Simulation of the p-generalized Gaussian distribution

S. Kalke and W.-D. Richter

University of Rostock, Mathematical Departement, Ulmenstraße 69, 18057 Rostock, Germany

Abstract. We introduce the p-generalized polar methods for the simulation of the p-generalized Gaussian distribution. On the basis of geometric measure representations, the well known Box-Muller method and the Marsaglia-Bray rejecting polar method for the simulation of the Gaussian distribution are generalized to simulate the p-generalized Gaussian distribution, which fits much more flexible to data than the Gaussian distribution and is already applied in various fields of modern sciences. To prove the correctness of the p-generalized polar methods, we give stochastic representations and to demonstrate their adequacy, we perform a comparison of six simulation techniques w.r.t. the goodness of fit and the complexity. The competing methods include adapted general methods and another special method. Further, we prove stochastic representations for all adapted methods.

Keywords: random numbers, simulation, p-generalized Gaussian distribution, power exponential distribution, exponential error distribution, p-generalized polar method, p-generalized rejecting polar method, stopping time, p-generalized uniform distribution on the p-circle, generalized arc-length measure, goodness of fit, Monte Carlo simulation, Monty Python method, Ziggurat method, tail algorithm

MSC2010: 65C10, 60G40, 68U20

1 Introduction

The p-generalized Gaussian distribution $\mathcal{N}_p$ is also known as power exponential, exponential error or p-generalized normal distribution and was introduced in (Subbotin, 1923) and later on studied , e.g., in (Box and Tiao, 1973) and in (Johnson et al., 1994). It is of special interest in statistical modelling because it fits much more flexible to data than the Gaussian distribution due to an additional parameter p. The role of this parameter can be seen from the $\mathcal{N}_p$-density

$$f_p(x) = \frac{p^{1-1/p}}{2\Gamma\left(\frac{1}{p}\right)} \exp\left[-\frac{|x|^p}{p}\right],\ x \in \mathbb{R},\ p > 0,$$

which is illustrated in Figure 1 and represents in case $p = 1$ the Laplace density, in case $p = 2$ the Gaussian density and approaches for $p \to \infty$ the density of the uniform distribution on $[-1, 1]$.

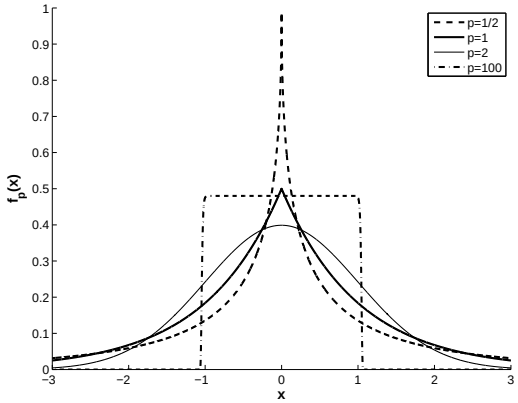

Fig. 1: Univariate density f_p

In the bivariate case the level sets $C_p(r) := \{x \in \mathbb{R}^2 : |x|_p = r\}$ of the product density $f_p(x_1)\, f_p(x_2)$, see Figure 2, are circles with respect to the functional $|(x_1, x_2)|_p := (|x_1|^p + |x_2|^p)^{1/p}$, $(x_1, x_2) \in \mathbb{R}^2$, which is the $l_{2,p}$-norm if $p \geq 1$ and the $l_{2,p}$-anti-norm if $p \in (0, 1)$, see (Moszyńska and Richter, 2011). For this reason, p will be called the form parameter, although it can also be understood as a scale parameter causing heavy or light tails for small or large values, respectively.

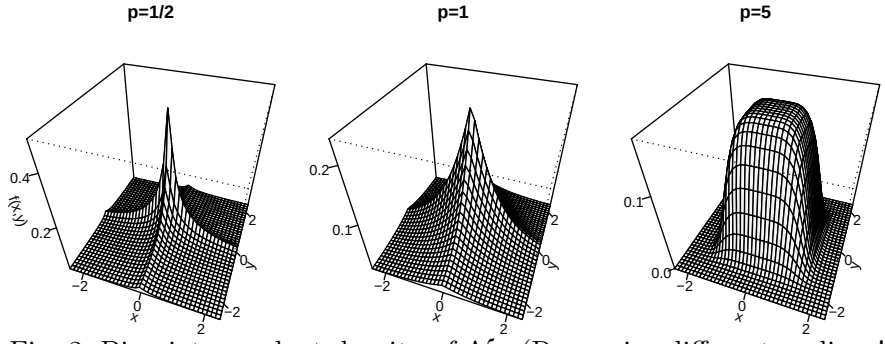

Fig. 2: Bivariate product density of $\mathcal{N}_p$. (Recognize different scalings!)

As a consequence of its attractive flexibility, the p-generalized Gaussian distribution has already received wide spread applications in various areas. According to (Pogány and Nadarajah, 2010) and (Nardon and Pianca, 2009), these areas include signal processing, quantitative finance, medicine, automation and remote control, telecommunication, information systems, biomedical engineering, physics, cybernetics, nanotechnology and many more.
The necessity of effective simulation techniques for the $\mathcal{N}_p$ distribution is highlighted by an application from signal processing, which is of great importance in the respiratory mechanics research area. In (Saatci and Akan, 2010) a respiratory system under non-invasive ventilation is modelled by using measured respiratory signals. Here, the generalized Gaussian distribution is used to model the measurement noise in the respiratory system identification problem. To investigate the adequacy of the model considered this way for different noise scenarios, the authors in (Saatci and Akan, 2010) generate Laplace, Gauss and uniformly distributed random numbers. Using now the p-generalized polar methods presented in this paper, one can simulate the $\mathcal{N}_p$-distribution for every $p > 0$ and is therefore able to test the model for an arbitrary noise scenario.

Let us briefly recall the main mathematical background of the standard polar method and the rejecting polar method for generating Gaussian random numbers. A standard Gaussian random vector (X, Y) satisfies a stochastic representation

$$(X, Y) \stackrel{d}{=} \mathcal{R} \cdot \mathcal{U} \ , \tag{1}$$

which is known as the Box-Muller transformation in the literature. In case of the polar method in (Box and Muller, 1958), $\mathcal{R} = \sqrt{-2 \ln V_1}$ and $\mathcal{U} = (\cos(2\pi V_2), \sin(2\pi V_2))$ are independent, where V_1, V_2 are independent and uniformly distributed random variables on $(0, 1)$.
The rejection method in (Marsaglia and Bray, 1964) uses a conditional probability for alternatively simulating the so called uniform basis vector $\mathcal{U}$, which is uniformly distributed on C_2, where $C_p := C_p(1)$. In this context, $\mathcal{U}$ belongs to a Borel-subset A of C_2 with probability

$$P\left(\mathcal{U} \in A\right) = P\left(\Sigma \frac{(V_1, V_2)}{|(V_1, V_2)|_2} \in A \;\middle|\; |(V_1, V_2)|_2 \leq 1\right) \ , \tag{2}$$

where $\Sigma = diag(S_1, S_2)$ is a random signature matrix which is independent from V_1, V_2 and S_1, S_2 are independent and uniformly distributed on $\{-1, 1\}$. According to this rejection method, random numbers v_1, v_2 are generated until they satisfy $|(v_1, v_2)|_2 \leq 1$. So in principal the generation of one pair of random numbers (u_1, u_2) involves an infinitely long sequence of random numbers which is randomly stopped. Here and throughout this paper random variables will be denoted by capital letters and random numbers by small letters.

The focus of this paper is to p-generalize the well known polar method in (Box and Muller, 1958) and the rejecting polar method in (Marsaglia and Bray, 1964) for the simulation of Gaussian random numbers. In addition, we set out to make various comparisons of our generalized methods with methods known from the literature. Our generalizations are essentially based upon stochastic representations which make use of the p-generalized trigonometric functions in (Richter, 2007), the non-Euclidean geometric measure representation for $l_{n,p}$-spherical symmetric distribution laws in (Richter, 2009) and which make also use of certain standard distributional transformations in (Rubinstein, 1981) and in (Jöhnk, 1964).
The paper is structured as follows. In Section 2, the p-generalized polar methods are introduced. The mathematical fundamentals for these new algorithms will be presented in Section 3. Thereby, conclusions from a non-Euclidean geometric measure representation are given in Subsection 3.1 and in Subsection 3.2, the p-generalized polar methods are proved on the basis of stochastic representations. In Section 4, we adapt general methods from the literature and also consider the special method in (Nardon and Pianca, 2009) for alternatively sampling from the $\mathcal{N}_p$-distribution. In this context, we demonstrate the adequacy and practicability of the new methods derived in Section 2 by comparing six simulation methods w.r.t. goodness of fit and complexity.

2 Two new algorithms

2.1 The p-generalized polar method

We present here the algorithm creating the p-generalized polar method for simulating pairs $(X_1,\, X_2)$ of independent p-generalized Gaussian distributed random variables. Consider therefore the $l_{2,p}$-generalized trigonometric functions $\cos_p$ and $\sin_p$, defined in (Richter, 2007) as

$$\cos_p(x) := \frac{\cos(x)}{N_p(x)} \quad , \quad \sin_p(x) := \frac{\sin(x)}{N_p(x)} \quad ,$$

where

$$N_p(x) := (|\sin(x)|^p + |\cos(x)|^p)^{1/p},\ x \in [0, 2\pi)\,.$$

As in all algorithms of this paper, the uniform choice of random numbers or random tuples is to be understood as a choice of independent random numbers or random tuples, respectively.

Algorithm 1 (The p-generalized polar method)**.**

1. Simulation of a generalized radius variable $\mathcal{R}_p$
   (a) Generate a Gamma distributed random number γ with parameter $2/p$.
   (b) Form $r_p = (p\,\gamma)^{1/p}$.
2. Simulation of a generalized uniform basis vector $\mathcal{U}_p$
   (a) Generate ϕ from the density $f_\Phi(\phi) = 1_{(0,2\pi)}(\phi) \quad \frac{p\,\Gamma\left(\frac{2}{p}\right)}{4\,\Gamma^2\left(\frac{1}{p}\right)} \quad \frac{1}{N_p^2(\phi)}$.
   (b) Build $u_p = (\cos_p \phi, \sin_p \phi)$.
3. Return $(x_1, x_2) = r_p \cdot u_p$.

Here, $\Gamma(.)$ denotes the Gamma function. For further details concerning this algorithm, we refer to Appendix B.

2.2 The p-generalized rejecting polar method

In dependence of the available hard and software, it may be of interest to analyze the influence which calculations of trigonometric functions have onto the time consumed by a simulation algorithm. This was the main reason for modifying the polar method in (Marsaglia and Bray, 1964). On the one hand, because of todays faster computers, the calculations of trigonometric functions may have become less important in a complexity study than at the times when (Box and Muller, 1958) and (Marsaglia and Bray, 1964) appeared. On the other hand, Algorithm 1 involves more general trigonometric functions and in step 2 the simulation of an angle ϕ which is not uniformly distributed unless for $p = 2$. So in the present situation, there is a new reason to study a rejection method modifying Algorithm 1.
In an analogous way as the polar method was modified in (Marsaglia and Bray, 1964) to the rejecting polar method, we introduce now the p-generalized rejecting polar method for generating pairs of independent p-generalized Gaussian distributed random numbers.

Algorithm 2 (The p-generalized rejecting polar method)**.**

1. Follow step 1 from Algorithm 1 to simulate the generalized radius variable $\mathcal{R}_p$.
2. Simulation of the generalized uniform basis vector $\mathcal{U}_p$
   (a) Choose (v_1, v_2) uniformly from $(0,1)^{\times 2}$ until $v_1^p + v_2^p \leq 1$.
   (b) Choose (s_1, s_2) uniformly from $\{-1,1\}^{\times 2}$ and form the signature matrix $\Sigma = diag(s_1, s_2)$.
   (c) Build $u_p = \frac{(v_1,v_2)}{|(v_1,v_2)|_p}\ \Sigma$.
3. Return $(x_1, x_2) = r_p \cdot u_p$.

For further details concerning this algorithm, we refer to Appendix B.

