Electronic Journal of Statistics
Vol. 17 (2023) 2180–2209
ISSN: 1935-7524
https://doi.org/10.1214/23-EJS2149

Efficient sampling from the PKBD distribution

Lukas Sablica, Kurt Hornik and Josef Leydold

Institute for Statistics and Mathematics, Vienna University of Economics and Business, Austria
url: https://www.wu.ac.at/en/statmath
e-mail: Lukas.Sablica@wu.ac.at; Kurt.Hornik@wu.ac.at; Josef.Leydold@wu.ac.at

Abstract: In this paper we present and analyze random number generators for the Poisson Kernel-Based Distribution (PKBD) on the sphere. We show that the only currently available sampling scheme presented in Golzy and Markatou (2020) can be improved by a better selection of hyperparameters but still yields an unbounded rejection constant as the concentration parameter approaches 1. Furthermore, we introduce two additional and superior sampling methods for which boundedness in the above mentioned case can be obtained. The first method proposes initial draws from angular central Gaussian distribution and offers uniformly bounded rejection constants for a significant part of the PKBD parameter space. The second method uses adaptive rejection sampling and the results of Ulrich (1984) to sample from the projected Saw distribution (Saw, 1978). Finally, both new methods are compared in a simulation study.

Keywords and phrases: Sampling, rejection sampling, simulation, Poisson kernel, PKBD distribution, spherical distributions.

Received July 2022.

Contents

1. Introduction

Directional statistics has attracted a fair amount of attention over the past years. New developments in the fields of mixture modeling, special function approximation, estimation and random sampling of direction distributions have allowed to develop new algorithms that are capable of recognizing various patterns on multidimensional spheres and cluster data with similar directions into groups using probabilistic methods. An example is the work of Hornik and Grün (2014), where von Mises-Fisher distributions are used to analyse sentiment of submitted abstracts.

Golzy and Markatou (2020) introduce the Poisson Kernel-Based Distribution (PKBD) family for spherical data and propose an efficient algorithm for its estimation. Unlike other commonly employed spherical distributions, the PKBD densities are straightforward to compute, making them particularly attractive for mixture modeling of spherical data. This also amplifies the need for a fast and efficient random number generator for the distribution. The samples from such a generator can then be used in various applications where simulation techniques are needed. An example would be the use of the random draws to accompany the estimation algorithms and assess the uncertainty of the parameters using methods such as the parametric boostrap (O'Hagan et al., 2019), which specifically was proposed in the context of mixture modeling for the likelihood ratio test to assess the number of components (McLachlan, 1987). In terms of Bayesian modeling, random number samples can directly be applied in the process of sampling from a predictive distribution or as a proposal distribution in a Metropolis-Hastings algorithm.

Random variate generation for distributions on the sphere involves generating random points on the surface of a sphere. This is useful in a variety of fields, including physics, astronomy, and geology. In order to generate random variates from a desired distribution on the sphere, one common technique is the rejection sampling method. This method involves generating random points over a larger, enclosing shape called hat function, and then rejecting points that are outside of the desired distribution. The ratio of the volume below the hat function and the given density is called the rejection constant, denoted by $\mathcal{R}$ in this contribution. This constant gives the expected number of iterations of the acceptance-rejection loop of the algorithm and equals the reciprocal of the acceptance probability (which is often used as an alternative characterization). Obviously the choice of the hat function is crucial for the performance of the rejection method, in particular for high dimensional sampling problems. For more details, see for example Devroye (1986).

Golzy and Markatou (2020) suggest to sample from the PKBD via rejection sampling using von Mises-Fisher (vMF) envelopes. In this paper we show that the proposed sampling scheme is not uniformly bounded and that it can be improved by a proper selection of the hyper parameters. Moreover, we show that better results can be achieved when other proposal distributions are selected. The first algorithm we present uses angular central Gaussian (ACG) envelopes as a proposal distribution, and compared to the vMF envelopes, it offers bounded rejection constant as the concentration parameter goes to the extreme case equal to 1. Furthermore, the presented sampling scheme offers high efficiency, where the sampling from the proposal distribution overcomes the curse of dimensionality and has cost which scales only linearly in the dimension. The second algorithm utilizes adaptive rejection sampling from the projected Saw distribution (Saw, 1978), based on the results presented in Ulrich (1984). In addition, both algorithms are compared in a brief simulation study, which shows a great balance between the algorithms, suggesting a large importance value for both algorithms depending on the various tasks in the application.

All proofs of the results in the following sections are given in an appendix.

2. PKBD distribution

With $\mathbb{S}^{d-1} = \{x \in \mathbb{R}^d : \|x\| = 1\}$ the unit sphere in $\mathbb{R}^d$ and υ_d the uniform distribution on $\mathbb{S}^{d-1}$, the density of the PKBD with parameters $0 \leq \rho < 1$ and $\mu \in \mathbb{S}^{d-1}$ with respect to υ_d is given by

$$f_{\mathrm{PKBD}}(x|\rho, \mu) = \frac{1-\rho^2}{\|x - \rho\mu\|^d}, \qquad x \in \mathbb{S}^{d-1}.$$

Clearly, for $\rho = 0$ we get the uniform distribution on the sphere, and using densities with respect to this needs no additional normalizing constant. For $\rho \to 1-$, the PKBD with parameters ρ and μ tends to the Dirac distribution at μ.

The distribution can be considered as a special case ($\xi = d$) of the family of densities

$$f(x) \propto \|x - \rho\mu\|^{-\xi} \quad x \in \mathbb{S}^{d-1}, \tag{1}$$

with $\xi > 0$. It arises for example as an exit distribution from Brownian Motion on unit sphere in $\mathbb{R}^d$ (Kato and Jones, 2013; Durrett, 1984). Another famous member of the family (1) is the spherical Cauchy distribution with $\xi = 2(d-1)$ (Kato and McCullagh, 2020). On the contrary to PKBD, spherical Cauchy distribution can be sampled directly by applying suitable Möbius transformation on the sphere to the uniform distribution on the sphere.

The following reparametrization allows to write the PKBD density in the form $g_d(\lambda\mu'x)/c_{d,\lambda}$, i.e., as a Saw distribution (Saw, 1978) with shape function g_d, direction parameter μ and concentration parameter λ.

Theorem 1. *Let*

$$\lambda = \lambda(\rho) = \frac{2\rho}{1+\rho^2}. \tag{2}$$

Then $\lambda(\rho)$ increases from 0 to 1 as ρ goes from 0 to 1, with inverse

$$\rho = \rho(\lambda) = \frac{\lambda}{1+\sqrt{1-\lambda^2}}, \tag{3}$$

and the PKBD density can be written as $g_d(\lambda\mu'x)/c_{d,\lambda}$ with

$$g_d(t) = (1-t)^{-d/2}$$

and

$$c_{d,\lambda} = \frac{2^{d-2}}{\sqrt{1-\lambda^2}(\sqrt{1+\lambda}+\sqrt{1-\lambda})^{d-2}}.$$

3. Simulating with vMF envelopes

A random vector in $\mathbb{S}^{d-1}$ has a von Mises-Fisher distribution with parameters $\kappa \geq 0$ and $\mu \in \mathbb{S}^{d-1}$ if its density with respect to the uniform distribution on the unit sphere is given by

$$f_{\mathrm{vMF}}(x|\kappa, \mu) = \frac{e^{\kappa \mu' x}}{H_{d/2-1}(\kappa)},$$

where $H_\nu(\kappa) = {}_0F_1(;\nu+1;\kappa^2/4) = \frac{\Gamma(\nu+1)}{(\kappa/2)^\nu} I_\nu(\kappa)$ with ${}_0F_1$ and I_ν being the confluent hypergeometric limit function (e.g., Mardia and Jupp, 2009, page 352) and modified Bessel function of the first kind (DLMF, Eq. 10.25.2), respectively.

Golzy and Markatou (2020) suggest sampling from the PBKD with parameter ρ and μ using vMF envelopes with concentration parameter

$$\kappa_\rho = \frac{d\rho}{1+\rho^2} = \frac{d\lambda}{2}$$

and direction parameter μ, observing however that this choice yields rejection constants which do not remain bounded as $\rho \to 1-$, so that these samplers are not feasible for large concentration parameters.

The following result shows that one can find concentration parameters with smaller rejection constants.

Theorem 2. *Using κ_ρ has rejection constant*

$$H_\nu(\kappa_\rho) e^{-\kappa_\rho} \frac{1+\rho}{(1-\rho)^{d-1}} \approx H_\nu(d/2) e^{-d/2} \frac{1+\rho}{(1-\rho)^{d-1}}$$

as $\rho \to 1-$. The smallest rejection constants for sampling from PKBD distribution using vMF envelopes are obtained using $\kappa_\rho^* = \frac{d}{2}\log\frac{1+\rho}{1-\rho}$, with value

$$\mathcal{R}_{d,\rho}^{\mathrm{vMF}} = \frac{H_\nu(\kappa_\rho^*)}{(1-\rho^2)^{d/2-1}}$$

and as $\rho \to 1-$,

$$\mathcal{R}_{d,\rho}^{\mathrm{vMF}} \approx \frac{\Gamma(d/2) 2^{d/2-1}}{\sqrt{2\pi}(d/2)^{(d-1)/2}} \frac{1}{\left(\log\frac{1+\rho}{1-\rho}\right)^{(d-1)/2}} \frac{1+\rho}{(1-\rho)^{d-1}} \to \infty.$$

We see that using κ_ρ^* gives rejection constants which diverge as $\rho \to 1-$ at a "slightly smaller" rate than the rejection constants for κ_ρ, but still are computationally infeasible for large concentration parameters. This behavior can for the most part be explained by the fact that the PKBD has heavier tails than vMF and hence vMF cannot provide an efficient envelope for sampling from PKBD under high concentration. A similar situation occurs also for the previously mentioned spherical Cauchy distribution, which has lighter tails than the PKBD and hence also cannot serve as an efficient proposal distribution.

The following two sections propose two new sampling algorithms that offer uniformly bounded rejection constants for fixed d and *all* ρ.

4. Simulating with angular central Gaussian envelopes

The ACG distribution with parameter Ω (a symmetric, positive definite matrix) has density

$$f_{\mathrm{ACG}}(x|\Omega) = \det(\Omega)^{1/2}(x'\Omega x)^{-d/2}, \qquad x \in \mathbb{S}^{d-1}.$$

Clearly,

$$\max_{x\in\mathbb{S}^{d-1}} \frac{f_{\mathrm{PKBD}}(x|\rho(\lambda),\mu)}{f_{\mathrm{ACG}}(x|\Omega)} = \frac{\det(\Omega)^{-1/2}}{c_{d,\lambda}} \max_{x\in\mathbb{S}^{d-1}} (1-\lambda\mu'x)^{-d/2}(x'\Omega x)^{d/2}$$

over all feasible Ω. Following Kent, Ganeiber and Mardia (2018), we restrict our attention to those Ω for which $x'\Omega x$ is a function of $\mu'x$ (i.e., where $\mu'x$ is a sufficient statistic for the ACG as it is for the PKBD), using $\Omega(\beta,\mu) = I - \beta\mu\mu'$ with $\beta < 1$. In this case, $\det(\Omega(\beta,\mu)) = 1-\beta$ and $x'\Omega(\beta,\mu)x = 1-\beta(\mu'x)^2$. Note that the ACG distribution generalizes the uniform distribution on the sphere and in the case of $\Omega = I$ (i.e., $\beta = 0$) simplifies to it.

Theorem 3. *Let $\beta_\lambda = \lambda/(2-\lambda)$. The optimal rejection constant of sampling from PKBD distribution using ACG envelopes is of the form*

$$\mathcal{R}^{\mathrm{ACG}}_{d,\rho(\lambda)} = \frac{2\sqrt{1-\lambda^2}}{1+\sqrt{1-\lambda^2}} \min_{\beta_\lambda\le\beta<1} \frac{1}{\sqrt{1-\beta}} \left(\frac{1+\sqrt{1-\lambda^2}}{1+\sqrt{1-\lambda^2/\beta}}\right)^{d/2}.$$

Proposition 1. *Let*

$$C_{d,\lambda}(\beta) = -4(d-1)\beta^3 + (4d-\lambda^2(d-2)^2)\beta^2 + 2d(d-2)\lambda^2\beta - d^2\lambda^2.$$

The optimal $\beta =: \beta^*_{d,\lambda}$ of

$$\min_{\beta_\lambda\le\beta<1} \frac{1}{\sqrt{1-\beta}} \left(\frac{1+\sqrt{1-\lambda^2}}{1+\sqrt{1-\lambda^2/\beta}}\right)^{d/2}$$

is given by the unique root of $C_{d,\lambda}$ in $(\beta_\lambda, 1)$.

