The Annals of Mathematical Statistics*, 25(4), 631–650. <https://doi.org/10.1214/aoms/1177728652>
- Edgeworth, F. Y. (1905). The law of error. *Cambridge Philosophical Society Transactions*, 20, 33–66.
- Elderton, W. P., & Johnson N. L. (1969). *Systems of frequency curves*. New York, NY: Cambridge University Press.
- Elliott, R. J., Chan, L., & Siu, T. K. (2005). Option pricing and Esscher transform under regime switching. *Annals of Finance*, 1(4), 423–432. <https://doi.org/10.1007/s10436-005-0013-z>
- Esscher, F. (1932). On the probability function in the collective theory of risk. *Skandinavisk Aktuarietidskrift*, 15(3), 175–195. <https://doi.org/10.1080/03461238.1932.10405883>
- Fisher, R. A., & Cornish, E. A. (1960). The percentile points of distributions having known cumulants. *Technometrics*, 2(2), 209–225. <https://doi.org/10.1080/00401706.1960.10489895>
- Frees, E. W. (2011). *Regression modeling with actuarial and financial applications*. New York, NY: Cambridge University Press.
- Gajek, L. (1986). On improving density estimators which are not bona fide functions. *The Annals of Statistics*, 14(4), 1612–1618. <https://doi.org/10.1214/aos/1176350182>
- Gerber, H. U., & Shiu, E. S. W. (1994). Option pricing by Esscher transforms. *Transactions of the Society of Actuaries*, 46, 99–191.
- Ghosh, S., & Resnick, S. (2010). A discussion on mean excess plots. *Stochastic Processes and their Applications*, 120(8), 1492–1517. <https://doi.org/10.1016/j.spa.2010.04.002>
- Goutis C., & Casella G. (1999). Explaining the saddlepoint approximation. *The American Statistician*, 53(3), 216–224. <https://doi.org/10.1080/00031305.1999.10474463>
- Ha, H. (2007). Moment-based density approximation algorithm for symmetric distributions. *The Korean Communica-*

- tions in Statistics, 14(3), 583–592. <https://doi.org/10.5351/ckss.2007.14.3.583>
- Ha, H., & Provost, S. B. (2007). A viable alternative to resorting to statistical tables. *Communications in Statistics–Simulation and Computation*, 36, 1135–1151. <https://doi.org/10.1080/03610910701569671>
- Ha, H., & Provost, S. B. (2008). On the inversion of certain moment matrices. *Linear Algebra and its Applications*, 430(10), 2650–2658. <https://doi.org/10.1016/j.laa.2008.10.024>
- Hill, G. W., & Davis, A. W. (1968). Generalized Asymptotic Expansions of Cornish–Fisher Type. *Annals of Mathematical Statistics*, 39(4), 1264–1273. <https://doi.org/10.1214/aoms/1177698251>
- Jensen, J. L. (1995). *Saddlepoint approximations (No. 16)*. Oxford: Oxford University Press.
- McNeil, A. (1997). Estimating the tails of loss severity distributions using extreme value theory. *ASTIN Bulletin: The Journal of the IAA*, 27(1), 117–137. <https://doi.org/10.2143/ast.27.1.563210>
- Nadarajah, S., & Kotz, S. (2006). On the Laplace transform of the Pareto distribution. *Queueing Systems*, 54, 243–244. <https://doi.org/10.1007/s11134-006-0299-1>
- Provost, S. B. (2005). Moment–based density approximants. *The Mathematica Journal*, 9, 727–756.
- Provost, S. B., & Jiang, M. (2012). Orthogonal polynomial density estimates: Alternative representation and degree selection. *International Journal of Computational and Mathematical Sciences*, 6, 12–29.
- Resnick, S. I. (1997). Discussion of the Danish data on large fire insurance losses. *Astin Bulletin*, 27(1), 139–151. <https://doi.org/10.2143/ast.27.1.563211>
- Siegmund, D. (1976). Importance sampling in the Monte Carlo study of sequential tests. *The Annals of Statistics*, 4(4), 673–684. <https://doi.org/10.1214/aos/1176343541>
- Solomon, H., & Stephens, M. (1978). Approximations to density functions using Pearson curves. *Journal of the American Statistical Association*, 73(361), 153–160. <https://doi.org/10.1080/01621459.1978.10480019>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).