3 Mathematical background

3.1 Conclusions from a geometric measure representation

The polar method for simulating a standard Gaussian distributed random vector (X, Y) in (Box and Muller, 1958) is essentially based on equation (1). A transformation of Cartesian coordinates into polar coordinates delivers a product representation of (X, Y) with a radius $\mathcal{R}$ and a uniform basis $\mathcal{U}$, where $\mathcal{R}$ and $\mathcal{U}$ are independent. Here, the vector $\mathcal{U}$ takes its values in C_2 and the corresponding random polar angle is uniformly distributed on $[0, 2\pi)$. The reason why one also calls $\mathcal{U}$ uniformly distributed is the geometric representation of its distribution

$$P\left(\mathcal{U} \in A\right) = \frac{\mathfrak{U}_2(D)}{\mathfrak{U}_2(C_2)} \quad , \ D \in \mathfrak{B}(\mathbb{R}^2) \cap C_2 \ ,$$

where $\mathfrak{U}_2$ denotes the Euclidean arclength measure. The p-generalized polar method is smilarly to the polar method based on a product representation of the random vector to be simulated. In this context, a tuple of independent p-generalized Gaussian distributed random variables X_1, X_2 satisfies

$$(X_1, X_2) \stackrel{d}{=} \mathcal{R}_p \mathcal{U}_p \, .$$

This may be considered as an analogue to (1) but with a radius $\mathcal{R}_p$ and a uniform basis $\mathcal{U}_p$, which are suitably generalized and still independent. Introducing $l_{2,p}$-generalized polar coordinates as in (Richter, 2007), we obtain

$$\mathcal{R}_p = |(X_1, X_2)|_p \ , \quad \mathcal{U}_p = (\cos_p(\Phi), \sin_p(\Phi)) \ . \tag{3}$$

According to (Richter, 2009), the density of the transformed vector $(\mathcal{R}_p, \Phi)$ satisfies the representation

$$f_{(\mathcal{R}_p,\Phi)}(r,\varphi) = 1_{(0,\infty)}(r)1_{(0,2\pi)}(\varphi) \left(\frac{p^{1-1/p}}{2\Gamma\left(\frac{1}{p}\right)}\right)^2 \frac{r}{N_p^2(\varphi)} \exp\left(-\frac{r^p}{p}\right) ,$$

and the marginal densities satisfy

$$f_{\mathcal{R}_p}(r) = 1_{(0,\infty)}(r) \frac{p^{1-2/p}}{\Gamma\left(\frac{2}{p}\right)} \quad r\exp\left(-\frac{r^p}{p}\right) ,$$

$$f_{\Phi}(\varphi) = 1_{(0,2\pi)}(\varphi) \quad \frac{p\,\Gamma\left(\frac{2}{p}\right)}{4\,\Gamma^2\left(\frac{1}{p}\right)} \quad \frac{1}{N_p^2(\varphi)} \quad . \tag{4}$$

If $p = 2$, f_Φ is the constant density of the uniform distribution on $[0, 2\pi)$ and Φ is the polar angle of $\mathcal{U} = (\cos\Phi, \sin\Phi)$ from equation (1). In the general case $p > 0$, f_Φ is an oscillating $\pi/2$-periodic function on $[0, 2\pi)$, see Figure 4.
While the polar method from (Box and Muller, 1958) is based on the uniform distribution on the circle C_2, it follows from the considerations in (Richter, 2007) and (Richter, 2009), that a certain p-generalized uniform distribution on the p-circle plays an analogous role when considering the p-generalized Gaussian distribution. To this end, it is shown in (Richter, 2009) that the uniform basis $\mathcal{U}_p = (\cos_p \Phi, \sin_p \Phi)$ satisfies the geometric representation

$$P\left(\mathcal{U}_p \in A\right) = \frac{\mathfrak{U}_p\left(A\right)}{\mathfrak{U}_p\left(C_p\right)} \quad , \; A \in \mathfrak{B}(\mathbb{R}^2) \cap C_p, \tag{5}$$

where $\mathfrak{U}_p$ denotes the $l_{2,p}$-generalized arclength measure. This measure differs from the Euclidean arclength measure if $p \neq 2$ and is according to (Richter, 2009) generated by the dual $l_{2,q}$-norm, if $p \geq 1$. In this case holds

$$\mathfrak{U}_p(D) = \int\limits_{G(D)} |N(x)|_q \, dx \quad ,$$

where $1/p + 1/q = 1$, $N(x)$ is the normal vector to the $l_{2,p}$-circle C_p at the point x and $G(D) = \{x_1 \in \mathbb{R} : \; |x_1|^p \leq 1 \wedge (x_1, x_2) \in D\}$. If $0 < p < 1$, $\mathfrak{U}_p$ is generated in the same way but then the functional $|.|_q$ is a semi-anti-norm. For the definition of anti-norms and semi-anti-norms we refer to (Moszyńska and Richter, 2011). A certain characteristic property of $\mathfrak{U}_p$ is demonstrated in Figure 3.

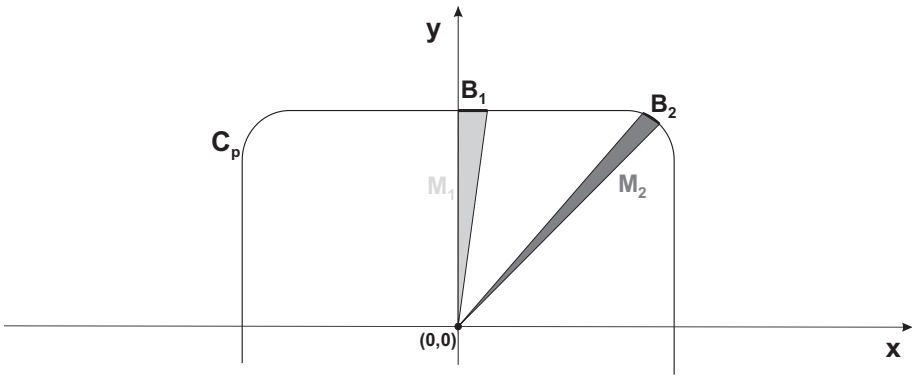

Fig. 3: Sets M_1, M_2 having the same area content correspond to arcs B_1, B_2 from C_p having the same $\mathfrak{U}_p$-lengths but different Euclidean lengths, unless for $p = 2$.

Definition 1.

a) The distribution of a random vector $\mathcal{U}_p$ satisfying equation (5) is called the $l_{2,p}$-generalized uniform distribution on the p-circle C_p.

b) The distribution of the random angle Φ from $\mathcal{U}_p = (\cos_p \Phi, \sin_p \Phi)$ in equation (3) is called the angular distribution corresponding to the $l_{2,p}$-generalized uniform distribution.

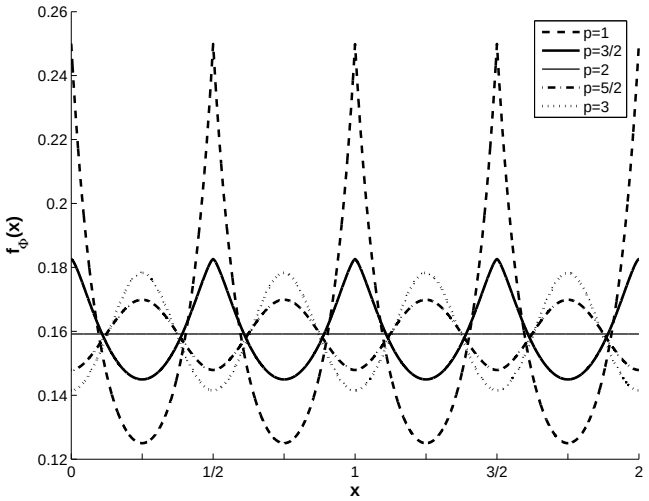

Fig. 4: Density f_Φ of the angular distribution for selected values of p. Values on the axis of abscissae are multiples of π.

The step 2 of Algorithm 1 includes the generation of a random number that follows the angular distribution according to Definition 1. The simulation of Φ will be further reduced to the generation of uniformly distributed random numbers $v_n \in (0,1)$ in Algorithm 7, see Appendix B, which is essentially based upon the following stochastic representation of Φ.

Theorem 1. Let S_1 and S_2 be uniformly distributed on $\{-1,1\}$ and on $\{0,1\}$, respectively, and let (Y_1, Y_2) be $l_{2,p}$-generalized uniformly distributed on $C_p^+ := C_p \cap \mathbb{R}_+^2$, i.e.

$$P\left((Y_1,Y_2) \in A\right) = \frac{\mathfrak{U}_p(A)}{\mathfrak{U}_p(C_p^+)} \quad , A \in C_p^+ \cap \mathfrak{B}(\mathbb{R}^2)\,.$$

If S_1, S_2 and (Y_1, Y_2) are independent random variables, then

$$\Phi \stackrel{d}{=} S_1 \arctan\left(\frac{Y_1}{Y_2}\right) + \pi S_2 + \frac{\pi}{2} \quad .$$

Proof. According to Definition 1, Φ can be interpreted as the polar angle of an $l_{2,p}$-generalized uniformly distributed random vector $\mathcal{U}_p$ taking values in C_p. Therefrom,

$$P\left(\Phi \in A \;\middle|\; \Phi \in (0,\pi/2)\right) = P\left(\arctan\left(\frac{Y_2}{Y_1}\right) \in A\right) \;, \forall A \in \mathfrak{B}\left([0,\pi/2)\right) \;.$$

It also follows from the $\pi/2$-periodicity of f_Φ, that

$$P\left(\Phi \in A \;\middle|\; \Phi \in (0,\pi/2)\right) = P\left(-\arctan\left(\frac{Y_1}{Y_2}\right) + \frac{\pi}{2} \in A\right) \;, A \in B\left((0,\pi/2)\right)$$

$$P\left(\Phi \in A \;\middle|\; \Phi \in (\pi/2,\pi)\right) = P\left(\arctan\left(\frac{Y_1}{Y_2}\right) + \frac{\pi}{2} \in A\right) \;, A \in B\left((\pi/2,\pi)\right)$$

$$P\left(\Phi \in A \;\middle|\; \Phi \in (\pi,3/2\pi)\right) = P\left(-\arctan\left(\frac{Y_1}{Y_2}\right) + \frac{\pi}{2} + \pi \in A\right) \;, A \in B\left((\pi,3/2\,\pi)\right)$$

$$P\left(\Phi \in A \;\middle|\; \Phi \in (3/2\pi,2\pi)\right) = P\left(\arctan\left(\frac{Y_1}{Y_2}\right) + \frac{\pi}{2} + \pi \in A\right) \;, A \in B\left((3/2\,\pi,2\pi)\right)$$

and hence

$$\Phi \stackrel{d}{=} S_1 \arctan\left(\frac{Y_1}{Y_2}\right) + \pi S_2 + \frac{\pi}{2} \quad .$$

□

3.2 Two stochastic representations

We are now in the position to show the correctness of the p-generalized polar method and the p-generalized rejecting polar method. To this end, we prove stochastic representations which correspond to the Algorithms 1 and 2, respectively.

Theorem 2. Let X_1, X_2 be p-generalized Gaussian distributed random variables, Y be a Gamma distributed random variable with parameter $2/p$ and assume the independence of X_1, X_2 and Y. Then

$$(X_1, X_2) \stackrel{d}{=} \mathcal{R}_p \cdot \mathcal{U}_p \quad ,$$

where

$$\mathcal{R}_p = (p\,Y)^{1/p}$$

and

$$\mathcal{U}_p = (\cos_p(\Phi), \sin_p(\Phi))$$

are independent and Φ follows the angular distribution according to Definition 1.

Proof. In view of our consideration in Subsection 3.1 it remains to determine the distribution of $\mathcal{R}_p$:

$$f_{\mathcal{R}_p}(r) = 1_{(0,\infty)}(r)\, \frac{p^{1-2/p}}{\Gamma\left(\frac{2}{p}\right)} \quad r \exp\left(-\frac{r^p}{p}\right) \ ,$$

so that

$$\begin{aligned} P\left(\frac{\mathcal{R}_p^p}{p} < r\right) &= 1_{(0,\infty)}(r)\, \frac{p^{1-2/p}}{\Gamma\left(\frac{2}{p}\right)} \int_0^{(pr)^{1/p}} s \exp\left(-\frac{s^p}{p}\right) ds \\ &= 1_{(0,\infty)}(r)\, \frac{1}{\Gamma\left(\frac{2}{p}\right)} \int_0^r y^{2/p-1} \exp\left(-y\right) dy \\ &= 1_{(0,\infty)}(r)\; P\left(Y < r\right) \quad . \end{aligned}$$

□

The p-generalized rejecting polar method is based upon the stochastic representation of Theorem 3 below. The second step of Algorithm 2 deals with a rejection method for the simulation of the uniform basis $\mathcal{U}_p$, that means random numbers are generated until a certain condition is fulfilled, see Figure 5. So in principal one deals with an infinetly long sequence $(V_n)_{n\in\mathbb{N}} = (V_{n,1}, V_{n,2})$ of independent random vectors which is stopped at the random time

$$\tau^A := \inf\left\{n \in \mathbb{N} : \; V_n \in A\right\} \quad , \tag{6}$$

where A is a Borel set denoting the acceptance region and $\mathbb{N} = \{0, 1, 2, \ldots\}$. Notice, that τ^A is a stopping time and that the stopping element

$$V_{\tau^A} = (V_{\tau^A,1}, V_{\tau^A,2}) := \sum_{n=0}^{\infty} 1_{\{\tau^A = n\}} V_n \tag{7}$$

is a measurable function, see Appendix A for more details.

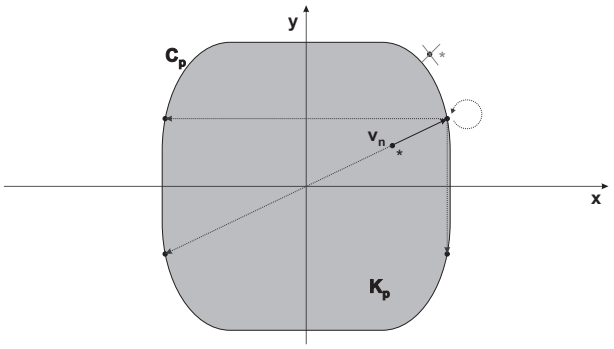

Fig. 5: Generating the $l_{2,p}$-generalized uniform distribution on the $l_{2,p}$-unit circle C_p by stopping, projecting and randomly reflecting $(v_n)_{n\in\mathbb{N}}$.

Additionally to the assumptions of Theorem 2, $\Sigma := diag(S_1, S_2)$ is a random signature matrix with S_1, S_2 uniformly distributed on $\{-1, 1\}$ and $(V_n)_{n\in\mathbb{N}} = (V_{n,1}, V_{n,2})_{n\in\mathbb{N}}$ is a sequence of uniformly on $(0, 1)^{\times 2}$ distributed random vectors. Let also X_1, X_2, Y, S_1, S_2 and $(V_n)_{n\in\mathbb{N}}$ be stochastically independent.