Proposition 2. *$\mathcal{R}^{\mathrm{ACG}}_{d,\rho}$ is increasing in d and as $d\to\infty$,*

$$\mathcal{R}^{\mathrm{ACG}}_{d,\rho}/\sqrt{d} \to \sqrt{e}\rho\sqrt{1-\rho^2}.$$

Proposition 3. *As $\rho\to 1-$,*

$$\mathcal{R}^{\mathrm{ACG}}_{d,\rho} \approx \frac{(1+\rho)\sqrt{1+\rho^2}}{\sqrt{2}} \to 2.$$

The above approaches allow to identify limits but fail to give sharp bounds for the rejection constant. These are investigated in the following theorems.

Theorem 4. *Let*

$$Z_{d,\lambda}(u) = \frac{2\sqrt{1-\lambda^2}}{1+\sqrt{1-\lambda^2}} \frac{1}{\sqrt{u-1}} \left(\frac{1+\sqrt{1-\lambda^2}}{1+\sqrt{1-u\lambda^2}} \right)^{d/2},$$

$\tilde{u}_{d,\lambda} = \left(1-\tilde{v}_{d,\lambda}^2\right)/\lambda^2$, $\tilde{v}_{d,\lambda} = \max(v_{d,\lambda}, 1-\lambda)$ *and* $v_{d,\lambda} = \frac{d(1-\lambda^2)}{1+\sqrt{1+d(d+2)(1-\lambda^2)}}$. *Then the following inequalities are satisfied*

$$Z_{d,\lambda}(\tilde{u}_{d,\lambda}) \le \mathcal{R}^{\mathrm{ACG}}_{d,\rho(\lambda)} \le \sqrt{\tilde{u}_{d,\lambda}} Z_{d,\lambda}(\tilde{u}_{d,\lambda}).$$

Furthermore, let $d \to \infty$ *and* $\lambda \to 1-$ *in a way that* $d(1-\rho(\lambda)) = \omega$. *Then*

$$\mathcal{R}^{\mathrm{ACG}}_{d,\rho(\lambda)} \approx Z_{d,\lambda}(\tilde{u}_{d,\lambda}) \to \sqrt{2}\sqrt{1+\sqrt{1+\omega^2}}\, e^{\frac{1}{2}(1+\omega-\sqrt{1+\omega^2})}$$

and thus is bounded for fixed ω.

Note that as $d \to \infty$, clearly $\tilde{v}_{d,\lambda} \to \sqrt{1-\lambda^2}$ and hence $\tilde{u}_{d,\lambda} \to 1$. Similarly, if $\lambda \to 1-$, $\tilde{v}_{d,\lambda} \to 0$ and again $\tilde{u}_{d,\lambda} \to 1$. So in both cases where d or λ approach the upper limit, the bounds are asymptotically sharp (and we can re-obtain the limits using the bounds). On the other hand, for $\lambda \le 2/(d+1)$, $\tilde{u}_{d,\lambda} = u_\lambda = 2/\lambda - 1$, which tends to infinity as $\lambda \to 0+$. In this situation, the lower bound is much too small.

Theorem 5. *The optimal rejection constant* $\mathcal{R}^{\mathrm{ACG}}_{d,\rho}$ *satisfies*

$$\mathcal{R}^{\mathrm{ACG}}_{d,\rho} \le \sqrt{2e}\sqrt{d}\frac{(1+\rho)\sqrt{1+\rho^2}}{2} \le 2\sqrt{e}\sqrt{d}.$$

According to Proposition 2, $\mathcal{R}^{\mathrm{ACG}}_{d,\rho}/\sqrt{d} \to \sqrt{e}\rho\sqrt{1-\rho^2}$ where $\rho\sqrt{1-\rho^2} = \sqrt{\rho^2(1-\rho^2)}$ is at most $1/2$, so the (worst case) upper bound is for $d \to \infty$ off by a factor of 4.

Coming back to the sampling algorithm, we have a rejection ratio of the form

$$\frac{\tilde{f}_{\mathrm{PKBD}}(x|\lambda,\mu)}{\tilde{f}_{\mathrm{ACG}}(x|\Omega)\,\tilde{M}} \le 1,$$

with $\tilde{f}_{\mathrm{PKBD}}(x|\lambda,\mu) = (1-\lambda\mu' x)^{-d/2}$ and $\tilde{f}_{\mathrm{ACG}}(x|\Omega) = (x'\Omega x)^{-d/2} = (x'(I - \beta\mu\mu')x)^{-d/2} = (1-\beta(\mu' x)^2)^{-d/2}$ being the unnormalized densities and

$$\tilde{M} = \left(\frac{2}{1+\sqrt{1-\lambda^2/\beta}}\right)^{d/2}.$$

According to Proposition 1, the optimal $\beta =: \beta^*_{d,\lambda}$ is given by the unique root of $C_{d,\lambda}(\beta)$ in $(\lambda(2-\lambda), 1)$. Thus, we accept the draw x if

$$\log(U) \le \frac{d}{2}\left(-\log(1-\lambda\mu' x) + \log\left(1-\beta^*_{d,\lambda}(\mu' x)^2\right) - \log\left(\frac{2}{1+\sqrt{1-\lambda^2/\beta^*_{d,\lambda}}}\right)\right), \tag{4}$$

where $U \sim U(0,1)$. Recall that if $y \sim N_p(0,\Sigma)$, then $x = y/\|y\| \sim \mathrm{ACG}(\Omega)$ with $\Omega = \Sigma^{-1}$ (Mardia and Jupp, 2009). Obviously, to evaluate the right-hand side of (4) one only needs the value for $\mu' x$, which can be easily calculated using the following proposition.

Proposition 4. *Let Ω be a rank-1 updated identity matrix of the form $\Omega = I - \beta_0\mu\mu'$, with $\|\mu\| = 1$. Then*

$$\Omega^{-1} = I + \beta_1\mu\mu', \qquad \Omega^{-1/2} = I + \beta_2\mu\mu',$$

with $\beta_1 = \beta_2(\beta_2+2) = \beta_0/(1-\beta_0)$ and $\beta_2 = -1 + 1/\sqrt{1-\beta_0}$.

Thus for $x = y/\|y\|$ with $y = \Omega^{-1/2}z$, where $z \sim N(0,I)$, we have

$$\mu' x = \frac{\mu' y}{\|y\|} = \frac{\mu'\Omega^{-1/2}z}{\sqrt{y'y}} = \frac{\mu' z + \beta_2\mu' z}{\sqrt{z'\Omega^{-1}z}} = \frac{\mu' z + \beta_2\mu' z}{\sqrt{z'z + \beta_1(\mu' z)^2}}.$$

The full sampling procedure is summarized in Algorithm 1.

Algorithm 1 Generator for PKBD distribution with parameters λ and μ using ACG envelopes

$(\beta_0{=})\ \beta^*_{d,\lambda} \leftarrow$ unique root of $C_{d,\lambda}(\beta)$ in $(\lambda(2-\lambda),1)$ ▷ e.g. using root solver
$\beta_1 \leftarrow \beta^*_{d,\lambda}/(1-\beta^*_{d,\lambda})$
$\beta_2 \leftarrow -1 + 1/\sqrt{(1-\beta^*_{d,\lambda})}$
repeat ▷ acceptance-rejection loop
  Sample $U \sim U(0,1)$
  Sample $Z_i \sim N(0,1)$ for $i = 1,2,\ldots,d$
  $z \leftarrow (Z_1, Z_2, \ldots, Z_d)$
  $q \leftarrow (\mu' z + \beta_2\mu' z)/\sqrt{z'z + \beta_1(\mu' z)^2}$
until $\log(U) \leq \frac{d}{2}\left(-\log(1-\lambda q) + \log\left(1 - \beta^*_{d,\lambda}q^2\right) - \log\left(\frac{2}{1+\sqrt{1-\lambda^2/\beta^*_{d,\lambda}}}\right)\right)$
return $x \leftarrow (z + \beta_2(\mu' z)\mu)/\sqrt{z'z + \beta_1(\mu' z)^2}$

5. Simulating with adaptive rejection sampling from projected Saw distributions

Following the results of Theorem 1, we can rewrite the PKBD density in the form $g_d(\lambda\mu' x)/c_{d,\lambda}$, which resembles the family of densities introduced in Saw (1978) of the form

$$\frac{g(\lambda u' x)}{c_{g;\lambda,(d-1)/2}},$$

where for $\gamma = (d-1)/2 > 0$

$$c_{g;\lambda,\gamma} = \frac{1}{\mathrm{B}(1/2,\gamma)}\int_{-1}^{1} g(\lambda t)(1-t^2)^{\gamma-1}\,dt.$$

Here $g(\cdot)$ is a function from $\mathbb{R}$ to $[0,\infty)$ controlling the shape of the distribution, $u \in \mathbb{S}^{d-1}$ is a direction parameter and $\lambda \geq 0$ is a concentration parameter. Note however that unlike for the vMF and Watson distributions, g_d depends on d, and that infinite concentration is obtained for $\lambda \to 1-$ (instead of $\lambda \to \infty$). This representation suggests exploring the possibility to obtain better samplers for the PKBD distribution using Theorem 1 of Ulrich (1984).

Theorem 6 (Ulrich, 1984)**.** *Let W be a random variable with density*

$$\frac{g(\lambda z_d)}{c_{g;\lambda,(d-1)/2}} \frac{(1-z_d^2)^{(d-3)/2}}{\mathrm{B}(1/2,(d-1)/2)} =: f_{g,\lambda,d}(z_d)$$

and let $Y \sim \upsilon_{d-1}$ be independent of W then the vector X, where

$$X' = \left(\sqrt{1-W^2}Y', W\right),$$

has density $\mathrm{Saw}_d(g,u,\lambda)$ *with modal vector* $u' = (0,0,\ldots,1)$.

We will refer to the distribution with density $f_{g;\lambda,d}$ as the *projected Saw distribution* with parameters d, g and λ, symbolically $\mathrm{pSaw}_d(g,\lambda)$. Thus, to simulate $X \sim \mathrm{Saw}_d(g,u,\lambda)$, we can draw from Saw-type distributions on the unit sphere $\mathbb{S}^{d-1}$ with densities proportional to $g(\lambda\mu' x)$ via

$$U(\sqrt{1-W^2}Y', W)',$$

where Y and W are independently drawn from, respectively, the uniform distribution on $\mathbb{S}^{d-2}$ and density proportional to $g(\lambda t)(1-t^2)^{(d-3)/2}$ on $(-1,1)$, and U is an orthogonal matrix for which $Ue_d = \mu$ (where e_d is the d-th Cartesian unit vector). We note that the uniformly distributed part Y can be efficiently sampled by $Y = Z/\|Z\|$, where the elements of Z are i.i.d. with standard normal distribution. For more informations see, e.g., Deak (1979) or Hörmann, Leydold and Derflinger (2004).