Theorem 3. The random vector (X_1, X_2) satisfies a representation with independent $\mathcal{R}_p$ and $\mathcal{U}_p$

$$(X_1, X_2) \stackrel{d}{=} \mathcal{R}_p \cdot \mathcal{U}_p \,,$$

where

$$\mathcal{R}_p = (p\, Y)^{1/p}$$

and

$$\mathcal{U}_p = \Sigma\ V_{\tau^A} / |V_{\tau^A}|_p \,.$$

Here, the acceptance region for defining the stopping time of the sequence $(V_n)_{n\in\mathbb{N}}$ is the p-disc $A = K_p = \left\{x \in \mathbb{R}^2 : |x|_p \leq 1\right\}$.

Proof. Let $T : (0, \infty) \times (0, 2\pi) \to \mathbb{R}^2$ be the $l_{2,p}$-generalized polar coordinates transformation

$$T(r, \varphi) = r\,(\cos_p \varphi, \sin_p \varphi)\ ,$$

respectively T^* it's restriction to C_p, and $B^+ \in \mathfrak{B}(\mathbb{R}^2_+) \cap C_p$. Assume $K_p^+(r) = \{x \in \mathbb{R}^2_+ : |x|_p \leq r\}$. The Jacobian of the transformation T is according to (Richter, 2007)

$$J(r, \varphi) = \frac{r}{N_p^2(\varphi)}\ ,\ (r, \varphi) \in (0, \infty) \times (0, 2\pi).$$

Defining the central projection cone $CPC(B^+) := \{(x,y) \in \mathbb{R}^2 : \ (x,y)/(|x|^p + |y|^p)^{1/p} \in B^+\}$, we obtain

$$P\left(V_0/|V_0|_p \in B^+ \ \middle| \ |V_0|_p \leq 1\right) = \left[\int\limits_{CPC(B^+)\cap K_p^+(1)} \mu(d(x,y))\right] \cdot \left[\int\limits_{K_p^+(1)} \mu(dz)\right]^{-1} \quad ,$$

where μ denotes the Lebesgue measure on $\mathfrak{B}(\mathbb{R}^2)$. Changing Cartesian with p-generalized polar coordinates from (Richter, 2007), it follows

$$\begin{aligned} P\left(V_0/|V_0|_p \in B^+ \ \middle| \ |V_0|_p \leq 1\right) &= \left[\int\limits_{r=0}^{1} \left(\int\limits_{(T^*)^{-1}(B^+)} \frac{r}{N_p^2(\varphi)} d\varphi\right) dr\right] \cdot \left[\int\limits_{r=0}^{1} \left(\int\limits_{(T^*)^{-1}\left(C_p\cap\mathbb{R}_+^2\right)} \frac{r}{N_p^2(\varphi)} d\varphi\right) dr\right]^{-1} \\ &= \left[\int\limits_{(T^*)^{-1}(B^+)} f_\Phi(\varphi) d\varphi\right] \cdot \left[\int\limits_{(0,\pi/2)} f_\Phi(\varphi) d\varphi\right]^{-1} \\ &= 4 \ P\left(\Phi \in (T^*)^{-1}(B^+)\right) \\ &= 4 \ P\left(\mathcal{U}_p \in B^+\right) \end{aligned}$$

and hence

$$P\left(\Sigma\, V_0/|V_0|_p \in B \,\middle|\, |V_0|_p \leq 1\right) = \ P\left(\mathcal{U}_p \in B\right) \quad , \forall B \in \mathfrak{B}\left(C_p\right) .$$

The proof is complete if we show that

$$P\left(V_0/|V_0|_p \in B \,\middle|\, |V_0|_p \leq 1\right) = P\left(V_{\tau^A}/|V_{\tau^A}|_p \in B\right) \quad , \forall B \in \mathfrak{B}\left(C_p\right) .$$

This will be done for a general acceptance region A in Lemma 1, see Appendix A. □

4 Comparison with the literature

4.1 Alternative simulation techniques

In this subsection, we refer to alternative algorithms for sampling from the p-generalized Gaussian distribution. On the one hand, there are general simulation methods which can be adapted to the special case of the $\mathcal{N}_p$-distribution. On the other hand, the simulation of the p-generalized Gaussian distribution has already been the subject of former studies. In this context, Nardon and Pianca introduced in (Nardon and Pianca, 2009) a special method which reduces the simulation of the $\mathcal{N}_p$-distribution to the generation of Gamma distributed random numbers. Moreover, Malham and Wiese introduced in (Malham and Wiese, 2010) a simulation method for n-tuples of independent, generalized Gaussian distributed random variables under the name generalized Marsaglia polar method. In fact, this method is strongly limited to the special case that the form parameter is equal to the tuple-dimension $n \in \mathbb{N}$ and will therefore play no further role in our investigations.
Besides the method of Nardon and Pianca, we will compare the p-generalized polar methods with adaptions of the Monty Python method, the Ziggurat method and a numerical version of the inversion method. The Monty Python method in (Marsaglia, 1998) and the Ziggurat method in (Marsaglia and Tsang, 2000) can in principal be used to sample from every symmetric density which is strictly montonic decreasing for $x > 0$. However, there is a considerable effort for setting up this algorithms, especially if $p < 1$. Among other things, this is due to the fact that both simulation techniques make use of the tail algorithm to sample from the tail of the distribution of interest, see (Marsaglia and Tsang, 1984).

The inversion method is probably the most common simulation method in this comparison. It is based upon an inversion of the distribution function F_X and can in principal be applied to every continuous univariate distribution. Practically, the adaption can be very difficult if an exact representation of the inverse distribution function is not available. In such situations, one often uses a numerical version of the inversion method due to a transformation of the inversion equation

$$F_X^{-1}(u) = x$$

to the equivalent equation

$$F_X(x) - u = 0\,. \tag{8}$$

In case of the $\mathcal{N}_p$-distribution, we choose a random number u uniformly from $(0,1)$ and solve equation (8) numerically with the help of the R-procedure uniroot, see (Brent, 1973).

The method of Nardon and Pianca, introduced in (Nardon and Pianca, 2009), is given in Algorithm 3.

Algorithm 3 (The method of Nardon and Pianca)**.**

1. (a) Generate a Gamma distributed random number γ with parameter $1/p$.
   (b) Form $r = (p\,\gamma)^{1/p}$.
2. Choose σ uniformly from $\{-1, 1\}$.
3. Return $x = \sigma\, r$.

Notice, that the random numbers γ in Algorithm 1 and 2 follow Gamma distributions with different parameters and that the outcomes of both algorithms possess different interpretations. This becomes obvious even from the fact that the method of Nardon and Pianca works in one dimension and generates single random numbers while the polar methods work in two dimensions and generate pairs of random numbers.

The adaptions of the Monty Python method and the Ziggurat method make both use of the adapted tail algorithm for sampling from a random variable Y, where Y follows the conditional distribution of the absolute value of $X \sim \mathcal{N}_p$ under the condition $|X| > \xi$. Let therefore $\beta > 1$ and $\psi > 0$ be suitably chosen constants, such that

$$f_p(\xi + t) \le f_p(\xi)\ (1 + \psi t)^{-\beta}\ ,\ \forall t \ge 0,$$

if $0 < p < 1$, see Appendix D for more details.

Algorithm 4 (The tail algorithm adapted to the p-generalized Gaussian distribution)**.**

1. If $p \ge 1$, then choose (v_1, v_2) uniformly from $(0,1)^{\times 2}$ until $v_2 < v_1 \exp\left[\frac{\xi^p}{p} - \frac{1}{p}\left(\xi - \frac{\ln v_1}{\xi^{p-1}}\right)^p\right]$. Then return $\left(\xi - \frac{\ln v_1}{\xi^{p-1}}\right)$.
2. If $0 < p < 1$, then choose (v_1, v_2) uniformly from $(0,1)^{\times 2}$ until $v_2\ e^{-\frac{\xi^p}{p}} v_1^{\beta/(\beta-1)} \le exp\left[-\frac{1}{p}\left(\xi + \frac{1}{\psi}\left(v_1^{1/(1-\beta)}\right)\right)\right]$. Then return $\left(\xi + \frac{1}{\psi}\left(v_1^{1/(1-\beta)}\right)\right)$.

The Monty Python and the Ziggurat method are based on a theorem that we prove for the sake of completeness in Appendix C. It says that the first component of a random vector $X = (X_1, X_2)$, which is uniformly distributed on the set $A(f) := \{(x, y) \in \mathbb{R}^2 : 0 < y < f(x)\}$, has density f.
In case of the Monty Python method, the uniform distribution on the region $A(f)$ is generated by choosing a random tuple uniformly from a rectangle $R(b) := (0, b) \times (0, 1/b)$ and mapping it to $A(f)$

with a suitable transformation $T = (T_1, T_2)$, see Figure 6. Here, f is a strictly monotonic decreasing density of a positive random variable. In case of a symmetric distributed random variable X, a generated realisation of $|X|$ will be multiplied with a random sign, uniformly chosen from $\{-1, 1\}$. To this end, the inverse of the map T rotates the cap $\tilde{H}$ of $A(f)$ w.r.t. the centre $(a, 1/b)$, scales it with factor $s := a/(b-a)$ and stretches it with factor $1/s$, so that the resulting set fits in the upper right corner of $R(b)$.

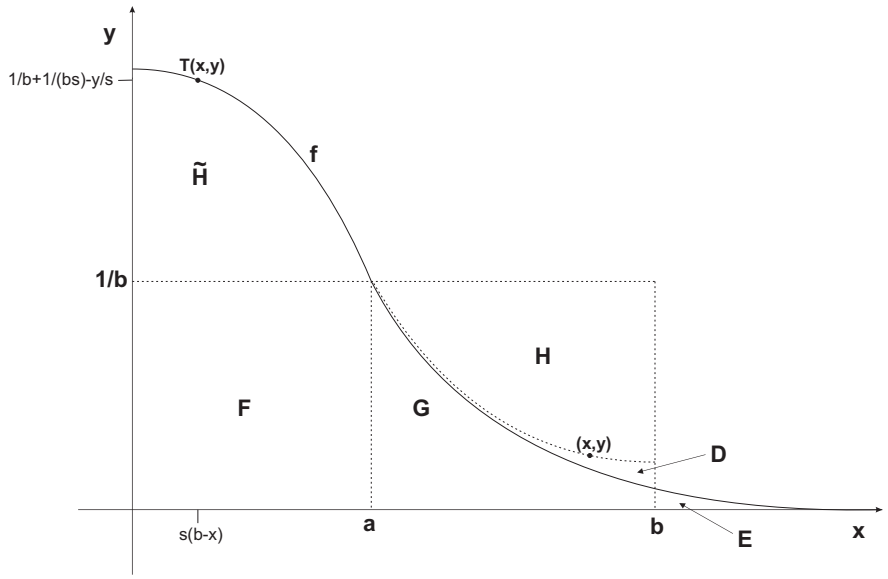

Fig. 6: Monty Python method for suitably chosen b.

The following algorithm reflects the adaption of the Monty Python method to the p-generalized Gaussian density f_p using a stretching factor $s := a/(b-a)$, see Figure 6, where $f = 2\, f_p$. Due to the fact that $X \sim \mathcal{N}_p$ has a symmetric distribution, a random sign σ has to be chosen uniformly from $\{-1, 1\}$ to transform a realisation of $|X|$ to a realisation of X.

Algorithm 5 (Adapted Monty Python method)**.**

1. Choose a random tuple $(\tilde{x}, \tilde{y})$ uniformly from $R(b)$ and a random sign σ uniformly from $\{-1, 1\}$.
2. If $(\tilde{x}, \tilde{y})$ is in F, i.e. $\tilde{x} < a$, then return $\sigma\, \tilde{x}$.
3. If $(\tilde{x}, \tilde{y})$ is in G, i.e. $\tilde{y} < 2\, f_p(\tilde{x})$, then return $\sigma\, \tilde{x}$.
4. If $(\tilde{x}, \tilde{y})$ is in H, i.e. $\tilde{y} > \frac{1}{b} + \frac{s}{b} - 2\, s\, f_p(s(b-\tilde{x}))$, then return $\sigma \cdot s \cdot (b-\tilde{x})$.
5. Otherwise generate the first component of the uniformly on E distributed tuple $(\hat{x}, \hat{y})$ with Algorithm 4 and return $\sigma\, \hat{x}$.

Let us finally consider the Ziggurat method, which originates from another basic theorem on a uniformly distributed random vector $X = (X_1, X_2)$. As can be seen in (Marsaglia and Tsang, 2000) and in Appendix C, the conditional distribution of X under the condition that $X \in A \subset IM(X)$ is the uniform distribution on A, if $\mu(A) > 0$. In this context, the algorithm in (Marsaglia and Tsang, 2000) overlaps the region $A(f)$ with m rectangles $R_1, \ldots, R_m$ and a tail area E, see Figure 7, where $R_1, \ldots, R_{m-1}, R_m \cup E$ have the same area content v. After selecting one of the areas $R_1, \ldots, R_{m-1}, R_m \cup E$ uniformly with probability $1/m$, one chooses (x, y) uniformly from there and returns x if $(x, y) \in A(f)$. Otherwise, a new area will be selected at random.
Here, f is a strictly monotonic decreasing density of a positive random variable X, and as in Algorithm 5, the simulation of a positive random variable can be generalized to the simulation of a symmetric distributed random variable by choosing a random sign σ uniformly from $\{-1, 1\}$.

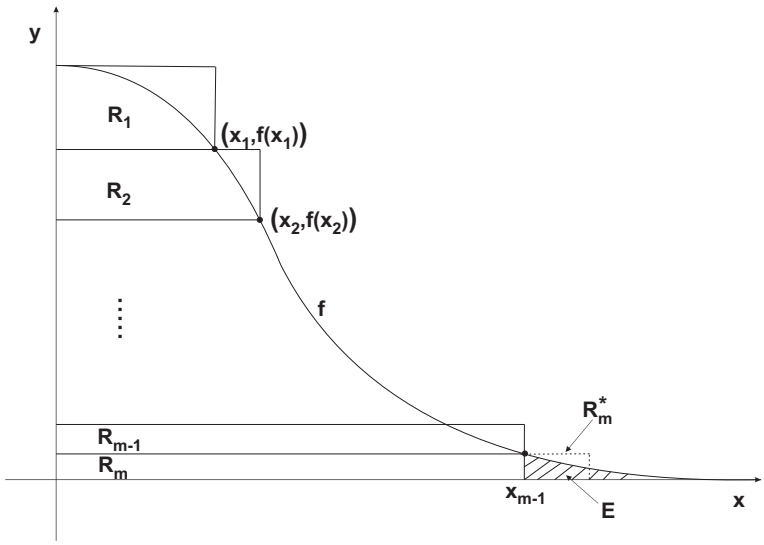

Fig. 7: Ziggurat method: m rectangles and E overlap $A(f)$.

The m-th area of the Ziggurat plays a special role in the algorithm from (Marsaglia and Tsang, 2000). Here, the random tuple (x, y) is chosen uniformly from the rectangle R_m^* (having the area content v) and will be rejected, if it is outside the subset R_m. Then the first component of a random tuple $(\tilde{x}, \tilde{y})$, uniformly distributed on E, will be generated with Algorithm 4.
Provided that $A(f_p^*)$ is overlapped with a Ziggurat in the described way, where $f_p^* = 2(f_p)\big|_{\mathbb{R}_+}$, Algorithm 6 reflects the adapted method from (Marsaglia and Tsang, 2000) for the simulation of the $\mathcal{N}_p$-distribution.

Algorithm 6 (Adapted Ziggurat algorithm)**.** Let $x_1, \ldots, x_{m-1}$ be the x-coordinates defining $R_1, \ldots, R_m$ and assume $v = 2\,x_{m-1} f_p(x_{m-1}) + 2 \int\limits_{x_{m-1}}^{\infty} f_p(x)\,dx$, $x_0 = 0$ and $x_m = v/(2\,f_p(x_{m-1}))$.

1. Choose z uniformly from $\{1, \ldots, m\}$ and σ uniformly from $\{-1, 1\}$.
2. Choose u_1 uniformly from $(0, 1)$.
3. If $x_z u_1 < x_{z-1}$, then return $\sigma\, x_z u_1$. Otherwise go on with 4 if $z = m$ and with 5 if $z < m$.
4. Generate $\tilde{x} \geq x_{m-1}$ with the tail algorithm, see Algorithm 4, and return $\sigma\, \tilde{x}$.
5. Choose u_2 uniformly from $(0, 1)$ and return $\sigma\, x_z u_1$, if $2\,f_p(x_z) + \frac{v}{x_z} u_2 < 2\,f_p(x_z u_1)$. Otherwise repeat 2.

4.2 Goodness of fit comparison

In this subsection, we compare the goodness of fit of six different techniques for simulating the p-generalized Gaussian distribution: the two p-generalized polar methods, the method of Nardon and Pianca, the Monty Python method, the Ziggurat method and the numerical inversion method. All simulation techniques are implemented in an R-program, which uses the Mersenne-Twister algorithm for the generation of the uniform distribution on $(0, 1)$. To verify the correctness of this program respectively of each simulation method, the Kolmogoroff Smirnoff test and a Monte Carlo Simulation will be used. In the last case, we compare exact probabilities of X belonging to a certain set with the corresponding approximations of a Monte Carlo simulation.

Test 1. In context of a Kolmogoroff-Smirnoff-test, we generated $m = 1000$ samples of size n with each method, i.e. we generated $(x_{i,j})_{i=1:1000, j=1:n}$, where $n \in \{10, 50, 5 \cdot 10^2, 10^3, 5 \cdot 10^3, 10^4, 5 \cdot$

$10^4, 10^5\}$ and $p \in \{0.25, 0.5, 0.75, 1, 1.5, 2, 2.5, 3, 5, 8\}$. Then the Kolmogoroff-Smirnoff test statistic $\hat{D}(x_{i,1}, \ldots, x_{i,n})$ of each sample was compared with the corresponding quantile $D_{1-\alpha,n}$ according to the significance level $\alpha \in \{0.01, 0.05\}$. The relative rejection frequency $h(p, m, n, \alpha)$ is for fixed $n \in \mathbb{N}$ and $p \in \mathbb{R}$ the mean of the sample $\tilde{x}_1, \ldots, \tilde{x}_{1000}$, where

$$\tilde{x}_i := \begin{cases} 1 & \text{, if } \hat{D}(x_{i,1}, \ldots, x_{i,n}) > D_{1-\alpha,n} \\ 0 & \text{, else} \end{cases} .$$

It approximates α if the simulation technique works correctly and can therefore be considered as a measure for the correctness of an algorithm w.r.t. the sample size n and the form parameter p. In this context, the variance $h(p, m, n, \alpha) \cdot (1 - h(p, m, n, \alpha))$ of the sample $\tilde{x}_1, \ldots, \tilde{x}_{1000}$ similarly reflects the goodness of fit of an algorithm and could alternatively be used for the following comparison.

n	10	50	10^2	$5 \cdot 10^2$	10^3	$5 \cdot 10^3$	10^4	$5 \cdot 10^4$	10^5
p-gen. polar method (pgp)	0.048	0.057	0.048	0.058	0.070	0.053	0.038	0.059	0.050
p-gen. rejecting polar method (pgrp)	0.049	0.050	0.042	0.043	0.048	0.056	0.052	0.044	0.067
Monty Python (mp)	0.041	0.055	0.051	0.045	0.041	0.056	0.041	0.062	0.042
Ziggurat (zig)	0.032	0.047	0.048	0.038	0.052	0.073	0.054	0.057	0.061
Nardon/ Pianca (np)	0.048	0.037	0.047	0.051	0.055	0.046	0.047	0.039	0.063
numerical inversion (ninv)	0.040	0.042	0.049	0.042	0.040	0.051	0.056	0.058	0.042

Table 1: Values of $h(5, 1000, n, 0.05)$ for selected values of n w.r.t. Test 1.

Result: As can be seen exemplarily in Table 1, systematic differences in the performance of the simulation methods are not obvious for any combination of α and p. On the one hand, Table 1 shows for fixed $n \in \mathbb{N}$ rankings of the six algorithms that differ for different sample sizes. On the other hand, a repetition of Test 1, that means another simulation of $(x_{i,j})_{i=1:1000, j=1:n}$, showed relative frequencies $h(p, m, n, \alpha)$ leading to a different ranking of the six simulation algorithms for the same fixed sample size $n \in \mathbb{N}$, see Table 2. In consequence, all simulation methods performed similarly in Test 1. Moreover, the little deviations of the relative rejection frequencies from α indicate the correctness of these methods.

n	10	50	10^2	$5 \cdot 10^2$	10^3	$5 \cdot 10^3$	10^4	$5 \cdot 10^4$	10^5
p-gen. polar method (pgp)	0.042	0.053	0.046	0.045	0.047	0.057	0.042	0.060	0.054
p-gen. rejecting polar method (pgrp)	0.057	0.053	0.042	0.055	0.060	0.041	0.046	0.053	0.050
Monty Python (mp)	0.054	0.063	0.057	0.042	0.051	0.042	0.051	0.051	0.051
Ziggurat (zig)	0.054	0.043	0.056	0.045	0.049	0.058	0.040	0.042	0.066
Nardon/ Pianca (np)	0.042	0.042	0.051	0.061	0.037	0.047	0.045	0.050	0.048
numerical inversion (ninv)	0.042	0.051	0.033	0.055	0.035	0.058	0.060	0.057	0.044

Table 2: Values of $h(5, 1000, n, 0.05)$ in a repetition of Test 1.

Test 2 (Monte Carlo Simulation)**.** Analogously to Test 1, we generated $m = 100$ samples $(x_{i,j})_{i=1:100, j=1:n}$ of size n with each simulation technique. We did this for $p \in \{0.25, 0.5, 0.75, 1, 1.5, 2, 2.5, 3, 5, 8\}$ and $n \in \{50, 250, 5 \cdot 10^2, 10^3, 2 \cdot 10^3, 5 \cdot 10^3, 10^4, 5 \cdot 10^4, 10^5\}$. The relative frequency

$$h(p, m, n) = \# \left\{ (x_{i,j})_{j=\overline{1,n}} : \, i \in \{1, \ldots, m\} \, , \, |h_{n,i}(A) - P((X_1, X_2) \in A)| < \epsilon \right\} / m$$

of adequate Monte Carlo approximations $h_{n,i}(A)$ for the exact probability $P((X_1, X_2) \in A)$ and approximation accuracies $\epsilon \in \{10^{-2}, 10^{-3}, 10^{-4}\}$ was observed, where X_1, X_2 are independent and p-generalized Gaussian distributed. In this context, two types of sets A were used. The first type was a set $K_p(1)$ close to the distribution centre, where $K_p(r) = \{x \in \mathbb{R}^2 : \, |x|_p \leq r\}$. The second type was a set $\mathbb{R}^2 \backslash K_p(r)$ from the distribution tail, where the exact probability of (X_1, X_2) belonging to $\mathbb{R}^2 \backslash K_{2,p}(r)$ is sufficiently small. To this end, we have chosen arbitrarily a radius $r > 0$ (for every fixed p) which satisfies the condition $10^{-3} < P\left((X_1, X_2) \in \mathbb{R}^2 \backslash K_{2,p}(r)\right) < 10^{-2}$, because smaller exact probabilities cause approximations with large variances for the selected sample sizes.

Let us consider the results in Table 3. The relative frequencies for sets of the first type are similar for every simulation technique in the case $p = 5$ and do not point out any significant and systematic goodness of fit differences between the six generation methods.

ϵ	10^{-2}						10^{-3}			10^{-4}
n	50	10^2	$2.5 \cdot 10^2$	$5 \cdot 10^2$	10^3	$2 \cdot 10^3$	$5 \cdot 10^3$	10^4	$5 \cdot 10^4$	10^5
pgp	0.12	0.21	0.29	0.49	0.51	0.70	0.12	0.21	0.45	0.10
pgrp	0.12	0.21	0.30	0.39	0.49	0.71	0.15	0.23	0.45	0.12
mp	0.08	0.23	0.37	0.36	0.53	0.65	0.16	0.22	0.40	0.05
zig	0.14	0.20	0.32	0.37	0.58	0.71	0.10	0.19	0.44	0.06
np	0.13	0.22	0.30	0.48	0.51	0.63	0.11	0.25	0.38	0.08
ninv	0.14	0.16	0.24	0.41	0.52	0.70	0.13	0.13	0.38	0.09

Table 3: Values of $h(5, 100, n)$ for selected values of n and approximation accuracies $\epsilon \in \{10^{-2}, 10^{-3}, 10^{-4}\}$ in case of a test set close to the distribution centre.

Similar effects can be seen in a test using sets of the second type (see Table 4) and also by varying the form parameter p within the above specified range. In consequence, the Monte Carlo Simulation showed that the six simulation techniques all work in nearly the same way correctly but it cannot be used to point out any goodness of fit differences.

ϵ	10^{-3}					10^{-4}	
n	$5 \cdot 10^2$	10^3	$2 \cdot 10^3$	$5 \cdot 10^3$	10^4	$5 \cdot 10^4$	10^5
pgp	0.40	0.50	0.74	0.98	1.00	0.44	0.58
pgrp	0.26	0.66	0.77	0.88	1.00	0.42	0.62
mp	0.34	0.55	0.84	0.93	0.99	0.47	0.55
zig	0.38	0.60	0.72	0.92	0.98	0.39	0.65
np	0.32	0.58	0.83	0.95	0.99	0.44	0.62
ninv	0.38	0.45	0.75	0.88	0.96	0.49	0.50

Table 4: Values of $h(5, 100, n)$ for selected values of n and approximation accuracies $\epsilon \in \{10^{-3}, 10^{-4}\}$ in case of a test set from the distribution tail.

4.3 Complexity comparison

In this subsection, the six methods for the simulation of the p-generalized Gaussian distribution will be compared w.r.t. to the complexity. In addition to the total computing time of each algorithm, the effort for the adaption of the general methods will be considered.

Remark 1. As can be seen in Appendix D, E and F, the use of the general methods for the simulation of the p-generalized Gaussian distribution is connected with some effort for the adaption of these methods.
In case of the Monty Python method, a suitable rectangle width b has to be found by solving an optimization problem, see Appendix E. The Ziggurat algorithm needs the rectangle coordinates $x_1, \ldots, x_m$, which have to be evaluated with a try and error method, see Appendix F.
Furthermore, the Monty Python method and the Ziggurat method both use the tail algorithm, which requires in the case $p \in (0, 1)$ the solution of another optimization problem, see Appendix D for more details. Even in case of the numerical inversion method there is an effort for the adaption. A range for the exact solution of equation (8) is needed for the computation of a numerical solution $\tilde{x}$.
The special methods are in consequence of this setting up efforts more efficient than the general methods in case of a small sample size, even if the total computing time of an already adapted general method is very small.