To use this approach for sampling from the PKBD, we thus have to find appropriate samplers for the density proportional to

$$f_{d,\lambda}(t) = (1-\lambda t)^{-d/2}(1-t^2)^{(d-3)/2}. \tag{5}$$

The scaled log-density on $(-1,1)$ is then

$$\begin{aligned} L_{d,\lambda}(t) &:= 2\log(f_{d,\lambda}(t)) \\ &= -d\log(1-\lambda t) + (d-3)\log(1-t^2) \\ &= -d\log(1-\lambda t) + (d-3)\left(\log(1+t) + \log(1-t)\right). \end{aligned} \tag{6}$$

Theorem 7. *Let $I = (-1,1)$, $f_{d,\lambda}(t)$ be the density* (5)*, $L_{d,\lambda}(t)$ be the log-density* (6) *and*

$$t_{d,\lambda,1} = \frac{d\lambda}{d-3-\sqrt{(d-3)^2 - d(d-6)\lambda^2}},$$

$$t_{d,\lambda,2} = \frac{d\lambda}{d - 3 + \sqrt{(d-3)^2 - d(d-6)\lambda^2}},$$

the following then holds:

(1) If $\lambda = 0$, $f_{d,0}$ is log-convex on I with minimum at $t = 0$ for $d < 3$, constant for $d = 3$, and log-concave on I with maximum at $t = 0$ for $d > 3$.
(2) If $d < 3$, $f_{d,\lambda}$ attains its minimum over I at $t_{d,\lambda,1}$. If $d > 3$, $f_{d,\lambda}$ attains its maximum over I at $t_{d,\lambda,2}$.
(3) If $d \leq 3$ and $\lambda > 0$, $f_{d,\lambda}$ is strictly log-convex on I.
(4) If $d > 3$ and $\lambda > 0$, the function $L_{d,\lambda}(t)$ has at most 2 inflection points in I.

Theorem 7 directly shows that the function $L_{d,\lambda}(t)$ is either concave, convex, or has at most 2 inflection points which can be easily calculated. Thus, in the case where these inflection points exist, by splitting the interval I with the arithmetical mean of the 2 roots, the two defined intervals have at most 1 inflection point. Note, that even in the case where the roots of $B_{d,\lambda}$ in I are complex conjugate pair or a double root, by performing this operation on the real parts, the defined intervals have 0 inflection points in the interior. For the log-concave and log-convex parts, adaptive rejection sampling (Gilks and Wild, 1992) can be deployed. For the intervals with exactly one inflection point, Botts, Hörmann and Leydold (2013) have proposed an algorithm where even the exact position of the inflection point is not required. The algorithm in Hörmann, Leydold and Derflinger (2004) replaces the stochastic method for finding construction points with a deterministic one called derandomized adaptive rejection sampling for finding construction points. This method is implemented in CRAN package Tinflex (Leydold, Botts and Hörmann, 2019).

In order to avoid numerical underflow and overflow when calculating the density $\exp(L_{d,\lambda}/2)$, the scaled log-density $L_{d,\lambda}$ can be normalized. This can be achieved by $L_{d,\lambda}(x) - L_{d,\lambda}(t_{d,\lambda,2})$, where $t_{d,\lambda,2}$ is the extremum from the previous theorem. The Tinflex algorithm has the advantage that the rejection constant can be predefined to some given value. In particular it can be selected close to one at the cost of a more expensive setup. In addition, this procedure allows to decide on the rejection after generating only one single random variate, so that overall $O(d)$ operations are needed to sample from the PKBD on $\mathbb{S}^{d-1}$, for any value of the concentration parameter. This is not true for the ACG method which requires to sample a full d-dimensional vector first and in combination with the result of Proposition 2 scales as $O(d^{3/2})$. This clearly indicates a big advantage of the Tinflex method for large dimensions d. On the other hand, this method requires some time consuming setup which makes it not a favorable choice over the ACG sampler if the dimension is small enough and small amount of random variates is needed or the input parameters (d, λ) vary with every iteration.

This method is summarized in Algorithm 2. Per iteration, the algorithm requires to sample the univariate sample from Tinflex and d times from standard normal distribution. Hornik and Grün (2014) presented a similar algorithm, with however less efficient postprocessing for d large. To the contrary, the algorithm in Hornik and Grün (2014) needs only $d-1$ samples from standard normal distribution, but from our experiments this is not worth the additional costs that come with the required QR decomposition.

Algorithm 2 Generator for PKBD distribution with parameters λ and μ using projected Saw distribution

Sample $W \sim \mathrm{pSaw}_d((1-t)^{-d/2}, \lambda)$ ▷ using Tinflex
Sample $Y \sim \upsilon_d$ with $Y = Z/\|Z\|$ and elements of Z i.i.d. standard normal
$Y \leftarrow Y - \mu' Y \mu$ ▷ projection to hyperspace orthogonal to μ
$Y \leftarrow Y/\sqrt{Y'Y}$
$X \leftarrow W\mu + \sqrt{1-W^2}Y$ ▷ projection of W in the direction of μ and addition of the uniform distribution in the orthogonal complement of it
return X

6. Simulation study

In the following simulation study we compare the two newly proposed samplers (ACG-envelope and Tinflex) for different sets of parameters and sample sizes n. The final time was calculated as an average of 300 measurements. The experiments were all initiated from R (R Core Team, 2018) on operating system Ubuntu 18.04 LTS and compiled using GNU Compiler Collection. The ACG algorithm (Algorithm 1) is written in C++ and integrated into R using Rcpp (Eddelbuettel and François, 2011). Algorithm 2 uses Tinflex package which is mainly written in C. The postprocessing was performed in R.

Table 1

Average times of an ACG-sampler (Algorithm 1, left table) and Tinflex-sampler (Algorithm 2, right table) in milliseconds for $n = 10$, for different dimensions d (as rows) and parameters ρ (as columns).

$d =$	3	5	10	20	50	100	200	1000
$\rho = 0.01$	0.02	0.02	0.02	0.02	0.04	0.06	0.13	0.68
0.1	0.02	0.02	0.02	0.03	0.05	0.12	0.27	2.43
0.25	0.02	0.02	0.02	0.04	0.09	0.20	0.53	5.38
0.4	0.02	0.02	0.03	0.04	0.12	0.29	0.77	8.19
0.6	0.02	0.02	0.03	0.05	0.15	0.37	1.00	10.67
0.75	0.02	0.02	0.03	0.05	0.15	0.40	1.02	10.85
0.9	0.02	0.02	0.03	0.05	0.12	0.32	0.81	8.50
0.99	0.02	0.02	0.02	0.04	0.07	0.15	0.34	3.20

$d =$	3	5	10	20	50	100	200	1000
$\rho = 0.01$	0.54	0.25	0.26	0.29	0.32	0.36	0.49	1.22
0.1	0.31	0.24	0.25	0.31	0.30	0.36	0.48	1.23
0.25	0.26	0.28	0.25	0.30	0.33	0.38	0.46	1.23
0.4	0.22	0.24	0.25	0.27	0.30	0.38	0.46	1.22
0.6	0.23	0.25	0.25	0.29	0.33	0.35	0.49	1.24
0.75	0.27	0.25	0.25	0.27	0.33	0.38	0.48	1.22
0.9	0.25	0.29	0.30	0.30	0.32	0.39	0.48	1.23
0.99	0.32	0.33	0.33	0.32	0.35	0.40	0.47	1.20

Table 2

Average times of an ACG-sampler (Algorithm 1, left table) and Tinflex-sampler (Algorithm 2, right table) in milliseconds for $n = 1000$, for different dimensions d (as rows) and parameters ρ (as columns).

$d =$	3	5	10	20	50	100	200	1000
$\rho = 0.01$	0.3	0.4	0.6	1.1	2.6	5.2	11.3	65.0
0.1	0.4	0.5	0.8	1.5	4.2	10.0	25.7	230.8
0.25	0.5	0.7	1.1	2.4	7.3	18.8	50.9	525.8
0.4	0.6	0.8	1.4	3.1	10.4	27.1	73.7	787.3
0.6	0.7	1.0	1.7	3.8	13.0	34.3	95.0	1027.5
0.75	0.7	1.0	1.8	3.8	13.2	35.4	96.9	1059.0
0.9	0.7	1.0	1.5	3.3	10.7	28.4	77.7	839.1
0.99	0.7	0.9	1.2	2.2	5.7	13.2	32.4	309.3

$d =$	3	5	10	20	50	100	200	1000
$\rho = 0.01$	0.6	1.0	1.4	2.4	5.1	10.0	20.6	103.4
0.1	0.7	1.0	1.4	2.6	5.3	9.9	20.3	99.6
0.25	0.8	1.0	1.4	2.4	5.3	10.0	20.3	101.6
0.4	0.8	1.0	1.4	2.4	5.3	10.0	20.4	101.8
0.6	0.8	1.0	1.4	2.4	5.2	9.8	20.1	112.4
0.75	0.8	1.0	1.4	2.4	5.0	9.9	20.3	102.3
0.9	0.8	1.0	1.4	2.3	5.1	9.7	20.0	100.4
0.99	0.8	1.0	1.5	2.4	5.4	10.0	20.0	104.0

The measured times were further compared relatively and visualized in the following graphics. We note that Tinflex, as an adaptive rejection sampling al-

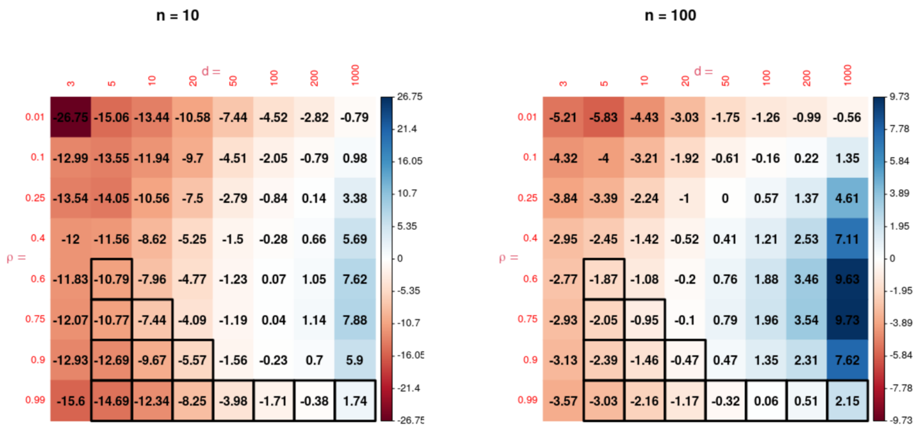

Fig 1. *Relative differences of ACG-sampler and Tinflex-sampler for* $n = 10$ *and* $n = 100$*, with reference value being the smaller of the two values. Negative and positive numbers (toned into red and blue color respectively) indicate the dominance of ACG-sampler and Tinflex-sampler respectively. Cases where the log-density of* $\mathrm{pSaw}_d(g, \lambda)$ *is neither concave nor convex on* $(0, 1)$ *are further annotated with a thick black border.*

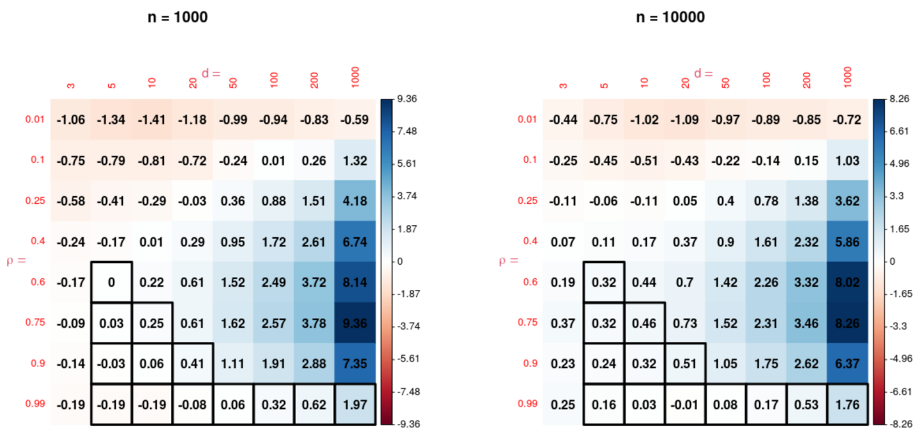

Fig 2. *Relative differences of ACG-sampler and Tinflex-sampler for* $n = 1000$ *and* $n = 10000$*, with reference value being the smaller of the two values. Negative and positive numbers (toned into red and blue color respectively) indicate the dominance of ACG-sampler and Tinflex-sampler respectively. Cases where the log-density of* $\mathrm{pSaw}_d(g, \lambda)$ *is neither concave nor convex on* $(0, 1)$ *are further annotated with a thick black border.*

gorithm, has much more demanding setup period and hence would be expected to be relatively slower for small sample size n. However it should offer superior speed for n large enough thanks to the univariate form of the marginal distribution and bounded rejection constant. This is also confirmed by the results of the simulation study, which show a very nice balance between the samplers.