Test 3. The total computing time for the generation of $n = 10^6$ p-generalized Gaussian distributed random numbers, i.e. $5 \cdot 10^5$ pairs of numbers in case of the p-generalized polar methods, was

determined for each simulation technique with the help of the R-function system.time, which returns for a valid R-instruction among other things the CPU time (in seconds) exclusively charged for the execution of the mentioned instruction, see (Becker et al., 1988) for more details. The adaption efforts for the general methods were not considered.

p	0.25	0.45	0.75	1	1.5	2	3	5	8
pgp	309.25	58.93	37.23	19.16	31.14	15.89	28.06	27.81	27.33
pgrp	313.38	59.34	37.00	19.16	30.80	15.89	27.77	27.51	27.00
mp	78.06	62.70	52.34	46.13	43.09	41.59	38.50	36.85	35.16
zig	36.44	35.19	34.39	33.67	33.28	33.03	32.74	32.47	32.34
np	16.17	42.84	43.15	15.52	40.55	39.98	40.33	39.56	39.07
ninv	614.89	531.77	456.36	404.46	382.36	359.09	326.17	314.64	291.91

Table 5: Total computing times (in seconds) for sample size $n = 10^6$ and selected values of the form parameter p.

As can be seen in Table 5, the numerical inversion method is for each value of p significantly slower than all other methods. Furthermore, the Ziggurat algorithm and the method of Nardon and Pianca seem to be more efficient than the p-generalized polar methods, if p is sufficiently small. For bigger values of p, the p-generalized polar methods are the most efficient simulation techniques.

4.4 Discussion

The focus in this section was to investigate the adequacy and practicability of the two new methods for the simulation of the $\mathcal{N}_p$-distribution introduced in Section 2. For this purpose, the p-generalized polar methods were compared in different tests with four other simulation methods for the p-generalized Gaussian distribution. On the one hand, the tests included goodness of fit comparisons studying the quality or adequacy of generated random numbers. On the other hand, the practicability of simulation methods is closely connected with their complexity, which was also subject of the mentioned tests.
Test 1 showed that all simulation techniques work correctly, because there were little deviations of the relative rejection frequencies from the rejection probability α in case of a correct simulation method. The quality of the random numbers generated by the six algorithms was also confirmed by Test 2, because exact probabilities of sets from the distribution centre and from the distribution tail could be approximated very accurate, if the sample size was large enough.
While both tests cannot be used to rank the generation methods due to a pretty similar performance, in contrast, Test 3 showed considerable differences in the practicability of the six methods. In this context, the numerical inversion method seems to be very unefficient even if the time for the set up of this method is neglected. It also becomes apparent that the p-generalized polar methods are not preferable in the case that p is sufficiently small. If this is not the case, the p-generalized polar methods, the method from Nardon and Pianca, the adapted Monty Python method and the adapted Ziggurat method perform on a similar level, where the two new methods from Section 2 were in all tests the most efficient ones if $p \geq 1.5$. Remembering that the adapted Monty Python method and the adapted Ziggurat method should not be used for small sample sizes, compare Remark 1, there are many situations in which the p-generalized polar methods are the preferred generation algorithms for sampling from the p-generalized Gaussian distribution. In this connection, the choice of a simulation technique w.r.t. the sample size and the form parameter could be made according to Table 6.

sample size	form parameter		
	$0 < p < 0.75$	$0.75 \leq p < 1.5$	$p \geq 1.5$
small	Nardon/Pianca	p-gen. polar methods	p-gen. polar methods
large	Ziggurat	Ziggurat	p-gen. polar methods

Table 6: Most efficient simulation technique w.r.t. the sample size n and the form parameter p.

A Rejection methods and stopping elements

Here, we give a probabilistic description of the announced connection between conditional probabilities as in (2) and the distribution of stopping elements.
As indicated in Subsection 3.2, rejection methods make use of stopping elements and conditional probabilities and include a successsive generation of random numbers until a certain condition is fulfilled. In case of the rejecting polar method, on $(0,1)^{\times 2}$ uniformly distributed (v_1, v_2) are generated until they are in the unit disc K_2, see (2), i.e. a sequence $(V_n)_{n\in\mathbb{N}} = (V_{n,1}, V_{n,2})_{n\in\mathbb{N}}$ of independent, on $(0,1)^{\times 2}$ uniformly distributed random vectors gets stopped at the random time

$$\tau^A := \inf\{n \in \mathbb{N} : \ V_n \in A\} \quad , \tag{9}$$

where $A = K_2$ represents the event of acceptance. The random index of the first element from the sequence $(V_n)_{n\in\mathbb{N}}$ which belongs to the acceptance region, τ^A, is a stopping time with respect to the canonical filtration $(\sigma(V_0, \ldots, V_n))_{n\in\mathbb{N}}$ for every Borel-set A, because

$$\{\tau^A = n\} = \{V_0 \notin A, \ V_1 \notin A, \ldots, \ V_{n-1} \notin A, \ V_n \in A\} \in \sigma(V_0, \ldots, V_n) \ .$$

Here, $\sigma(V_0, \ldots, V_n)$ denotes the smallest σ-Borel field generated by the random variables $V_0, \ldots, V_n$. The accepted random vector has a random index and can be considered as the stopping element

$$V_{\tau^A} = (V_{\tau^A,1}, V_{\tau^A,2}) := \sum_{n=0}^{\infty} 1_{\{\tau^A = n\}} V_n \ . \tag{10}$$

Remark 2. The stopping time τ^A is a.s. finite iff $P(V_0 \in A) > 0$, because

$$\begin{aligned} P\left(\tau^A = \infty\right) &= P\left(V_0 \notin A, \ V_1 \notin A, \ldots\right) = P\left(\bigcap_{n\in\mathbb{N}} \{(V_0 \notin A) \cap \ldots \cap (V_n \notin A)\}\right) \\ &= \lim_{n\to\infty} P\left(\{(V_0 \notin A) \cap \ldots \cap (V_n \notin A)\}\right) = \lim_{n\to\infty} \left(1 - P(V_0 \in A)\right)^{n+1} \ . \end{aligned}$$

Example 1. Choosing the region of acceptance $A = K_p$, as it was done in step 2 of Algorithm 2, one gets an a.s. finite stopping time τ^A. To see this, let us consider the joint density of the power-function distributed random variables $V_{0,1}^l$ and $V_{0,2}^m$

$$h(x,y) = \frac{1}{ml} \ x^{1/l-1} y^{1/m-1}, \ (x,y) \in (0,1)^2, \ l, m > 0 \, .$$

Following (Jöhnk, 1964), it holds

$$\begin{aligned} P\left(V_{0,1}^l + V_{0,2}^m \leq 1\right) &= \int_0^1 \int_0^{1-x} \frac{1}{ml} \ x^{1/l-1} y^{1/m-1} \ dy \ dx = \frac{1}{ml} \int_0^1 x^{1/l-1} \left[\frac{y^{1/m}}{1/m}\right]_0^{1-x} \ dx \\ &= \frac{1}{l} \int_0^1 x^{1/l-1} (1-x)^{1/m} \ dx = \frac{\Gamma\left(\frac{1}{l}+1\right)\Gamma\left(\frac{1}{m}+1\right)}{\Gamma\left(\frac{1}{l}+\frac{1}{m}+1\right)} \ , \end{aligned}$$

so that

$$P\left(V_{0,1}^p + V_{0,2}^p \leq 1\right) = \frac{\Gamma\left(\frac{1}{p}+1\right)^2}{\Gamma\left(\frac{2}{p}+1\right)} \ .$$

Thus, step 2 of Algorithm 2 stopps almost surely at a finite time.

The introduction of a stopping time τ^A and a stopping element V_{τ^A} enables one to reduce various algorithms to stochastic representations of the underlying random vectors. Such representations can be used to prove the correctness of several generation methods. We show in the next lemma that the distribution of the stopping element V_{τ^A} coincides with the conditional distribution of the random vector V_i under the acceptance condition that $V_i \in A$, $\forall i \in \mathbb{N}$.

Lemma 1. If $A, B \in \mathfrak{B}(\mathbb{R}^2)$ and $P(V_0 \in A) > 0$, then

$$P(V_{\tau^A} \in B) = P(V_0 \in B |\ V_0 \in A) \quad .$$

Proof. Assuming $B \in \mathfrak{B}(\mathbb{R}^2)$, it holds

$$\begin{aligned} P(V_{\tau^A} \in B) &= P\left(\sum_{n=0}^{\infty} 1_{\{\tau^A = n\}} V_n \in B\right) = \sum_{n=0}^{\infty} P\left(\tau^A = n,\ V_n \in B\right) \\ &= P(V_0 \in A \cap B) + P(V_0 \in A \cap B) \sum_{n=1}^{\infty} P(V_0 \notin A)^n \\ &= P(V_0 \in B |\ V_0 \in A) \quad . \end{aligned}$$

□

Example 2. The p-generalized rejecting polar method is according to Theorem 3 based upon a stochastic representation for the random vector $\mathcal{U}_p$

$$\mathcal{U}_p \stackrel{d}{=} \Sigma \ \frac{V_{\tau^A}}{|V_{\tau^A}|_p} \quad , \tag{11}$$

where the acceptance region in the definition of the stopping element is $A = K_p$, the stopping time τ^A is a.s. finite, see Example 1, and the random signature matrix $\Sigma = diag(S_1, S_2)$ is based upon the random variables S_1, S_2 being independent and uniformly distributed on $\{-1, 1\}$. Here, equation (11) is a consequence of Lemma 1 and

$$P\left(\Sigma\, V_0/|V_0|_p \in B \,\middle|\, |V_0|_p \leq 1\right) = P(\mathcal{U}_p \in B) \quad , \forall B \in \mathfrak{B}(C_p) \ ,$$

which was shown in the proof of Theorem 3.

Example 3. Let (Y_1, Y_2) be $l_{2,p}$-generalized uniformly distributed on $C_p \cap \mathbb{R}^2_+$. Then, according to the proof of Theorem 3,

$$P\left(V_0/|V_0|_p \in B_+ \,\middle|\, |V_0|_p \leq 1\right) = P((Y_1, Y_2) \in B_+) \quad , \forall B_+ \in C_p \cap \mathbb{R}^2_+,$$

if $(V_n)_{n\in\mathbb{N}}$ is a sequence of uniformly on $(0,1)^{\times 2}$ distributed random vectors which is independent from (Y_1, Y_2). Hence

$$P\left(\frac{V_{0,1}}{V_{0,2}} \in B \,\middle|\, |V_0|_p \leq 1\right) = P\left(\frac{Y_1}{Y_2} \in B\right) \quad , \forall B \in \mathfrak{B}(\mathbb{R}_+),$$

and consequently

$$\frac{V_{\tau^A,1}}{V_{\tau^A,2}} \stackrel{d}{=} \frac{Y_1}{Y_2}\,,$$

if $A = K_p$. For this reason, the stochastic representation from Theorem 1 for the random angle Φ is equivalent to

$$\Phi \stackrel{d}{=} S_1 \arctan\left(\frac{V_{\tau^A,1}}{V_{\tau^A,2}}\right) + \pi S_2 + \frac{\pi}{2} \quad , \tag{12}$$

that means the simulation of Φ can be reduced to the simulation of uniformly distributed random variables, see step 2 of Algorithm 7.

Example 4. Let Y be Beta distributed with positive parameters l and m, i.e. $Y \sim B(l,m)$. Then, according to (Jöhnk, 1964),

$$P\left(Y \in B\right) = P\left(\frac{V_{0,1}^{1/l}}{V_{0,1}^{1/l} + V_{0,1}^{1/m}} \in B \,\middle|\, V_{0,1}^{1/l} + V_{0,1}^{1/m}\right) \quad , \forall B \in \mathfrak{B}(\mathbb{R}) \cap (0,1]\,.$$

With Lemma 1, we obtain the stochastic representation

$$Y \stackrel{d}{=} \frac{V_{\tau^A,1}^{1/l}}{V_{\tau^A,1}^{1/l} + V_{\tau^A,2}^{1/m}}\,, \tag{13}$$

where the acceptance region A is defined as $A = \left\{(x,y) \in (0,1)^{\times 2} : x^{1/l} + y^{1/m} \leq 1\right\}$.

B The completed algorithms of the p-generalized polar methods

The p-generalized polar method according to Algorithm 1 includes in step 1 the generation of a Gamma distributed random number and in step 2 the simulation of the angular distribution according to Definition 1. We now present completed versions of Algorithm 1 and Algorithm 2, which are exclusively based on the generation of uniformly distributed random numbers. To this end, let k be the largest natural number less than or equal to $2/p$, i.e. $k := [2/p]$, and $\tilde{p} = 2/p - k$.

Algorithm 7 (Completed p-generalized polar method)**.**

1. Simulation of $\mathcal{R}_p$
   - (i) Choose $z_1, \ldots, z_k$ uniformly from $(0,1)$. Set $\gamma = \ln z_1 + \ldots + \ln z_k$ and go on with (iv) , if $\tilde{p} = 0$.
   - (ii) Choose (u_1, u_2) uniformly from $(0,1)^{\times 2}$ until $u_1^{1/\tilde{p}} + u_2^{1/(1-\tilde{p})} \leq 1$.
   - (iii) Choose z_{k+1} uniformly from $(0,1)$ and set $\gamma = \gamma + \ln z_{k+1} \frac{u_1^{1/\tilde{p}}}{u_1^{1/\tilde{p}} + u_2^{1/(1-\tilde{p})}}$.
   - (iv) Form $r_p = (-p\,\gamma)^{1/p}$.
2. Simulation of $\mathcal{U}_p$
   - (i) Choose (v_1, v_2) uniformly from $(0,1)^{\times 2}$ until $|(v_1, v_2)|_p \leq 1$.
   - (ii) Choose σ_1 uniformly from $\{-1, 1\}$.
   - (iii) Choose σ_2 uniformly from $\{0, 1\}$.
   - (iv) Build $\phi = \sigma_1 \arctan\left(\frac{v_1}{v_2}\right) + \pi\sigma_2 + \frac{\pi}{2}$ and $u_p = (\cos_p \phi, \sin_p \phi)$.
3. Return $(x_1, x_2) = r_p \cdot u_p$.

Algorithm 8 (Completed p-generalized rejecting polar method)**.**

1. Follow step 1 from Algorithm 7 to simulate $\mathcal{R}_p$.
2. Simulation of $\mathcal{U}_p$
   (a) Choose (v_1, v_2) uniformly from $(0,1)^{\times 2}$ until $v_1^p + v_2^p \leq 1$.
   (b) Choose (s_1, s_2) uniformly from $\{-1,1\}^{\times 2}$ and form the signature matrix $\Sigma = diag(s_1, s_2)$.
   (c) Build $u_p = \frac{(v_1,v_2)}{|(v_1,v_2)|_p}\,\Sigma$.
3. Return $(x_1, x_2) = r_p \cdot u_p$.

Remark 3. Let X_1, X_2 be an independent pair of p-generalized Gaussian distributed random variables and let $Z_1, \ldots, Z_{[2/p]+1}$ be independent random variables following the uniform distribution on $(0,1)$. The random variable $\mathcal{R}_p^p/p = |(X_1, X_2)|_p^p/p$ is according to Theorem 2 Gamma distributed with parameter $2/p$. If $2/p$ is a natural number, then

$$\frac{\mathcal{R}_p^p}{p} \stackrel{d}{=} -\sum_{i=1}^{2/p} \ln Z_i \quad ,$$

i.e. $\mathcal{R}_p^p/p$ can be represented as the sum of $2/p$ independent exponential distributed random variables, see (Jöhnk, 1964). If $\tilde{p} = 2/p - [2/p] > 0$, then

$$-\frac{\mathcal{R}_p^p}{p} \stackrel{d}{=} \sum_{i=1}^{[2/p]} \ln Z_i + \ln Z_{[2/p]+1}\, Y \,,$$

where Y is beta distributed with parameters $\tilde{p}$ and $1-\tilde{p}$. Consequently,

$$-\frac{\mathcal{R}_p^p}{p} \stackrel{d}{=} \sum_{i=1}^{[2/p]} \ln Z_i + \ln Z_{[2/p]+1}\, \frac{V_{\tau^A,1}^{1/\tilde{p}}}{V_{\tau^A,1}^{1/\tilde{p}} + V_{\tau^A,2}^{1/(1-\tilde{p})}} \,,$$

if $\tilde{p} > 0$, $(V_n)_{n\in\mathbb{N}}$ is a sequence of independent random vectors which follow the uniform distribution on $(0,1)^{\times 2}$ and $A = \left\{(x,y) \in (0,1)^{\times 2} : x^{1/\tilde{p}} + y^{1/(1-\tilde{p})} \leq 1\right\}$, compare with Example 4.

Remark 4. In step 2 of Algorithm 7, the angular distribution corresponding to the $l_{2,p}$-generalized uniform distribution on C_p is simulated with a rejection method corresponding to the stochastic representation (12) from Example 3.

C Auxiliary lemmas

Lemma 2. If $\tilde{f}(x) = \exp\left(-\frac{x^p}{p}\right)$ and $g(x) = \exp\left(-\xi^{p-1}x\right)$, then

$$\tilde{f}(\xi + x) \leq \tilde{f}(\xi) g(x) \,,\ \forall x \geq 0,\ p \geq 1.$$

Proof. We notice that $\exp\left(-\frac{(\xi+x)^p}{p}\right) \leq \exp\left(-\frac{\xi^p}{p} - \xi^{p-1}x\right)$ iff $(\xi+x)^p \geq \xi^p + p\xi^{p-1}x$. A Taylor expansion of the continuously differentiable function $h(x) = x^p$, $x \in \mathbb{R}_+$, yields

$$h(\xi + x) = \xi^p + p\xi^{p-1}x + R_1(\xi + x) \,,$$

where

$$R_1(\xi + x) = p\,(p-1) \int_{\xi}^{\xi+x} (\xi + x - t)\, t^{p-2}\, dt \geq 0 \,.$$

□

Lemma 3. Let V be a continuous random variable with density $f : \mathbb{R} \to \mathbb{R}_+$ and let $X = (X_1, X_2)$ be uniformly distributed on $A(f) := \{(x, y) \in \mathbb{R}^2 : \ 0 < y < f(x)\}$. Then

$$X_1 \stackrel{d}{=} V \ .$$

Proof. X has the density $g_X(x, y) = 1_{A(f)}(x, y)$ and the marginal density of X_1 is

$$\begin{aligned} g_{X_1}(x_1) &= \int\limits_{\mathbb{R}} 1_{A(f)}(x_1, x_2) \ dx_2 = \int\limits_0^{f(x_1)} dx_2 \\ &= f(x_1) \quad . \end{aligned}$$

□

Lemma 4. Let X be uniformly distributed on $A \in \mathfrak{B}(\mathbb{R}^2)$ with $\mu(A) = 1/c > 0$. Let also $T : A \to B$, $B \in \mathfrak{B}(\mathbb{R}^2)$, be a bijective map satisfying $\mu(\tilde{A}) = \mu(T(\tilde{A}))$, $\forall \tilde{A} \in \mathfrak{B}(A)$. Then $T(X)$ follows a uniform distribution on B and has the density

$$f_{T(X)}(x) = c \ 1_B(x) \quad .$$

Proof. Let $\tilde{B} \in \mathfrak{B}(\mathbb{R}^2) \cap B$, then

$$\begin{aligned} P\left(T(X) \in \tilde{B}\right) = P\left(X \in T^{-1}(\tilde{B})\right) &= \int\limits_{T^{-1}(\tilde{B})} c \ 1_A(x, y) \, \mu(d(x, y)) \\ &= c \int\limits_{T^{-1}(\tilde{B})} \left(1_A \circ T^{-1} \circ T\right)(x, y) \, \mu(d(x, y)) = c \int\limits_{\tilde{B}} \left(1_A \circ T^{-1}\right)(x, y) \ d(T\,\mu) \\ &= \int\limits_{\tilde{B}} c \ 1_B(x, y) \, \mu(d(x, y)) \end{aligned}$$

□

Lemma 5. Let Y be uniformly distributed on $A \subseteq B \in \mathfrak{B}(\mathbb{R}^2)$ and let X be uniformly distributed on B, where $\mu(A) > 0$. Then

$$P\left(X \in C |\ X \in A\right) = P\left(Y \in C\right) \ , \ \forall C \in \mathfrak{B}(\mathbb{R}^2) \ .$$

Proof.

$$\begin{aligned} P\left(X \in C |\ X \in A\right) &= \frac{\int\limits_{A \cap C} \frac{1}{\mu(B)} 1_B(x) \ dx}{(\mu(A)/\mu(B))} = \int\limits_{A \cap C} \frac{1}{\mu(A)} 1_B(x) \ dx \\ &= \int\limits_C \frac{1}{\mu(A)} 1_A(x) \ dx = P\left(Y \in C\right) \ . \end{aligned}$$

□

Lemma 6. Let $A_1, \ldots, A_m$ be pairwise disjoint with $A_i \in \mathfrak{B}(\mathbb{R}^2)$ and $\mu(A_i) = v > 0$, $\forall i \in \{1, \ldots, m\}$. Let further X_i be uniformly distributed on A_i, $i \in \{1, \ldots, m\}$, and assume that Z is uniformly distributed on $\{1, \ldots, m\}$ and independent from $X_1, \ldots, X_m$. Then $\sum_{i=1}^m 1_{\{Z=i\}} X_i$ follows a uniform distribution on $A_1 \cup \ldots \cup A_m$.

Proof.

$$P\left(\sum_{i=1}^{m} 1_{\{Z=i\}} X_i \in A\right) = P\left(\bigcup_{i=1}^{m} \{Z=i,\ X_i \in A\}\right) = \sum_{i=1}^{m} P\left(Z=i,\ X_i \in A\right)$$
$$= \sum_{i=1}^{m} P\left(Z=i\right) P\left(X_i \in A\right) = \frac{1}{m} \sum_{i=1}^{m} \frac{\mu(A \cap A_i)}{\mu(A_i)}$$
$$= \frac{1}{m\ v} \sum_{i=1}^{m} \mu\left(A \cap A_i\right) = \frac{\mu\left(A \cap (A_1 \cup \ldots \cup A_m)\right)}{\mu\left(A_1 \cup \ldots \cup A_m\right)}$$

□

D The stochastic representation behind the adapted tail algorithm

In this supplementary section, we reflect the basic mathematical relations behind the tail algorithm, see (Marsaglia and Bray, 1964), and its adaption to the p-generalized Gaussian distribution. Consider a positive random variable X with a strictly monotonic decreasing density f and a random vector $Z = (Z_1, Z_2)$, which is uniformly distributed on the tail area $A(f,\xi) := \{(x,y) \in \mathbb{R}_+^2 : x > \xi,\ 0 < y < f(x)\}$ for $\xi > 0$. It follows from Lemma 3, that

$$P\left(Z_1 < t\right) = P\left(X < t \,\middle|\, X \geq \xi\right),\ \forall t \geq \xi\,,$$

i.e. Z_1 follows the conditional distribution of X under the condition that $X \geq \xi$. To simulate Z_1, the tail agorithm generates a random tuple (z_1, z_2) that is uniformly distributed on a superset $\tilde{A}(f,\xi)$ of $A(f,\xi)$, see Figure 8. According to Lemma 5, z_1 is a realisation of Z_1 if $(z_1, z_2) \in A(f,\xi)$. In this context, the set $\tilde{A}(f,\xi)$ is constructed with the help of a function g of the type $g(t) = e^{-\beta t}$ or $g(t) = (1+\psi t)^{-\beta}$, such that

$$f(\xi + t) \leq f(\xi) g(t)\ ,\ \forall t \geq 0\,. \tag{14}$$

In case of a p-generalized Gaussian distributed random variable X and the density f_p^* of it's absolute value, the function $g(x) = \exp\left(-\xi^{p-1} x\right)$ satisfies the condition in (14) for every $p \geq 1$. Otherwise constants $\beta > 1$ and $\psi > 0$ have to be found for every p from the range $(0,1)$, such that $g(t) = (1+\psi t)^{-\beta}$ fulfilles the condition in (14) and $\int_0^\infty g(t)\ dt$ is as small as possible. The last requirement is due to the fact that the rejection probability of (z_1, z_2) attains its minimal value if $\int_0^\infty g(t)\ dt$ is as small as possible.

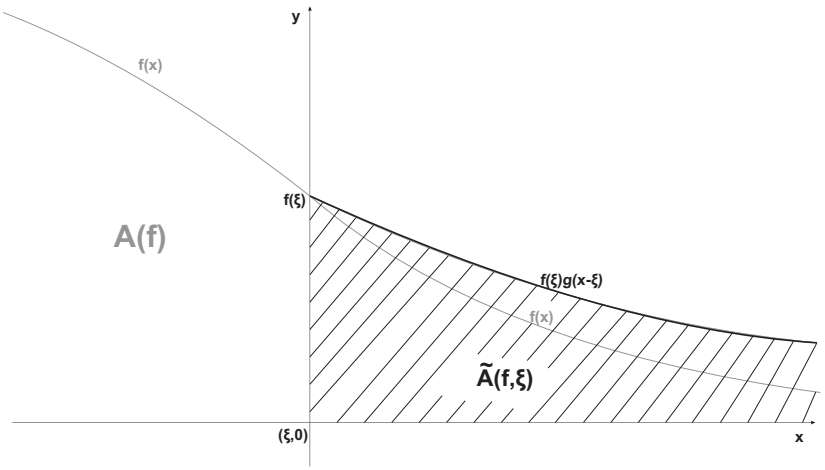

Fig. 8: Simulation of a uniform distribution on $\tilde{A}(f,\xi)$.

The uniform distribution on the set $\tilde{A}(f,\xi)$ is finally simulated with the help of a suitable one-to-one map $T:(0,1)^2 \to \tilde{A}(f,\xi)$. In this context, a uniformly from $(0,1)^{\times 2}$ chosen random tuple (v_1,v_2) gets mapped with T onto $\tilde{A}(f,\xi)$, so that the tuple $T(v_1,v_2)$ is a realization of the random vector Z.

Remark 5. In case of the p-generalized Gaussian distribution, the map $T=(T_1,T_2)$ may be chosen as

$$T(v_1,v_2)=\begin{cases}\left(\xi-\frac{\ln v_1}{\xi^{p-1}}, v_2 f_p^*(\xi) g\left(-\frac{\ln v_1}{\xi^{p-1}}\right)\right) & \text{, if } p\geq 1\\ \left(\xi+\frac{1}{\psi}\left(v_1^{1/(1-\beta)}-1\right), v_2 f_p^*(\xi) g\left(\frac{1}{\psi}\left(v_1^{1/(1-\beta)}-1\right)\right)\right) & \text{, if } 0<p<1\end{cases}.$$

Algorithm 4 corresponds to the stochastic representations given in the next two theorems. Here, it is shown that $T_1\left(V_{\tau^B}\right)$ has the needed conditional distribution, if T is chosen as in Remark 5 and B is chosen as in Theorems 4 and 5, respectively. Notice, that the Lebesgue measure is in this case not invariant up to T. Consequently, Lemma 4 cannot be used to show the correctness of Algorithm 4.