7. Conclusions

In this paper we presented and analyzed random number generators for the Poisson Kernel-Based Distribution. Altogether we compared tree proposal distributions for the rejection sampling and analyze the efficiency of the corresponding algorithms.

The first method proposes random draws from vMF distribution. We showed that there exists an optimal choice of the concentration parameter κ, which dominates the currently only available rejection sampling scheme of PKBD distribution. Furthermore, Theorem 2 implied that the rejection constant diverges as $\rho \to 1-$, which motivates the need for samplers with other proposal distributions where higher efficiency can be obtained.

As a result, we proposed a rejection sampling algorithm which uses ACG envelopes. This gives rejection constants for which as $\rho \to 1-$, $\mathcal{R}_{d,\rho}^{\text{ACG}} \to 2$ (Proposition 3). What is more, uniformly bounded rejection constants can be achieved on a much larger set of parameter space, because for $d(1-\rho) = \omega$ as $d \to \infty$ and $\rho \to 1-$,

$$\mathcal{R}_{d,\rho}^{\text{ACG}} \to \sqrt{2}\sqrt{1+\sqrt{1+\omega^2}}e^{\frac{1}{2}(1+\omega-\sqrt{1+\omega^2})}$$

(Theorem 4). Furthermore, we approximated and bounded the rejection constant by multiple simple expressions, which can be used for a quick analysis of the worst case scenarios.

The last algorithm uses adaptive rejection sampling and the projection results for the Saw distribution family. This simplifies the whole procedure to a sampling from a univariate distribution, for which it has been shown that its log-density has at most two inflection point and that we can always split the support in a way that there is at most one inflection point in it. This allows to use the adaptive rejection sampling algorithm as for example Tinflex (Leydold, Botts and Hörmann, 2019).

Finally, both new sampling methods are compared in a simulation study for different sets of parameters, showing that adaptive rejection sampling via projected Saw distributions becomes increasingly attractive for large sample sizes (where the additional setup costs become increasingly negligible).

Both new sampling schemes allow for very efficient sampling from the PKBD, adding to its attractiveness for (mixture) modeling of spherical data due to the numerical simplicity of fitting such models established in Golzy and Markatou (2020).

Appendix A: Proofs

Proof of Theorem 1. Clearly, if $x, \mu \in \mathbb{S}^{d-1}$,

$$\|x-\rho\mu\|^2 = 1 - 2\rho\mu'x + \rho^2 = (1+\rho^2)\left(1 - \frac{2\rho}{1+\rho^2}\mu'x\right)$$

so that with

$$\lambda(\rho) = \frac{2\rho}{1+\rho^2}, \qquad g_d(t) = (1-t)^{-d/2} \tag{7}$$

we get the desired form.

Furthermore, $\lambda(\rho)$ increases from 0 to 1 as ρ increases from 0 to 1. Inverting the transformation gives the quadratic equation $q(\rho) = \lambda\rho^2 - 2\rho + \lambda = 0$, which has solutions

$$\frac{2 \pm \sqrt{4-4\lambda^2}}{2\lambda} = \frac{1 \pm \sqrt{1-\lambda^2}}{\lambda}.$$

As for $0 < \lambda < 1$, $q(0) = \lambda > 0$ and $q(1) = 2(\lambda - 1) < 0$, the inverse is given by the smaller root

$$\rho(\lambda) = \frac{1-\sqrt{1-\lambda^2}}{\lambda} = \frac{1-\sqrt{1-\lambda^2}}{\lambda}\frac{1+\sqrt{1-\lambda^2}}{1+\sqrt{1-\lambda^2}} = \frac{\lambda}{1+\sqrt{1-\lambda^2}}.$$

Then

$$\rho(\lambda)^2 = \frac{1-\sqrt{1-\lambda^2}}{\lambda}\frac{\lambda}{1+\sqrt{1-\lambda^2}} = \frac{1-\sqrt{1-\lambda^2}}{1+\sqrt{1-\lambda^2}}$$

so that

$$1-\rho(\lambda)^2 = \frac{2\sqrt{1-\lambda^2}}{1+\sqrt{1-\lambda^2}}, \qquad 1+\rho(\lambda)^2 = \frac{2}{1+\sqrt{1-\lambda^2}} \tag{8}$$

and hence for the reparametrization,

$$\begin{aligned} f_{\mathrm{PKBD}}(x|\rho(\lambda),\mu) &= \frac{1-\rho(\lambda)^2}{(1+\rho(\lambda)^2)^{d/2}(1-\lambda\mu'x)^{d/2}} \\ &= \sqrt{1-\lambda^2}\left(\frac{1+\sqrt{1-\lambda^2}}{2}\right)^{d/2-1} g_d(\lambda\mu'x). \end{aligned}$$

Finally, as

$$\left(\frac{1+\sqrt{1-\lambda^2}}{2}\right)^{d/2-1} = \left(\frac{\sqrt{1+\lambda}+\sqrt{1-\lambda}}{2}\right)^{d-2}$$

we have

$$c_{d,\lambda} = \frac{(1+\rho(\lambda)^2)^{d/2}}{1-\rho(\lambda)^2} = \frac{2^{d-2}}{\sqrt{1-\lambda^2}(\sqrt{1+\lambda}+\sqrt{1-\lambda})^{d-2}}.$$ □

Proof of Theorem 2. We aim to derive the rejection constant of the sampling from PKBD using vMF envelopes. This can be written as

$$\frac{H_\nu(\kappa)}{c_{d,\lambda}} \max_{x\in\mathbb{S}^{d-1}} \frac{(1-\lambda\mu'x)^{-d/2}}{e^{\kappa\mu'x}} = \frac{H_\nu(\kappa)}{c_{d,\lambda}} \frac{1}{\min_{-1\le t\le 1} e^{\kappa t}(1-\lambda t)^{d/2}}.$$

Now $L(t) = \log(e^{\kappa t}(1-\lambda t)^{d/2}) = \kappa t + (d/2)\log(1-\lambda t)$ has first and second derivatives

$$\kappa - \frac{d}{2}\frac{\lambda}{1-\lambda t}, \qquad -\frac{d}{2}\frac{\lambda^2}{(1-\lambda t)^2},$$

hence is concave on $[-1, 1]$ and attains its minimum at ± 1. The values of L at -1 and 1 are, respectively, given by

$$L(-1) = -\kappa + \frac{d}{2}\log(1+\lambda), \qquad L(1) = \kappa + \frac{d}{2}\log(1-\lambda)$$

so that the minimum is attained at $t = 1$ iff

$$\kappa + \frac{d}{2}\log(1-\lambda) \leq -\kappa + \frac{d}{2}\log(1+\lambda) \quad \Leftrightarrow \quad 2\kappa \leq \frac{d}{2}\log\frac{1+\lambda}{1-\lambda}.$$

Equivalently, using

$$1+\lambda = 1 + \frac{2\rho}{1+\rho^2} = \frac{(1+\rho)^2}{1+\rho^2}, \qquad 1-\lambda = 1 - \frac{2\rho}{1+\rho^2} = \frac{(1-\rho)^2}{1+\rho^2},$$

the minimum is attained at $t = 1$ iff

$$\kappa \leq \frac{d}{2}\log\frac{1+\rho}{1-\rho} =: \kappa_\rho^*.$$

For the suggested choice of κ_ρ, we find that

$$2\kappa_\rho - d\log\frac{1+\rho}{1-\rho} = d\left(\frac{2\rho}{1+\rho^2} - \log\frac{1+\rho}{1-\rho}\right)$$

is decreasing in ρ for $0 \leq \rho < 1$ with maximal value 0 attained for $\rho = 0$, so that for κ_ρ, the minimum is attained at $t = 1$.

In general, we obtain with $\rho = \rho(\lambda)$ that for $0 \leq \kappa \leq \kappa_\rho^*$, the rejection constant equals

$$H_\nu(\kappa)\frac{1-\rho^2}{(1+\rho^2)^{d/2}}e^{-\kappa}(1-\lambda)^{-d/2} = H_\nu(\kappa)e^{-\kappa}\frac{1+\rho}{(1-\rho)^{d-1}}$$

whereas for $\kappa \geq \kappa_\rho^*$ it equals

$$H_\nu(\kappa)\frac{1-\rho^2}{(1+\rho^2)^{d/2}}e^{\kappa}(1+\lambda)^{-d/2} = H_\nu(\kappa)e^{\kappa}\frac{1-\rho}{(1+\rho)^{d-1}}.$$

Now $\log(H_\nu(\kappa)e^{-\kappa}) = \log(H_\nu(\kappa)) - \kappa$ has derivative $R_\nu(\kappa) - 1 < 0$, with $R_\nu(\kappa) = I_{\nu+1}(\kappa)/I_\nu(\kappa)$ the Bessel function ratio and hence is decreasing in κ, whereas $\log(H_\nu(\kappa)e^{\kappa})$ has derivative $R_\nu(\kappa) + 1 > 0$ and hence is increasing in κ (Hornik and Grün, 2013). Thus, $\kappa = \kappa_\rho^*$ gives the smallest rejection constant, with value

$$H_\nu(\kappa_\rho^*)\left(\frac{1+\rho}{1-\rho}\right)^{-d/2}\frac{1+\rho}{(1-\rho)^{d-1}} = \frac{H_\nu(\kappa_\rho^*)}{(1-\rho^2)^{d/2-1}}.$$

As $\rho \to 1-$, clearly $\kappa_\rho^* \to \infty$. Using (DLMF, Eq. 10.40.1), for $\kappa \to \infty$ with ν fixed,

$$H_\nu(\kappa) \approx \frac{\Gamma(d/2)2^{d/2-1}e^{\kappa}}{\sqrt{2\pi}\kappa^{(d-1)/2}}$$

so that as $\rho \to 1-$, the optimal vMF rejection constant is

$$H_\nu(\kappa_\rho^*)e^{-\kappa_\rho^*}\frac{1+\rho}{(1-\rho)^{d-1}} \approx \frac{\Gamma(d/2)2^{d/2-1}}{\sqrt{2\pi}(d/2)^{(d-1)/2}}\frac{1}{\left(\log\frac{1+\rho}{1-\rho}\right)^{(d-1)/2}}\frac{1+\rho}{(1-\rho)^{d-1}}.$$

As $u^\epsilon \log(u) \to 0$ as $u \to 0+$ for all $\epsilon > 0$, this in essence tends to ∞ as $\rho \to 1-$ like $(1-\rho)^{1-d}$, and hence is not computationally feasible for large ρ. $\square$

Proof of Theorem 3. We want to express the rejection constant of the sampling from PKBD using ACG envelopes. It follows that

$$\begin{aligned}
&\max_{x\in\mathbb{S}^{d-1}}(1-\lambda\mu' x)^{-d/2}(x'\Omega(\beta,\mu)x)^{d/2} \\
&= \max_{-1\le t\le 1}(1-\lambda t)^{-d/2}(1-\beta t^2)^{d/2} \\
&= \left(\max_{-1\le t\le 1} R_{\beta,\lambda}(t)\right)^{d/2}, \qquad R_{\beta,\lambda}(t) = \frac{1-\beta t^2}{1-\lambda t}.
\end{aligned}$$

We have $\log(R_{\beta,\lambda}(t)) = \log(1-\beta t^2) - \log(1-\lambda t)$ which has first derivative

$$\frac{-2\beta t}{1-\beta t^2} + \frac{\lambda}{1-\lambda t} = \frac{-2\beta t(1-\lambda t)+\lambda(1-\beta t^2)}{(1-\beta t^2)(1-\lambda t)} = \frac{\beta\lambda t^2 - 2\beta t + \lambda}{(1-\beta t^2)(1-\lambda t)}.$$