Theorem 4. Assume $X\sim\mathcal{N}_p$, $p\geq 1$, $\xi>0$ and let the region of acceptance in the definition of the stopping time be $B:=\left\{(x,y)\in\mathbb{R}^2 : y<x^{-1}\exp\left[\frac{\xi^p}{p}-\frac{1}{p}\left(\xi-\frac{\ln x}{\xi^{p-1}}\right)^p\right]\right\}$. Then

$$P\left(\xi-\frac{\ln V_{\tau^B,1}}{\xi^{p-1}}<z\right)=P\left(|X|<z \;\middle|\; |X|>\xi\right)\ ,\ \forall z>\xi\ ,$$

where τ^B is an a.s. finite stopping time.

Proof. It follows from Lemma 2 that $B\subseteq(0,1)^2$, because

$$\exp\left[-\frac{1}{p}\left(\xi-\frac{\ln x}{\xi^{p-1}}\right)^p\right]=\tilde{f}\left(\xi-\frac{\ln x}{\xi^{p-1}}\right)\leq\tilde{f}(\xi)g\left(-\frac{\ln x}{\xi^{p-1}}\right)=\exp\left[-\frac{\xi^p}{p}\right]x,\ \forall x\in(0,1)\ .$$

For this reason,

$$\begin{aligned}P\left(V_{0,1}<t \;\middle|\; V_0\in B\right)&=\tilde{c}\int\limits_{x\in(0,t)}\ \int\limits_{y=0}^{x^{-1}\exp[\xi^p/p-1/p(\xi-\ln x/\xi^{p-1})^p]} dy\ dx\\ &=\tilde{c}\int\limits_{x\in(0,t)} x^{-1}\exp\left[\frac{\xi^p}{p}-\frac{1}{p}\left(\xi-\frac{\ln x}{\xi^{p-1}}\right)^p\right]\ dx\end{aligned}$$

and

$$\frac{d}{dt}P\left(V_{0,1}<t \;\middle|\; V_0\in B\right)=1_{(0,1)}(t)\ \tilde{c}\ t^{-1}\exp\left[\frac{\xi^p}{p}-\frac{1}{p}\left(\xi-\frac{\ln t}{\xi^{p-1}}\right)^p\right]\ .$$

Thus we obtain

$$\begin{aligned}\frac{d}{dz}\left[P\left(\xi-\frac{\ln V_{0,1}}{\xi^{p-1}}<z \;\middle|\; V_0\in B\right)\right]&=\frac{d}{dz}\left[P\left(V_{0,1}>\exp(\xi^p-\xi^{p-1}z) \;\middle|\; V_0\in B\right)\right]\\ &=\frac{d}{dz}\left[\int\limits_{\exp(\xi^p-\xi^{p-1}z)}^{1}\tilde{c}\ t^{-1}\exp\left[\frac{\xi^p}{p}-\frac{1}{p}\left(\xi-\frac{\ln t}{\xi^{p-1}}\right)^p\right]dt\right]\\ &=-\tilde{c}\ \exp(\xi^{p-1}z-\xi^p)\exp\left[\frac{\xi^p}{p}-\frac{1}{p}(\xi-(\xi-z))^p\right]\exp\left[\xi^p-\xi^{p-1}z\right](-\xi^{p-1})\\ &=\tilde{c}\ \xi^{1-p}\exp\left[\frac{\xi^p}{p}\right]\ \exp\left[-\frac{z^p}{p}\right]\\ &=c\ \exp\left[-\frac{z^p}{p}\right]\ ,\ z>\xi\ .\end{aligned}$$

The final conclusion follows from Lemma 1 and the fact, that

$$P\left(V_0 \in B\right) = \tilde{c} \int\limits_{x\in(0,1)} \underbrace{x^{-1} \exp\left[\frac{\xi^p}{p} - \frac{1}{p}\left(\xi - \frac{\ln x}{\xi^{p-1}}\right)^p\right]}_{>0} \, dx > 0 \ .$$

□

The following Theorem 5 proves the correctness of Algorithm 4 in the case $p \in (0,1)$. We assume therefore that the defining constants $\beta > 1$ and $\psi > 0$ of the function $g(t) = (1+\psi t)^{-\beta}$ are chosen according to condition (14).

Theorem 5. Assume $X \sim \mathcal{N}_p$, $p \in (0,1)$, $\xi > 0$ and let

$$B := \left\{(x,y) \in \mathbb{R}^2 : y e^{-\frac{\xi^p}{p}} x^{\frac{\beta}{\beta-1}} < exp\left[-\frac{1}{p}\left(\xi + \frac{1}{\psi}\left(u_1^{\frac{1}{(1-\beta)}} - 1\right)\right)\right]\right\}$$

denote the acceptance region in the definition of the stopping time τ^B of $(V_n)_{n\in\mathbb{N}}$. Then

$$P\left(\xi + \frac{1}{\psi}\left(V_{\tau^B,1}^{\frac{1}{(1-\beta)}} - 1\right) < z\right) = P\left(|X| < z \;\middle|\; |X| > \xi\right) \ , \ \forall z > \xi \ ,$$

where τ^B is a.s. finite.

Proof. We assumed that (14) is fulfilled, so $B \subset (0,1)^2$. It holds

$$\begin{aligned} P\left(V_{0,1} < t \mid V_0 \in B\right) &= \int\limits_{x\in(0,t)} \int\limits_{0}^{x^{\beta/(1-\beta)} \exp\left[\frac{\xi^p}{p} - \frac{1}{p}\left(\xi + \frac{1}{\psi}\left(x^{1/(1-\beta)} - 1\right)\right)^p\right]} dy \; dx \\ &= \int\limits_0^t x^{\beta/(1-\beta)} \exp\left[\frac{\xi^p}{p} - \frac{1}{p}\left(\xi + \frac{1}{\psi}\left(x^{1/(1-\beta)} - 1\right)\right)^p\right] \, dx \end{aligned}$$

and

$$\frac{d}{dt} P\left(V_{0,1} < t \mid V_0 \in B\right) = 1_{(0,1)}(t) e^{\frac{\xi^p}{p}} t^{\frac{\beta}{1-\beta}} \exp\left[-\frac{1}{p}\left(\xi + \frac{1}{\psi}\left(t^{1/(1-\beta)} - 1\right)\right)^p\right] \quad .$$

Hence

$$\begin{aligned} \frac{d}{dz}\left[P\left(\xi + \frac{1}{\psi}\left(V_{0,1}^{1/(1-\beta)} - 1\right) < z \mid V_0 \in B\right)\right] &= \frac{d}{dz}\left[P\left(V_{0,1} > (\psi(z-\xi)+1)^{1-\beta} \mid V_0 \in B\right)\right] \\ &= \frac{d}{dz}\left[\int\limits_{(\psi(z-\xi)+1)^{1-\beta}}^{1} e^{\frac{\xi^p}{p}} t^{\frac{\beta}{1-\beta}} e^{-\frac{1}{p}\left(\xi + \frac{1}{\psi}\left(t^{1/(1-\beta)} - 1\right)\right)^p} \, dt\right] \\ &= (\beta - 1)\, e^{\frac{\xi^p}{p}} \psi \; e^{-\frac{z^p}{p}} \\ &= c \; e^{-\frac{z^p}{p}} \quad . \end{aligned}$$

□

E The stochastic representation behind the adapted Monty Python method

We already described the main idea behind the Monty Python method and the role of a suitable map $T: R(b) \to A(f)$ in subsection 4.1, see especially Figure 6 for a definition of the sets D, E, F, G, H and $\tilde{H}$. With respect to Lemma 4, T has to be a bijective map and its images $T(A)$ should have the same Lebesgue-measure $\mu(T(A))$ as the origin sets $A \in \mathfrak{B}(R(b))$.

Definition 2. Let f be the strictly monotonic decreasing density of a positive random variable X and let μ denote the Lebesgue measure on $\mathfrak{B}(\mathbb{R}^2)$. The map $T: R(b) \to A(f)$ is defined by

$$T(x,y) = (T_1(x,y), T_2(x,y)) = \begin{cases} (x,y) & \text{, if } (x,y) \in F \dot{\cup} G \\ \left(s(b-x), \frac{1}{b} + \frac{1}{bs} - \frac{y}{s}\right) & \text{, if } (x,y) \in H \\ S(x,y) & \text{, otherwise} \end{cases},$$

where $a := f^{-1}(1/b)$, $s := a/(b-a)$ and $S: D \to E$ is bijective and satisfies the condition $\mu(\tilde{A}) = \mu(S(\tilde{A}))$, $\forall \tilde{A} \in \mathfrak{B}(D)$.

The map T corresponds to the projection of uniformly on $R(b)$ distributed random vectors in Algorithm 5 and fulfilles the conditions of Lemma 4. Consequently, the random vector $T(b\,V_{0,1}, V_{0,2}/b)$ is uniformly distributed on $A(f)$.

Theorem 6. Let (Z_1, Z_2) be uniformly distributed on E and independent from $V_0 = (V_{0,1}, V_{0,2})$. Then

$$X \overset{d}{=} 1_{R(b)\setminus D}(b \cdot V_{0,1}, 1/b \cdot V_{0,2})\; T_1(b \cdot V_{0,1}, 1/b \cdot V_{0,2}) + 1_D(b \cdot V_{0,1}, 1/b \cdot V_{0,2})\; Z_1.$$

Proof. The random vector $Y = (b \cdot V_{0,1}, 1/b \cdot V_{0,2})$ follows a uniform distribution on $R(b)$. It follows from Lemma 4, that $T(Y)$ is uniformly distributed on $A(f)$. According to Lemma 3, we obtain $T_1(Y) \overset{d}{=} X$ and

$$\begin{aligned} P\left(Y \in A | Y \in D\right) &= \frac{\int\limits_{A \cap D} \mu(d(x,y))}{\mu(D)} \\ &= \frac{1}{\mu(D)} \int\limits_A 1_D(x,y)\, \mu(d(x,y)) \\ &= P(\tilde{Z} \in A) \quad , \; A \in \mathfrak{B}(\mathbb{R}^2), \end{aligned}$$

where $\tilde{Z}$ is uniformly distributed on D. Hence

$$1_D(Y) T(Y) \overset{d}{=} 1_D(Y) T(\tilde{Z}),$$

where $T(\tilde{Z})$ follows according to Lemma 4 a uniform distribution on E. □

Let us remark that a problem comes up when choosing the width b. As can be seen in Definition 2, bijectivity is one of the requirements for T. If b is too large, the images of the disjoint sets $\tilde{H}$ and G under T^{-1} aren't disjoint, see Figure 9, where points from G are mapped to themselves. Therefore, we have to assume that b is chosen according to the condition

$$T_2^{-1}(x, f(x)) > f\left(T_1^{-1}(x, f(x))\right) \;, \forall x \in (0, a), \tag{15}$$

where $a = f^{-1}(1/b)$.

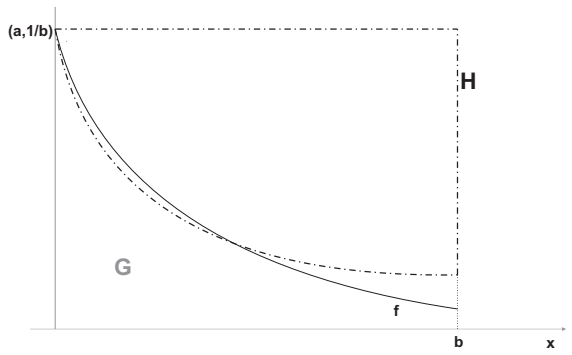

Fig. 9: Monthy Python: no bijectivity of T if b is too large.

Remark 6. The choice of the width b according to condition (15) requires for every p the solution of an optimization problem. From the condition $1/b < f(0)$, we obtain a lower bound $b \geq p^{1/p-1}\Gamma\left(\frac{1}{p}\right)$. The upper bound follows from the condition in (15), which is in case of the p-generalized Gaussian distribution equivalent to

$$1 - \frac{p^{1-1/p}}{\Gamma\left(\frac{1}{p}\right)}\, a \exp\left[-\frac{1}{p}x^p\right] > \frac{p^{1-1/p}}{\Gamma\left(\frac{1}{p}\right)}\,(b-a)\exp\left[-\frac{1}{p}\left(b-\frac{x}{s}\right)^p\right] \quad , \forall x \in (0,a)\ . \tag{16}$$

The optimization problem rises from the interest in the maximum value of the admissible range of b, to use the expensive tail algorithm as rarely as possible.

F The stochastic representation behind the adapted Ziggurat method

The main idea behind the Ziggurat method was already given in Subsection 4.1. For the sake of completeness, we prove here the stochastic representation corresponding to Algorithm 6 and reflect a method from (Marsaglia and Tsang, 2000) for the set up of the Ziggurat. Let us firstly assume that the Ziggurat for the strictly monotonic density f of a positive random variable X is already defined by the coordinates x_1, x_2, ..., x_{m-1}, see Figure 7.
Additionally to the sequence $(V_n)_{n\in\mathbb{N}}$ of uniformly on $(0,1)^{\times 2}$ distributed random vectors defined in subsection A, we consider here a sequence $(Y_n)_{n\in\mathbb{N}}$ of uniformly on $\{1,\ldots,m\}$ distributed random variables. A realisation y_n of Y_n represents the uniformly chosen area in Algorithm 6. The sequence $(Y_n)_{n\in\mathbb{N}}$ gets stopped if the uniformly chosen area is $R_m \cup E$, see Figure 7, or if the uniformly chosen tuple from R_{y_n} is in $A(f)$ in the case that $y_n \in \{1,\ldots,m-1\}$. For this reason, the stopping time τ of $(Y_n)_{n\in\mathbb{N}}$ stopps also the sequence $(V_n)_{n\in\mathbb{N}}$, because a uniformly chosen tuple from the rectangle R_i is represented by $S_i(v_n)$, where

$$S_i(x,y) := \begin{cases} \left(x_i x, \frac{v}{x_i} y\right) & , \text{if } i \in \{1,\ldots,m-1\} \\ \left(\frac{v}{f(x_{m-1})} x, f(x_{m-1}) y\right) & , \text{if } i = m \end{cases} .$$

The following theorem presents a stochastic representation of X that corresponds to Algorithm 6 in the case that $X \stackrel{d}{=} |Y|$ and $Y \sim \mathcal{N}_p$.

Theorem 7. Assume $(Y_n)_{n\in\mathbb{N}}$ is a sequence of independent random variables, uniformly distributed on $\{1,\ldots,m\}$, which is independent from $(V_n)_{n\in\mathbb{N}}$. Let also $Z = (Z_1, Z_2)$ be uniformly distributed on E and τ be a stopping time satisfying

$$\tau = \inf\left\{ n \in \mathbb{N} :\ Y_n = m \vee \sum_{i=1}^{m-1} 1_{\{Y_n = i\}} S_i(V_n) \in A(f) \right\} \quad ,$$

where

$$S_i(x,y) := \begin{cases} \left(x_i x, \frac{v}{x_i} y\right) & \text{, if } i \in \{1,\dots,m-1\} \\ \left(\frac{v}{f(x_{m-1})} x, f(x_{m-1}) y\right) & \text{, if } i = m \end{cases} .$$

Then

$$X \overset{d}{=} \sum_{i=1}^{m-1} 1_{\{Y_\tau = i\}} x_i V_{\tau,1} + 1_{\{Y_\tau = m\}} \left[1_{R_m}(S_m(V_\tau)) \frac{v}{f(x_{m-1})} V_{\tau,1} + 1_{R_m^* \setminus R_m}(S_m(V_\tau)) Z_1 \right] .$$

Proof. Assuming $A \in \mathfrak{B}(\mathbb{R}^2)$, it holds

$$\begin{aligned} P\left(\sum_{i=1}^{m-1} 1_{\{Y_\tau=i\}} S_i(V_\tau) \in A\right) &= P\left(\bigcup_{i=1}^{m-1} \{Y_\tau = i,\ S_i(V_\tau) \in A\}\right) = \sum_{i=1}^{m-1} P\left(Y_\tau = i,\ S_i(V_\tau) \in A\right) \\ &= \sum_{i=1}^{m-1} P\left(\bigcup_{j=0}^{\infty} \{\tau = j,\ Y_\tau = i,\ S_i(V_\tau) \in A\}\right) \\ &= \sum_{i=1}^{m-1} \sum_{j=0}^{\infty} P\left(Y_j = i,\ S_i(V_j) \in A \cap A(f)\right) P\left(\tau > j-1\right) \\ &= \sum_{i=1}^{m-1} P\left(Y_0 = i,\ S_i(V_0) \in A \cap A(f)\right) \sum_{j=0}^{\infty} P\left(\tau > j-1\right) = (*) , \end{aligned}$$

where

$$\begin{aligned} \sum_{j=0}^{\infty} P\left(\tau > j-1\right) &= \sum_{j=0}^{\infty} P\left(Y_0 \neq m,\ \sum_{i=1}^{m-1} 1_{\{Y_0=i\}} S_i(V_0) \notin A(f)\right)^j \\ &= \left(1 - P\left(Y_0 \neq m,\ \sum_{i=1}^{m-1} 1_{\{Y_0=i\}} S_i(V_0) \notin A(f)\right)\right)^{-1} \\ &= \left(P\left(\{Y_0 = m\} \cup \left\{\sum_{i=1}^{m-1} 1_{\{Y_0=i\}} S_i(V_0) \in A(f)\right\}\right)\right)^{-1} . \end{aligned}$$

Hence

$$(*) = \frac{P\left(\sum_{i=1}^{m-1} 1_{\{Y_0=i\}} S_i(V_0) \in A \cap A(f)\right)}{P\left(\{Y_0 = m\} \cup \left\{\sum_{i=1}^{m-1} 1_{\{Y_0=i\}} S_i(V_0) \in A(f)\right\}\right)} . \tag{17}$$

If V^* is uniformly distributed on $(0,1)^2$, then $S_m(V^*)$ follows a uniform distribution on R_m^*. Consider the bijective, Lebesgue-measure invariant map $W : R_m^* \to R_m \cup E$ satisfying

$$(W_1(x,y), W_2(x,y)) = \begin{cases} (x,y) & \text{, if } (x,y) \in R_m \\ H(x,y) & \text{, else} \end{cases} ,$$

where $H : R_m^* \backslash R(m) \to E$ is bijective and invariant with respect to the Lebesgue-measure. In this case $W\left(S_m(V^*)\right)$ is uniformly distributed on $R_m \cup E$ and it holds with respect to Lemma 3

$$1_{R_m^* \setminus R_m}\left(S_m(V^*)\right) H_1\left(S_m(V^*)\right) \overset{d}{=} 1_{R_m^* \setminus R_m}\left(S_m(V^*)\right) Z_1 .$$

Hence

$$\begin{aligned}
&1_{\{Y_\tau=m\}}\left[1_{R_m}\left(S_m(V_\tau)\right)\frac{v}{f(x_{m-1})}V_{\tau,1}+1_{R_m^*\setminus R_m}\left(S_m(V_\tau)\right)Z_1\right]\\
\overset{d}{=}&1_{\{Y_\tau=m\}}\underbrace{\left[1_{R_m}\left(S_m(V_\tau)\right)\frac{v}{f(x_{m-1})}V_{\tau,1}+1_{R_m^*\setminus R_m}\left(S_m(V_\tau)\right)H_1\left(S_m(V_\tau)\right)\right]}_{=W_1(S_m(V_\tau))}
\end{aligned}$$

and

$$\begin{aligned}
P\left(1_{\{Y_\tau=m\}}W\left(S_m(V_\tau)\right)\in A\right)&=P\left(Y_\tau=m,\ W\left(S_m(V_\tau)\right)\in A\right)\\
&=P\left(\bigcup_{j=0}^{\infty}\{\tau=j,\ Y_\tau=m,\ W\left(S_m(V_\tau)\right)\in A\}\right)\\
&=P\left(Y_0=m,\ W\left(S_m(V_0)\right)\in A\right)\sum_{j=0}^{\infty}P\left(\tau>j-1\right)\\
&=\frac{P\left(Y_0=m,\ W\left(S_m(V_0)\right)\in A\cap A(f)\right)}{P\left(\{Y_0=m\}\cup\left\{\sum_{i=1}^{m-1}1_{\{Y_0=i\}}S_i(V_0)\in A(f)\right\}\right)}\quad .
\end{aligned}$$

With (17), we finally obtain

$$\begin{aligned}
&P\left(\sum_{i=1}^{m-1}1_{\{Y_\tau=i\}}S_i(V_\tau)+1_{\{Y_\tau=m\}}W\left(S_m(V_\tau)\right)\in A\right)\\
=&\frac{P\left(\sum_{i=1}^{m-1}1_{\{Y_0=i\}}S_i(V_0)+1_{\{Y_0=m\}}W\left(S_m(V_0)\right)\in A\cap A(f)\right)}{P\left(\underbrace{\sum_{i=1}^{m-1}1_{\{Y_0=i\}}S_i(V_0)+1_{\{Y_0=m\}}W\left(S_m(V_0)\right)}_{=:\tilde{X}}\in A(f)\right)}\quad .
\end{aligned}\tag{18}$$

Due to Lemma 6, $\tilde{X}$ follows a uniform distribution on $R_1\cup\ldots\cup R_m\cup E$ and (18) is equivalent to

$$\frac{P\left(\tilde{X}\in A\cap A(f)\right)}{P\left(\tilde{X}\in A(f)\right)}\quad .$$

We know from Lemma 5 that this is the uniform distribution on $A(f)$ applied to A. The first component of

$$\sum_{i=1}^{m-1}1_{\{Y_\tau=i\}}S_i(V_\tau)+1_{\{Y_\tau=m\}}W\left(S_m(V_\tau)\right)$$

is in consequence and with respect to Lemma 3 distributed as X. □

Let us now refer to the concept introduced in (Marsaglia and Tsang, 2000) for the set up of the Ziggurat, i.e. for the calculation of the coordinates $x_1,\ldots,x_{m-1}$.

Remark 7. As $R_1,\ldots,R_{m-1},R_m^*$ have all the same area content v, it holds

$$v=x_{m-1}f(x_{m-1})+\int_{x_{m-1}}^{\infty}f(t)\ dt$$

and

$$v = x_i \left[f(x_{i-1}) - f(x_i) \right] \ , \forall i \in \{1, \ldots, m-1\}.$$

Therefore the coordinates $x_1, \ldots, x_{m-1}$ have to be calculated by choosing x_{m-1} suitably and using

$$x_{m-(k+1)} = f^{-1}\left(\frac{v}{x_{m-k}} + f(x_{m-k}) \right) \ , \ k = 1, \ldots, m.$$

The following algorithm reflects the idea for the set up of the Ziggurat presented in (Marsaglia and Tsang, 2000). Due to the fact that a correct or adequate value of x_{m-1} is previously unknown, a try and error method is used which finally accepts the coordinates $x_{m-1}, \ldots, x_1$ if x_0 equals approximately 0.

Algorithm 9 (Nested intervals method for the set up of the Ziggurat)**.**

1. Set $r_U = 10$ and $r_L = 0$. Put $r_U = 10\ r_U$ until $r_U f(r_U) + \int_{r_U}^{\infty} f(x)\ dx < 1/m$.
2. Form $x_{m-1} = (r_U + r_L)/2$, $v = x_{m-1} f(x_{m-1}) + \int_{x_{m-1}}^{\infty} f(t)\ dt$.
3. Calculate $x_0, \ldots, x_{m-2}$ with $x_{m-(k+1)} = f^{-1}\left(\frac{v}{x_{m-k}} + f(x_{m-k}) \right)$.
4. If $|x_0| < \epsilon$, then return $x_1, \ldots, x_{m-1}$.
5. If $x_0 \leq -\epsilon$, then put $r_L = x_{m-1}$ and go on with 2.
6. If $x_0 \geq \epsilon$, then put $r_U = x_{m-1}$ and go on with 2.

References

Becker, R. A., Chambers, J. M., and Wilks, A. R. (1988). *The new S Language.* Wadsworth and Brooks/Cole.

Box, G. E. P. and Muller, M. E. (1958). A note on the generation of random normal deviates. *Annals of Mathematical Statistics*, 29(2):610–611.

Box, G. E. P. and Tiao, C. G. (1973). *Bayesian inference in statistical analysis.* John Wiley & Sons, Inc.

Brent, R. (1973). *Algorithms for minimization without derivatives.* Englewood Cliffs, NJ: Prentice-Hall.

Jöhnk, M. D. (1964). Erzeugung von betaverteilten und gammaverteilten Zufallszahlen. *Metrika*, 8:5–15.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1994). *Continuous Univariate Distributions*, volume 1. Wiley, 2nd edition.

Malham, S. J. A. and Wiese, A. (2010). Positive volatility simulation in the Heston model. arXiv:0802.4411v3.

Marsaglia, G. (1998). The Monty Python method for generating random variables. *ACM Transactions on Mathematical Software*, 24(3):341–350.

Marsaglia, G. and Bray, T. (1964). A convenient method for generating normal variables. *SIAM Review*, 6:260–264.

Marsaglia, G. and Tsang, W. W. (1984). A fast, easily implemented method for sampling from decreasing or symmetric unimodal density functions. *SIAM Journal of Scientific and Statistical Computing*, 5:349–359.

Marsaglia, G. and Tsang, W. W. (2000). The Ziggurat method for generating random variables. *Journal of Statistical Software*, 5:1–7.

Moszyńska, M. and Richter, W.-D. (2011). Reverse triangle inequality. Anti-norms and semi-anti-norms. *submitted*.

Nardon, M. and Pianca, P. (2009). Simulation techniques for generalized Gaussian densities. *Journal of Statistical Software and Simulation*, 79(11):1317–1329.

Pogány, T. K. and Nadarajah, S. (2010). On the characteristic function of the generalized normal distribution. *Comptes Rendus Mathématique*, 348(3):203–206.

Richter, W.-D. (2007). Generalized spherical and simplicial coordinates. *Journal of Mathematical Analysis and Applications*, 336:1187–1202.

Richter, W.-D. (2009). Continuous $l_{n,p}$-symmetric distributions. *Lithuanian Mathematical Journal*, 49:93–108.

Rubinstein, R. Y. (1981). *Simulation and the Monte Carlo method*. John Wiley & Sons, Inc, 1st edition.

Saatci, E. and Akan, A. (2010). Respiratory parameter estimation in non-invasive ventilation based on generalized Gaussian noise models. *Signal processing*, 90(2):480–489.

Subbotin, M. T. (1923). On the law of frequency errors. *Mat. Sbornik*, 31:296–301.