The numerator $N(t) = \beta\lambda t^2 - 2\beta t + \lambda$ has $N(0) = \lambda \ge 0$ and $N'(t) = 2\beta(\lambda t - 1)$ which is non-decreasing with $N'(1) = 2\beta(\lambda - 1) \le 0$. Hence, N is non-increasing on $I = [-1, 1]$ with minimal value $N(1) = \beta(\lambda - 2) + \lambda$. If $N(1) \ge 0$, or equivalently if $\beta \le \beta_\lambda = \lambda/(2-\lambda)$, N is non-negative on I so that $R_{\beta,\lambda}$ is non-decreasing on I and attains its maximum for $t = 1$. Otherwise, there is a unique $t = t_{\beta,\lambda} \in I$ for which $N(t) = 0$ and $R_{\beta,\lambda}$ attains its maximum. Writing

$$0 = N(t) = \beta\lambda\left(t^2 - \frac{2}{\lambda}t + \frac{1}{\beta}\right)$$

we obtain

$$t_{\beta,\lambda} = \frac{1}{\lambda} - \sqrt{\frac{1}{\lambda^2} - \frac{1}{\beta}}.$$

(clearly, $1/\lambda + \sqrt{1/\lambda^2 - 1/\beta} \ge 1/\lambda > 1$). Using

$$t_{\beta,\lambda} = \frac{\sqrt{\beta} - \sqrt{\beta-\lambda^2}}{\lambda\sqrt{\beta}}\frac{\sqrt{\beta}+\sqrt{\beta-\lambda^2}}{\sqrt{\beta}+\sqrt{\beta-\lambda^2}} = \frac{\lambda}{\sqrt{\beta}(\sqrt{\beta}+\sqrt{\beta-\lambda^2})},$$

we have

$$\begin{aligned}
1-\beta t_{\beta,\lambda}^2 &= 1-\beta\left(\frac{2}{\lambda}t_{\beta,\lambda} - \frac{1}{\beta}\right) = 2\left(1 - \frac{\sqrt{\beta}}{\sqrt{\beta}+\sqrt{\beta-\lambda^2}}\right) \\
&= 2\frac{\sqrt{\beta-\lambda^2}}{\sqrt{\beta}+\sqrt{\beta-\lambda^2}}
\end{aligned}$$

and

$$1 - \lambda t_{\beta,\lambda} = 1 - \left(1 - \sqrt{1 - \lambda^2/\beta}\right) = \sqrt{1 - \lambda^2/\beta} = \frac{\sqrt{\beta - \lambda^2}}{\sqrt{\beta}}$$

so that

$$\frac{1 - \beta t_{\beta,\lambda}^2}{1 - \lambda t_{\beta,\lambda}} = 2\frac{\sqrt{\beta}}{\sqrt{\beta} + \sqrt{\beta - \lambda^2}} = \frac{2}{1 + \sqrt{1 - \lambda^2/\beta}}.$$

Altogether, if $\beta \leq \beta_\lambda$,

$$\begin{aligned}(1-\beta)^{-1/2}\left(\max_{-1\leq t\leq 1} R_{\beta,\lambda}(t)\right)^{d/2} &= (1-\beta)^{-1/2}(R_{\beta,\lambda}(1))^{d/2} \\ &= \frac{(1-\beta)^{(d-1)/2}}{(1-\lambda)^{d/2}}\end{aligned}$$

which is decreasing in β, so that (as $t_{\beta_\lambda,\lambda} = 1$)

$$\min_{\beta<1}(1-\beta)^{-1/2}\left(\max_{-1\leq t\leq 1} R_{\beta,\lambda}(t)\right)^{d/2} = \min_{\beta_\lambda\leq\beta<1} \frac{1}{\sqrt{1-\beta}} R_{\beta,\lambda}(t_{\beta,\lambda})^{d/2}.$$

As

$$\begin{aligned}&\frac{1}{c_{d,\lambda}}\frac{1}{\sqrt{1-\beta}} R_{\beta,\lambda}(t_{\beta,\lambda})^{d/2} \\ &= \sqrt{1-\lambda^2}\left(\frac{1+\sqrt{1-\lambda^2}}{2}\right)^{d/2-1}\frac{1}{\sqrt{1-\beta}}\left(\frac{2}{1+\sqrt{1-\lambda^2/\beta}}\right)^{d/2} \\ &= \frac{2\sqrt{1-\lambda^2}}{1+\sqrt{1-\lambda^2}}\frac{1}{\sqrt{1-\beta}}\left(\frac{1+\sqrt{1-\lambda^2}}{1+\sqrt{1-\lambda^2/\beta}}\right)^{d/2},\end{aligned}$$

the optimal simple ACG envelopes yield rejection constants

$$\mathcal{R}^{\mathrm{ACG}}_{d,\rho(\lambda)} := \frac{2\sqrt{1-\lambda^2}}{1+\sqrt{1-\lambda^2}} \min_{\beta_\lambda\leq\beta<1} \frac{1}{\sqrt{1-\beta}}\left(\frac{1+\sqrt{1-\lambda^2}}{1+\sqrt{1-\lambda^2/\beta}}\right)^{d/2}. \qquad \square$$

For $\lambda = 0$, the above clearly equals one (as the uniform distribution is also a special case of the ACG distribution), so assume $\lambda > 0$.

Proof of Proposition 1. Apart from an additive constant not depending on β, twice the log of the above equals

$$-\log(1-\beta) - d\log(1 + \sqrt{1 - \lambda^2/\beta})$$

the derivative of which with respect to β equals

$$\frac{1}{1-\beta} - \frac{d}{1+\sqrt{1-\lambda^2/\beta}} \times \frac{1}{2}(1-\lambda^2/\beta)^{-1/2} \times \frac{\lambda^2}{\beta^2}$$

$$
\begin{aligned}
&= \frac{1}{1-\beta} - \frac{d\lambda^2}{2\beta^2\sqrt{1-\lambda^2/\beta}} \frac{1}{1+\sqrt{1-\lambda^2/\beta}} \\
&= \frac{1}{1-\beta} - \frac{d\lambda^2}{2\beta^2\sqrt{1-\lambda^2/\beta}} \frac{1-\sqrt{1-\lambda^2/\beta}}{\lambda^2/\beta} \\
&= \frac{1}{1-\beta} - \frac{d}{2}\frac{1-\sqrt{1-\lambda^2/\beta}}{\beta\sqrt{1-\lambda^2/\beta}} \\
&= \frac{2\beta\sqrt{1-\lambda^2/\beta} - d(1-\beta)(1-\sqrt{1-\lambda^2/\beta})}{2\beta(1-\beta)\sqrt{1-\lambda^2/\beta}}
\end{aligned}
$$

where the numerator equals

$$
\begin{aligned}
&(2\beta + d(1-\beta))\sqrt{1-\lambda^2/\beta} - d(1-\beta) \\
&= \frac{(2\beta + d(1-\beta))^2(1-\lambda^2/\beta) - d^2(1-\beta)^2}{(2\beta + d(1-\beta))\sqrt{1-\lambda^2/\beta} + d(1-\beta)} \\
&= \frac{((d-2)\beta - d)^2(\beta - \lambda^2) - d^2\beta(1-\beta)^2}{\beta\left((2\beta + d(1-\beta))\sqrt{1-\lambda^2/\beta} + d(1-\beta)\right)}
\end{aligned}
$$

where in turn the numerator equals

$$
\begin{aligned}
&((d-2)^2\beta^2 - 2d(d-2)\beta + d^2)(\beta - \lambda^2) - d^2\beta(1 - 2\beta + \beta^2) \\
&= \beta^3((d-2)^2 - d^2) + \beta^2(-2d(d-2) - \lambda^2(d-2)^2 + 2d^2) \\
&\quad + \beta(d^2 + 2d(d-2)\lambda^2 - d^2) - d^2\lambda^2 \\
&= -4(d-1)\beta^3 + (4d - \lambda^2(d-2)^2)\beta^2 + 2d(d-2)\lambda^2\beta - d^2\lambda^2 \\
&= C_{d,\lambda}(\beta)
\end{aligned}
$$

so that the sign of the derivative equals the sign of $C_{d,\lambda}$. By straightforward computation,

$$
C_{d,\lambda}(1) = 4(1-\lambda^2) > 0
$$

and

$$
C_{d,\lambda}(\beta_\lambda) = -\frac{4\lambda^2(d-1)(1-\lambda)^2(2d + (1-d)\lambda)}{(2-\lambda)^3} < 0.
$$

As clearly $\lim_{\beta\to-\infty} C_{d,\lambda}(\beta) = \infty$, $C_{d,\lambda}(0) = -d^2\lambda^2 < 0$, and $\lim_{\beta\to\infty} C_{d,\lambda}(\beta) = -\infty$, $Q_{d,\lambda}$ has three real roots in $(-\infty, 0)$, $(\beta_\lambda, 1)$ and $(1, \infty)$, and the optimal $\beta =: \beta^*_{d,\lambda}$ is given by the unique root of $C_{d,\lambda}$ in $(\beta_\lambda, 1)$. □

Proof of Proposition 2. Clearly, for $\beta_\lambda \le \beta < 1$

$$
\frac{1+\sqrt{1-\lambda^2}}{1+\sqrt{1-\lambda^2/\beta}} > 1
$$

so that $\mathcal{R}^{\mathrm{ACG}}_{d,\rho(\lambda)}$ is increasing in d for fixed $\lambda > 0$. To understand the behavior of $\mathcal{R}^{\mathrm{ACG}}_{d,\rho(\lambda)}$ for large d, we proceed as follows. We have

$$C_{d,\lambda}(\beta) = -\lambda^2(1-\beta)^2 d^2 + 4\beta(1-\beta)(\beta-\lambda^2)d + 4\beta^2(\beta-\lambda^2),$$

so that as $d \to \infty$ we must have $\beta^*_{d,\lambda} \to 1$. We can derive an asymptotic approximation via the ansatz $\beta^*_{d,\lambda} = 1 - \alpha_\lambda/d + O(d^{-2})$. As

$$\begin{aligned} C_{d,\lambda}(\beta = 1-\alpha/d) &= -\lambda^2\alpha^2 + 4\beta\alpha(\beta-\lambda^2) + 4\beta^2(\beta-\lambda^2) \\ &= -\lambda^2\alpha^2 + 4\beta(\beta-\lambda^2)(\alpha+\beta) \\ &= -\lambda^2\alpha^2 + 4\left(1-\frac{\alpha}{d}\right)\left(1-\lambda^2-\frac{\alpha}{d}\right)\left(1+\alpha-\frac{\alpha}{d}\right) \\ &= -\lambda^2\alpha^2 + 4(1-\lambda^2)(1+\alpha) + O(d^{-1}), \end{aligned}$$

α_λ must be the positive solution of the quadratic equation $\lambda^2\alpha^2 - 4(1-\lambda^2)\alpha - 4(1-\lambda^2) = 0$, so that

$$\begin{aligned} \alpha_\lambda &= \frac{4(1-\lambda^2) + \sqrt{4^2(1-\lambda^2)^2 + 4^2\lambda^2(1-\lambda^2)}}{2\lambda^2} \\ &= \frac{2}{\lambda^2}\left((1-\lambda^2) + \sqrt{(1-\lambda^2)(1-\lambda^2+\lambda^2)}\right) \\ &= \frac{2\sqrt{1-\lambda^2}}{\lambda^2}(1+\sqrt{1-\lambda^2}). \end{aligned}$$

Thus, as $d \to \infty$,

$$\begin{aligned} \mathcal{R}^{\mathrm{ACG}}_{d,\rho(\lambda)} &\approx \frac{2\sqrt{1-\lambda^2}}{1+\sqrt{1-\lambda^2}}\ \frac{1}{\sqrt{1-\beta}}\left(\frac{1+\sqrt{1-\lambda^2}}{1+\sqrt{1-\lambda^2/\beta}}\right)^{d/2}\Bigg|_{\beta=1-\alpha_\lambda/d} \\ &= \frac{2\sqrt{1-\lambda^2}}{1+\sqrt{1-\lambda^2}}\frac{\sqrt{d}}{\sqrt{\alpha_\lambda}}\left(\frac{1+\sqrt{1-\lambda^2}}{1+\sqrt{1-\lambda^2/\beta}}\right)^{d/2}\Bigg|_{\beta=1-\alpha_\lambda/d}. \end{aligned}$$

As

$$\frac{d\log(1+\sqrt{1-\lambda^2/\beta})}{d\beta} = \frac{1}{1+\sqrt{1-\lambda^2/\beta}} \times \frac{1}{2\sqrt{1-\lambda^2/\beta}} \times \frac{\lambda^2}{\beta^2}$$

we have that as $\beta \to 1$,

$$\log\frac{1+\sqrt{1-\lambda^2}}{1+\sqrt{1-\lambda^2/\beta}} = -\frac{\lambda^2}{2\sqrt{1-\lambda^2}(1+\sqrt{1-\lambda^2})}(\beta-1) + O((\beta-1)^2)$$

and hence as $d \to \infty$,

$$\left(\frac{1+\sqrt{1-\lambda^2}}{1+\sqrt{1-\lambda^2/\beta}}\right)^{d/2}\Bigg|_{\beta=1-\alpha_\lambda/d} = \exp\left(\frac{d}{2}\left(-\frac{1}{\alpha_\lambda}\frac{-\alpha_\lambda}{d} + O(d^{-2})\right)\right)$$

$$\to \sqrt{e}$$

so that

$$\begin{aligned}
\lim_{d\to\infty} \frac{\mathcal{R}^{\mathrm{ACG}}_{d,\rho(\lambda)}}{\sqrt{d}} &= \frac{2\sqrt{1-\lambda^2}}{1+\sqrt{1-\lambda^2}} \frac{\sqrt{e}}{\sqrt{\alpha_\lambda}} \\
&= \sqrt{e}\frac{2\sqrt{1-\lambda^2}}{1+\sqrt{1-\lambda^2}} \frac{\lambda}{\sqrt{2}(1-\lambda^2)^{1/4}(1+\sqrt{1-\lambda^2})^{1/2}} \\
&= \sqrt{2e}\lambda \frac{(1-\lambda^2)^{1/4}}{(1+\sqrt{1-\lambda^2})^{3/2}}.
\end{aligned}$$

This can be further simplified by using the ρ parameter instead of λ. From $\lambda = 2\rho/(1+\rho^2)$,

$$1-\lambda^2 = 1 - \frac{4\rho^2}{(1+\rho^2)^2} = \frac{(1-\rho^2)^2}{(1+\rho^2)^2},$$

so that

$$\sqrt{1-\lambda^2} = \frac{1-\rho^2}{1+\rho^2}, \qquad 1+\sqrt{1-\lambda^2} = \frac{2}{1+\rho^2}$$

and the above limit becomes

$$\sqrt{e}\frac{2^{3/2}\rho}{1+\rho^2}\sqrt{\frac{1-\rho^2}{1+\rho^2}}\frac{(1+\rho^2)^{3/2}}{2^{3/2}} = \sqrt{e}\rho\sqrt{1-\rho^2},$$

which can also be written as $\sqrt{e}\sqrt{\rho^2(1-\rho^2)}$ and hence interestingly is maximized for $\rho^2 = 1/2$, or equivalently, $\rho = 1/\sqrt{2}$. $\square$

Proof of Proposition 3. The range $\beta_\lambda \le \beta < 1$ can be reparametrized as, e.g.,

$$\beta(u) = (1-u) + u\frac{\lambda}{2-\lambda}, \quad 0 < u \le 1.$$

Then

$$1-\beta(u) = u - u\frac{\lambda}{2-\lambda} = 2u\frac{1-\lambda}{2-\lambda}$$

and

$$\begin{aligned}
\mathcal{R}^{\mathrm{ACG}}_{d,\rho(\lambda)} &= \frac{2\sqrt{1-\lambda^2}}{1+\sqrt{1-\lambda^2}} \min_{0<u\le 1} \frac{1}{\sqrt{1-\beta(u)}} \left(\frac{1+\sqrt{1-\lambda^2}}{1+\sqrt{1-\lambda^2/\beta(u)}}\right)^{d/2} \\
&= \frac{2\sqrt{1-\lambda^2}}{1+\sqrt{1-\lambda^2}} \min_{0<u\le 1} \sqrt{\frac{2-\lambda}{2(1-\lambda)u}} \left(\frac{1+\sqrt{1-\lambda^2}}{1+\sqrt{1-\lambda^2/\beta(u)}}\right)^{d/2} \\
&= \frac{\sqrt{2(1+\lambda)(2-\lambda)}}{1+\sqrt{1-\lambda^2}} \min_{0<u\le 1} \frac{1}{\sqrt{u}} \left(\frac{1+\sqrt{1-\lambda^2}}{1+\sqrt{1-\lambda^2/\beta(u)}}\right)^{d/2}.
\end{aligned}$$

As $\lambda \to 1-$, the terms involving λ tend to 2 uniformly in $0 \le u \le 1$, from which

$$\lim_{\lambda \to 1-} \mathcal{R}^{\mathrm{ACG}}_{d,\rho(\lambda)} = 2.$$

Note, that the acceptance rate of 50% under high concentration is an unsurprising result considering the form and bimodality of ACG distribution.

Alternatively, as

$$C_{d,1}(\beta) = -(1-\beta)^2 d^2 + 4\beta(1-\beta)(\beta-1)d + 4\beta^2(\beta-1),$$

we see that $\beta^*_{d,\lambda} \to 1$ as $\lambda \to 1-$. Making the ansatz $\beta^*_{d,\lambda} = 1 - \alpha_d(1-\lambda) + O((1-\lambda)^2)$ gives $\alpha_d = 2$ so that as $\lambda \to 1-$, $1 - \beta^*_{d,\lambda} \approx 2(1-\lambda)$ and

$$\mathcal{R}^{\mathrm{ACG}}_{d,\rho(\lambda)} \approx \frac{2\sqrt{1-\lambda^2}}{1+\sqrt{1-\lambda^2}} \frac{1}{\sqrt{2(1-\lambda)}} = \frac{\sqrt{2(1+\lambda)}}{1+\sqrt{1-\lambda^2}} \to 2.$$

This can be also further rewritten using the ρ parameter. Trivially, if $\lambda \to 1$ then also $\rho \to 1$ and hence

$$\mathcal{R}^{\mathrm{ACG}}_{d,\rho(\lambda)} \approx \frac{\sqrt{2(1+\lambda)}}{1+\sqrt{1-\lambda^2}} = \frac{1-\rho^2}{\sqrt{2}\sqrt{(1-\lambda)}} = \frac{(1+\rho)\sqrt{1+\rho^2}}{\sqrt{2}} \to 2,$$

where (8) was used. □

Proof of Proposition 4. Trivial. □

Proof of Theorem 4. First, write $\beta = 1/u$ and $u_\lambda = 1/\beta_\lambda = 2/\lambda - 1$. As

$$\frac{1}{1-\beta} = \frac{1/\beta}{1/\beta - 1},$$

we have

$$\mathcal{R}^{\mathrm{ACG}}_{d,\rho(\lambda)} = \frac{2\sqrt{1-\lambda^2}}{1+\sqrt{1-\lambda^2}} \min_{1<u\le u_\lambda} \sqrt{\frac{u}{u-1}} \left(\frac{1+\sqrt{1-\lambda^2}}{1+\sqrt{1-u\lambda^2}} \right)^{d/2}.$$

Thus,

$$\mathcal{R}^{\mathrm{ACG}}_{d,\rho(\lambda)} \ge \frac{2\sqrt{1-\lambda^2}}{1+\sqrt{1-\lambda^2}} \min_{1<u\le u_\lambda} \frac{1}{\sqrt{u-1}} \left(\frac{1+\sqrt{1-\lambda^2}}{1+\sqrt{1-u\lambda^2}} \right)^{d/2}.$$

To compute the minimum, substitute $v = \sqrt{1-u\lambda^2}$. Then $v^2 = 1 - u\lambda^2$ so that $u = (1-v^2)/\lambda^2$. Up to an additive constant not depending on v, twice the log of the above is $-\log(1-\lambda^2-v^2) - d\log(1+v)$ which has derivative

$$\frac{2v}{1-\lambda^2-v^2} - \frac{d}{1+v} = \frac{2v(1+v) - d(1-\lambda^2-v^2)}{(1-\lambda^2-v^2)(1+v)}$$

$$= \frac{(d+2)v^2 + 2v - d(1-\lambda^2)}{(1-\lambda^2-v^2)(1+v)},$$

with positive critical point given by

$$\begin{aligned} v_{d,\lambda} &= \frac{-2+\sqrt{4+4d(d+2)(1-\lambda^2)}}{2(d+2)} = \frac{-1+\sqrt{1+d(d+2)(1-\lambda^2)}}{d+2} \\ &= \frac{d(d+2)(1-\lambda^2)}{(d+2)(1+\sqrt{1+d(d+2)(1-\lambda^2)})} = \frac{d(1-\lambda^2)}{1+\sqrt{1+d(d+2)(1-\lambda^2)}}. \end{aligned}$$

As $(d+2)v^2+2v-d(1-\lambda^2)$ is negative for $v=0$ and positive for $v=\sqrt{1-\lambda^2}$, $v_{d,\lambda}$ indeed gives the minimizer of the lower bound over $0<v<\sqrt{1-\lambda^2}$. Now as u goes from 1 to u_λ, $v=\sqrt{1-u\lambda^2}$ goes from $\sqrt{1-\lambda^2}$ to $1-\lambda$, so in fact we need the minimum over $1-\lambda \le v < \sqrt{1-\lambda^2}$, which is thus attained at $\max(v_{d,\lambda}, 1-\lambda)$, where

$$\begin{aligned} & v_{d,\lambda} \ge 1-\lambda \\ \Leftrightarrow\quad & \frac{-1+\sqrt{1+d(d+2)(1-\lambda^2)}}{d+2} \ge 1-\lambda \\ \Leftrightarrow\quad & \sqrt{1+d(d+2)(1-\lambda^2)} \ge 1+(d+2)(1-\lambda) \\ \Leftrightarrow\quad & 1+d(d+2)(1-\lambda^2) \ge 1+2(d+2)(1-\lambda)+(d+2)^2(1-\lambda)^2 \\ \Leftrightarrow\quad & (d+2)(1-\lambda)\left(d(1+\lambda)-2-(d+2)(1-\lambda)\right) \ge 0 \\ \Leftrightarrow\quad & 2(d+1)\lambda - 4 \ge 0 \\ \Leftrightarrow\quad & \lambda \ge 2/(d+1). \end{aligned}$$

We thus have the following. Let

$$\tilde{v}_{d,\lambda} = \max(v_{d,\lambda}, 1-\lambda) = \begin{cases} v_{d,\lambda}, & 2/(d+1) \le \lambda < 1, \\ 1-\lambda, & 0 < \lambda \le 2/(d+1). \end{cases}$$

and

$$Z_{d,\lambda}(u) = \frac{2\sqrt{1-\lambda^2}}{1+\sqrt{1-\lambda^2}} \frac{1}{\sqrt{u-1}} \left(\frac{1+\sqrt{1-\lambda^2}}{1+\sqrt{1-u\lambda^2}} \right)^{d/2}.$$

Then for

$$\tilde{u}_{d,\lambda} = \frac{1-\tilde{v}_{d,\lambda}^2}{\lambda^2}$$

we have

$$Z_{d,\lambda}(\tilde{u}_{d,\lambda}) \le \mathcal{R}^{\mathrm{ACG}}_{d,\rho(\lambda)} \le \sqrt{\tilde{u}_{d,\lambda}} Z_{d,\lambda}(\tilde{u}_{d,\lambda}).$$

Now clearly as $d\to\infty$ and $\rho\to 1$, for $d(1-\rho)=\omega$

$$\xi_{d,\lambda} := d\sqrt{1-\lambda^2} = d(1-\rho^2)/(1+\rho^2) \to \omega,$$

and $\lambda \geq 2/(d+1)$ is satisfied. Thus

$$\tilde{v}_{d,\lambda} = v_{d,\lambda} = \frac{\sqrt{1-\lambda^2}\xi_{d,\lambda}}{1+\sqrt{1+(1+2/d)\xi_{d,\lambda}^2}} \approx \frac{\sqrt{1-\lambda^2}\omega}{1+\sqrt{1+\omega^2}} \to 0,$$

and hence $\tilde{u}_{d,\lambda} \to 1$, making the lower and upper bounds sharp for this scenario.

Furthermore we have that

$$\left(\frac{1+\sqrt{1-\lambda^2}}{1+\sqrt{1-\tilde{u}_{d,\lambda}\lambda^2}}\right)^{d/2} = \left(\frac{1+\sqrt{1-\lambda^2}}{1+\tilde{v}_{d,\lambda}}\right)^{d/2} = \left(\frac{1+\omega/d}{1+\frac{(\omega/d)\xi_{d,\lambda}}{1+\sqrt{1+(1+2/d)\xi_{d,\lambda}^2}}}\right)^{d/2},$$

from which

$$\frac{1+\omega/d}{1+\frac{(\omega/d)\xi_{d,\lambda}}{1+\sqrt{1+(1+2/d)\xi_{d,\lambda}^2}}} = \frac{d+\omega}{d+\frac{\omega\xi_{d,\lambda}}{1+\sqrt{1+(1+2/d)\xi_{d,\lambda}^2}}},$$

where

$$\frac{\omega\xi_{d,\lambda}}{1+\sqrt{1+(1+2/d)\xi_{d,\lambda}^2}} \to \frac{\omega^2}{1+\sqrt{1+\omega^2}}.$$

Moreover

$$\begin{aligned}\frac{d+\omega}{d+\frac{\omega^2}{1+\sqrt{1+\omega^2}}} = \frac{d+\omega}{d-1+\sqrt{1+\omega^2}} &= \left(\frac{d-1+\sqrt{1+\omega^2}}{d+\omega}\right)^{-1} \\ &= \left(1+\frac{\sqrt{1+\omega^2}-\omega-1}{d+\omega}\right)^{-1}\end{aligned}$$

and hence

$$\left(\frac{1+\sqrt{1-\lambda^2}}{1+\sqrt{1-\tilde{u}_{d,\lambda}\lambda^2}}\right)^{d/2} \to e^{\frac{1}{2}\left(1+\omega-\sqrt{1+\omega^2}\right)}.$$

Note that as $\omega \to \infty$ (which is the case if $d \to \infty$ for fixed λ) and as $\omega \to 0$ (which is the case if $\lambda \to 1$ for fixed d) this goes to $\sqrt{e}$ and 1, respectively, confirming the previous results.

For

$$\begin{aligned}\frac{2\sqrt{1-\lambda^2}}{1+\sqrt{1-\lambda^2}}\frac{1}{\sqrt{\tilde{u}_{d,\lambda}-1}} &= \frac{2\sqrt{1-\lambda^2}}{1+\sqrt{1-\lambda^2}}\frac{1}{\sqrt{\frac{1-\tilde{v}_{d,\lambda}^2-\lambda^2}{\lambda^2}}} \\ &= \frac{2\sqrt{1-\lambda^2}}{1+\sqrt{1-\lambda^2}}\frac{\lambda}{\sqrt{1-\tilde{v}_{d,\lambda}^2-\lambda^2}}\end{aligned}$$

it follows that

$$\frac{2\sqrt{1-\lambda^2}}{1+\sqrt{1-\lambda^2}}\frac{\lambda}{\sqrt{1-\tilde{v}_{d,\lambda}^2-\lambda^2}} = \frac{2}{1+\sqrt{1-\lambda^2}}\frac{\lambda}{\sqrt{1-\frac{\xi_{d,\lambda}^2}{(1+\sqrt{1+(1+2/d)\xi_{d,\lambda}^2})^2}}}$$

$$\to 2\frac{1}{\sqrt{1-\frac{\omega^2}{(1+\sqrt{1+\omega^2})^2}}},$$

with

$$2\frac{1}{\sqrt{1-\frac{\omega^2}{(1+\sqrt{1+\omega^2})^2}}} = 2\frac{1+\sqrt{1+\omega^2}}{\sqrt{2(1+\sqrt{1+\omega^2})}} = \sqrt{2}\sqrt{1+\sqrt{1+\omega^2}}.$$

Again, as $\omega \to 0$, $\sqrt{2}\sqrt{1+\sqrt{1+\omega^2}} \to 2$ and as $\omega \to \infty$, $\sqrt{2}\sqrt{1+\sqrt{1+\omega^2}} \approx \sqrt{2\omega} = \sqrt{2d(1-\rho)}$. Note that

$$\sqrt{d}\rho\sqrt{(1-\rho^2)} = \sqrt{d}\rho\sqrt{(1-\rho)(1+\rho)} \approx \sqrt{2d(1-\rho)}.$$

Thus, overall we have that as $d \to \infty$ and $\rho \to 1$ for $d(1-\rho) = \omega$,

$$\mathcal{R}^{\mathrm{ACG}}_{d,\rho(\lambda)} \approx Z_{d,\lambda}(\tilde{u}_{d,\lambda}) \to \sqrt{2}\sqrt{1+\sqrt{1+\omega^2}}e^{\frac{1}{2}(1+\omega-\sqrt{1+\omega^2})}. \qquad \square$$

Proof of Theorem 5. To bound the rejection constant from above, we proceed as follows. Since

$$\frac{d}{du}\log(1+\sqrt{1-u\lambda^2}) = \frac{-\lambda^2}{2(1+\sqrt{1-u\lambda^2})\sqrt{1-u\lambda^2}},$$

we have for $u > 1$ by the mean value theorem that

$$\log\frac{1+\sqrt{1-\lambda^2}}{1+\sqrt{1-u\lambda^2}} = \frac{\lambda^2(u-1)}{2(1+\sqrt{1-\xi\lambda^2})\sqrt{1-\xi\lambda^2}}$$

for some $1 \le \xi \le u$, and hence for $1 < u \le u_\lambda$

$$\log\frac{1+\sqrt{1-\lambda^2}}{1+\sqrt{1-u\lambda^2}} \le \frac{\lambda^2(u-1)}{2(1+\sqrt{1-u_\lambda\lambda^2})\sqrt{1-u_\lambda\lambda^2}} = \frac{\lambda^2(u-1)}{2(2-\lambda)(1-\lambda)}.$$

So if $u = 1+\alpha(1-\lambda)/d \le u_\lambda$,

$$\begin{aligned}
\mathcal{R}^{\mathrm{ACG}}_{d,\rho(\lambda)} &\le \frac{2\sqrt{1-\lambda^2}}{1+\sqrt{1-\lambda^2}}\sqrt{\frac{u}{u-1}}\left(\frac{1+\sqrt{1-\lambda^2}}{1+\sqrt{1-u\lambda^2}}\right)^{d/2} \\
&\le \frac{2\sqrt{1-\lambda^2}}{1+\sqrt{1-\lambda^2}}\sqrt{\frac{1+\alpha(1-\lambda)/d}{\alpha(1-\lambda)/d}}\exp\left(\frac{d}{2}\frac{\lambda^2\alpha(1-\lambda)}{2(2-\lambda)(1-\lambda)d}\right) \\
&= \sqrt{d}\frac{2\sqrt{1+\lambda}}{1+\sqrt{1-\lambda^2}}\sqrt{\frac{1+\alpha(1-\lambda)/d}{\alpha}}\exp\left(\frac{\lambda^2\alpha}{4(2-\lambda)}\right) \\
&\le \sqrt{d}\frac{2\sqrt{1+\lambda}}{1+\sqrt{1-\lambda^2}}\sqrt{\frac{1+\alpha(1-\lambda)/d}{\alpha}}e^{\lambda^2\alpha/4}
\end{aligned}$$

(retaining $2-\lambda$ does not improve the worst case constants). Up to an additive constant not depending on α, twice the log of the above equals

$$\log(1+\alpha(1-\lambda)/d) - \log(\alpha) + \lambda^2\alpha/2$$

which has derivative

$$\begin{aligned}
&\frac{(1-\lambda)/d}{1+\alpha(1-\lambda)/d} - \frac{1}{\alpha} + \frac{\lambda^2}{2} \\
&= \frac{2(\alpha(1-\lambda)/d - (1+\alpha(1-\lambda)/d)) + \lambda^2\alpha(1+\alpha(1-\lambda)/d)}{2\alpha(1+\alpha(1-\lambda)/d)}
\end{aligned}$$

where the numerator equals

$$\frac{\lambda^2(1-\lambda)}{d}\alpha^2 + \lambda^2\alpha - 2 = \frac{\lambda^2(1-\lambda)\alpha^2 + \lambda^2 d\alpha - 2d}{d}.$$

This has its positive root at

$$\begin{aligned}
\alpha &= \frac{-\lambda^2 d + \sqrt{\lambda^4 d^2 + 8d\lambda^2(1-\lambda)}}{2\lambda^2(1-\lambda)} \\
&= \frac{8d\lambda^2(1-\lambda)}{2\lambda^2(1-\lambda)\left(\lambda^2 d + \sqrt{\lambda^4 d^2 + 8d\lambda^2(1-\lambda)}\right)} \\
&= \frac{4d}{\lambda^2 d + \sqrt{\lambda^4 d^2 + 8d\lambda^2(1-\lambda)}} \\
&= \frac{4d}{\lambda\left(\lambda d + \sqrt{\lambda^2 d^2 + 8d(1-\lambda)}\right)} \\
&=: \alpha_{d,\lambda}.
\end{aligned}$$

As the derivative at $\alpha = 0$ is negative, $\alpha_{d,\lambda}$ must give the minimum of the upper bound. In general,

$$\begin{aligned}
& 1+\alpha\frac{1-\lambda}{d} \le u_\lambda = \frac{2}{\lambda} - 1 \\
\Leftrightarrow \quad & \alpha\frac{1-\lambda}{d} \le \frac{2}{\lambda} - 2 = \frac{2(1-\lambda)}{\lambda} \\
\Leftrightarrow \quad & \alpha \le \frac{2d}{\lambda}.
\end{aligned}$$

For $\alpha_{d,\lambda}$ determined above, this is equivalent to

$$2 \le \lambda d + \sqrt{\lambda^2 d^2 + 8d(1-\lambda)}.$$

The right hand side is clearly increasing in d, with minimal value for $d = 2$ given by

$$2\lambda + \sqrt{4\lambda^2 + 16(1-\lambda)} = 2\lambda + \sqrt{4(\lambda^2 - 4\lambda + 4)}$$

$$= 2\lambda + 2(2-\lambda) = 4$$

so indeed we always have $u = 1 + \alpha_{d,\lambda}(1-\lambda)/d \le u_\lambda$, and thus

$$\mathcal{R}^{\mathrm{ACG}}_{d,\rho(\lambda)} \le \sqrt{d}\frac{2\sqrt{1+\lambda}}{1+\sqrt{1-\lambda^2}}\sqrt{\frac{1+\alpha_{d,\lambda}(1-\lambda)/d}{\alpha_{d,\lambda}}}e^{\lambda^2\alpha_{d,\lambda}/4}.$$

Now

$$\begin{aligned}&\frac{1+\alpha_{d,\lambda}(1-\lambda)/d}{\alpha_{d,\lambda}}\\ &= \frac{\lambda\left(\lambda d + \sqrt{\lambda^2 d^2 + 8d(1-\lambda)}\right) + 4(1-\lambda)}{4d}\\ &= \frac{\lambda(\lambda + \sqrt{\lambda^2 + 8(1-\lambda)/d}) + 4(1-\lambda)/d}{4}.\end{aligned}$$

This is clearly decreasing in d, with minimal value for $d = 2$ given by

$$\frac{\lambda(\lambda + \sqrt{\lambda^2 + 4(1-\lambda)}) + 2(1-\lambda)}{4} = \frac{\lambda(\lambda + (2-\lambda)) + 2(1-\lambda)}{4} = \frac{1}{2}.$$

Also,

$$\begin{aligned}\frac{\lambda^2\alpha_{d,\lambda}}{4} &= \frac{\lambda^2}{4}\frac{4d}{\lambda\left(\lambda d + \sqrt{\lambda^2 d^2 + 8d(1-\lambda)}\right)}\\ &= \frac{\lambda d}{\lambda d + \sqrt{\lambda^2 d^2 + 8d(1-\lambda)}} \le \frac{1}{2}.\end{aligned}$$

Altogether, we obtain

$$\mathcal{R}^{\mathrm{ACG}}_{d,\rho(\lambda)} \le \sqrt{d}\frac{2\sqrt{1+\lambda}}{1+\sqrt{1-\lambda^2}}\sqrt{\frac{1}{2}}\sqrt{e} = \sqrt{2e}\frac{\sqrt{1+\lambda}}{1+\sqrt{1-\lambda^2}}\sqrt{d}.$$

With $\lambda = 2\rho/(1+\rho^2)$, $1+\lambda = (1+\rho)^2/(1+\rho^2)$ and

$$\frac{\sqrt{1+\lambda}}{1+\sqrt{1-\lambda^2}} = \frac{1+\rho}{\sqrt{1+\rho^2}}\frac{1+\rho^2}{2} = \frac{(1+\rho)\sqrt{1+\rho^2}}{2}.$$

This is clearly increasing in ρ, and hence for $0 \le \rho < 1$ at most the value at $\rho = 1$, which is $\sqrt{2}$. Thus,

$$\mathcal{R}^{\mathrm{ACG}}_{d,\rho(\lambda)} \le 2\sqrt{e}\sqrt{d}.$$

□

Proof of Theorem 7. $L_{d,\lambda}(t)$ has first derivative

$$L'_{d,\lambda}(t) = d\frac{\lambda}{1-\lambda t} + (d-3)\left(\frac{1}{1+t} - \frac{1}{1-t}\right)$$

and second derivative

$$L''_{d,\lambda}(t) = d\frac{\lambda^2}{(1-\lambda t)^2} - (d-3)\left(\frac{1}{(1+t)^2} + \frac{1}{(1-t)^2}\right).$$

Then

$$L'_{d,\lambda}(t) = d\frac{\lambda}{1-\lambda t} - (d-3)\frac{2t}{1-t^2} = \frac{Q_{d,\lambda}(t)}{(1-\lambda t)(1-t^2)}$$

where the denominator is positive on $(-1,1)$ and the numerator equals

$$\begin{aligned} Q_{d,\lambda}(t) &= d\lambda(1-t^2) - 2(d-3)t(1-\lambda t) \\ &= \lambda t^2(2(d-3)-d) - 2(d-3)t + d\lambda \\ &= (d-6)\lambda t^2 - 2(d-3)t + d\lambda. \end{aligned}$$

Similarly,

$$L''_{d,\lambda}(t) = d\frac{\lambda^2}{(1-\lambda t)^2} - 2(d-3)\frac{1+t^2}{(1-t^2)^2} = \frac{B_{d,\lambda}(t)}{(1-\lambda t)^2(1-t^2)^2},$$

where the denominator is positive on $(-1,1)$ and the numerator equals

$$B_{d,\lambda}(t) = d\lambda^2(1-t^2)^2 - 2(d-3)(1+t^2)(1-\lambda t)^2. \tag{9}$$

Using

$$(1+t^2)(1-\lambda t)^2 = 1 - 2\lambda t + (1+\lambda^2)t^2 - 2\lambda t^3 + \lambda^2 t^4,$$

we find

$$\begin{aligned} B_{d,\lambda}(t) = (6-d)\lambda^2 t^4 + 4(d-3)\lambda t^3 + ((6-4d)\lambda^2 - 2d + 6)t^2 \\ + 4(d-3)\lambda t + (d\lambda^2 - 2d + 6). \end{aligned}$$

From this (1) and (3) are trivial.

Consider $0 < \lambda < 1$. If $d = 3$, $L'_{3,\lambda}(t) > 0$ on I, so $f_{3,\lambda}$ is strictly increasing on I. If $d = 6$, $Q_{6,\lambda}(t) = -6t + 6\lambda$, which has its unique zero at $t = \lambda$. As $Q_{6,\lambda}(-1) > 0$ and $Q_{6,\lambda}(1) < 0$, $f_{6,\lambda}$ has its maximum at $t = \lambda$.

Clearly,

$$Q_{d,\lambda}(-1) = 2(d-3)(1+\lambda), \quad Q_{d,\lambda}(0) = d\lambda, \quad Q_{d,\lambda}(1) = -2(d-3)(1-\lambda).$$

Hence,

$$\operatorname{sgn}(Q_{d,\lambda}(-1)) = \operatorname{sgn}(d-3), \quad Q_{d,\lambda}(0) > 0, \quad \operatorname{sgn}(Q_{d,\lambda}(1)) = -\operatorname{sgn}(d-3).$$

As $Q_{d,\lambda}(t)$ goes to ∞ for $t \to \pm\infty$ if $d > 6$ and to $-\infty$ for $d < 6$, we thus have the following.

- If $d < 3$, $Q_{d,\lambda}$ has one root in $(-1,0)$ and one root in $(1,\infty)$, and the former gives the minimum of $f_{d,\lambda}$ over I.

- If $3 < d < 6$, $Q_{d,\lambda}$ has one root in $(-\infty, -1)$ and one root in $(0, 1)$, and the latter gives the maximum of $f_{d,\lambda}$ over I.
- If $d > 6$, $Q_{d,\lambda}$ has one root in $(0, 1)$ and one root in $(1, \infty)$, and the former gives the maximum of $f_{d,\lambda}$ over I.

Now if $d \neq 6$, the roots of $Q_{d,\lambda}$ are given by

$$\begin{aligned} t_{d,\lambda,1} &= \frac{(d-3) + \sqrt{(d-3)^2 - d(d-6)\lambda^2}}{(d-6)\lambda} \\ &= \frac{(d-3)^2 - ((d-3)^2 - d(d-6)\lambda^2)}{(d-6)\lambda(d-3-\sqrt{(d-3)^2 - d(d-6)\lambda^2}} \\ &= \frac{d\lambda}{d-3-\sqrt{(d-3)^2 - d(d-6)\lambda^2}} \end{aligned}$$

and

$$\begin{aligned} t_{d,\lambda,2} &= \frac{(d-3) - \sqrt{(d-3)^2 - d(d-6)\lambda^2}}{(d-6)\lambda} \\ &= \frac{d\lambda}{d-3+\sqrt{(d-3)^2 - d(d-6)\lambda^2}}. \end{aligned}$$

Clearly, $t_{d,\lambda,1} < 0$ if $d < 6$, and if $d > 6$, $t_{d,\lambda,1} > t_{d,\lambda,2}$. As

$$t_{6,\lambda,2} = \frac{6\lambda}{3+3} = \lambda,$$

(2) follows.

To analyze $B_{d,\lambda}$, we first note that for $\lambda = 0$,

$$B_{d,0}(t) = -2(d-3)(1+t^2)$$

which is always negative for $d > 3$.

Again using Equation (9),

$$B_{d,\lambda}(-1) = -4(d-3)(1+\lambda)^2, \qquad B_{d,\lambda}(1) = -4(d-3)(1-\lambda)^2$$

which are both negative if $d > 3$ and $0 \leq \lambda < 1$. Using

$$\begin{aligned} \frac{d}{dt}(1+t^2)(1-\lambda t)^2 &= 2t(1-\lambda t)^2 - 2\lambda(1+t^2)(1-\lambda t) \\ &= 2(1-\lambda t)\left(t(1-\lambda t) - \lambda(1+t^2)\right) \\ &= 2(1-\lambda t)(-\lambda + t - 2\lambda t^2), \end{aligned}$$

we obtain that

$$B'_{d,\lambda}(t) = -4d\lambda^2 t(1-t^2) - 4(d-3)(1-\lambda t)(-\lambda + t - 2\lambda t^2) \tag{10}$$

so that $B'_{d,\lambda}(0) = 4(d-3)\lambda$,

$$B'_{d,\lambda}(-1) = -4(d-3)(1+\lambda)(-1-3\lambda) = 4(d-3)(1+\lambda)(1+3\lambda),$$

and

$$B'_{d,\lambda}(\lambda) = -4d\lambda^3(1-\lambda^2) + 8(d-3)(1-\lambda^2)\lambda^3 = 4(d-6)\lambda^3(1-\lambda^2).$$

If $3 < d < 6$, $B_{d,\lambda}(t)$ tends to ∞ for $t \to \pm\infty$. Hence, it must have one root in $(-\infty, -1)$ and one root in $(1, \infty)$, and thus can have at most two roots in I. Similarly, as $\lim_{t\to-\infty} B'_{d,\lambda}(t) = -\infty$, $B'_{d,\lambda}(-1) > 0$, $B'_{d,\lambda}(0) > 0$, $B'_{d,\lambda}(\lambda) < 0$, and $\lim_{t\to\infty} B'_{d,\lambda}(t) = \infty$, $B'_{d,\lambda}$ must have one root each in $(-\infty, -1)$, $(0, \lambda)$, and (λ, ∞). In particular, $B_{d,\lambda}$ is always increasing on $[-1, 0]$. Clearly, since $B_{d,\lambda}(-1) < 0$, $B'_{d,\lambda}(-1) > 0$, and $B_{d,\lambda}(\lambda) < 0$, $B'_{d,\lambda}(\lambda) < 0$, the remaining 2 roots are in I, either as a conjugate pair or real roots.

If $d \geq 6$,

$$(6-4d)\lambda^2 - 2d + 6 \leq -2d + 6 < 0, \qquad d\lambda^2 - 2d + 6 < -d + 6 \leq 0$$

so that the coefficients in $B_{d,\lambda}$ for even powers of t are non-positive and the ones for odd powers of t are positive. Hence for $t \leq 0$, $B_{d,\lambda}(t) \leq B_{d,\lambda}(0) = d\lambda^2 - 2d + 6 < 0$, so that $B_{d,\lambda}$ has no zeros on $(-\infty, 0]$. Similarly, $B'_{d,\lambda}$ can have no zeros on $(-\infty, 0]$, so that in particular, $B_{d,\lambda}$ is always increasing on $[-1, 0]$. With dots indicating terms at most quadratic in s,

$$\begin{aligned}
B_{d,\lambda}(1-s) &= (6-d)\lambda^2(1-s)^4 + 4(d-3)\lambda(1-s)^3 + \cdots \\
&= (6-d)\lambda^2(s^4 - 4s^3 + \cdots) + 4(d-3)\lambda(-s^3 + \cdots) + \cdots \\
&= (6-d)\lambda^2 s^4 + (4(d-6)\lambda^2 - 4(d-3)\lambda)s^3 + \cdots \\
&= (6-d)\lambda^2 s^4 + 4\lambda((d-6)\lambda - d + 3)s^3 + \cdots .
\end{aligned}$$

If $d \geq 6$ and $\lambda \leq 1$,

$$(d-6)\lambda - d + 3 \leq (d-6) - d + 3 = -3.$$

Thus, if $d \geq 6$ and $0 < \lambda < 1$, the coefficients of s^4, s^3 and s^0 in $B_{d,\lambda}(1-s)$ are non-positive, negative and (as $B_{d,\lambda}(1) < 0$) negative, so that the number of sign changes in the non-zero coefficients of $B_{d,\lambda}(1-s)$ is at most two. By exact investigation of the $B_{d,\lambda}(1-s)$ polynomial, it is possible to show that for any feasible λ, at least one of the quadratic and linear coefficient is positive. By Descartes' rule of signs, $B_{d,\lambda}(1-s)$ has two non-negative roots, or equivalently, $B_{d,\lambda}(t)$ can have at most two roots in $(-\infty, 1)$. Furthermore, since $B_{d,\lambda}(1) < 0$, $B_{d,\lambda}(1/\lambda) = d\lambda^2(1-(1/\lambda)^2)^{>}0$ and $\lim_{t\to\infty} B'_{d,\lambda}(t) = -\infty$, the remaining two roots are in $(1, 1/\lambda)$ and $(1/\lambda, \infty)$. Thus again, $B_{d,\lambda}$ has 2 roots in I and at most two real roots in I (and if it has, these must be positive). Similarly, we can infer that $B'_{d,\lambda}$ can at most have two zeros in I (and again, if it has these must be positive). ☐

References

Botts, C., Hörmann, W. and Leydold, J. (2013). Transformed density rejection with inflection points. *Statistics and Computing* **23** 251–260. MR3016942

Deak, I. (1979). Comparison of methods for generating uniformly distributed random points in and on a hypersphere. *Problems of Control and Information Theory* **8** 105–113. MR0533925

Devroye, L. (1986). *Non-Uniform Random Variate Generation*. Springer-Verlag, New-York. MR0836973

DLMF *NIST Digital Library of Mathematical Functions*. http://dlmf.nist.gov/, Release 1.0.19 of 2018-06-22. F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller and B. V. Saunders, eds.

Durrett, R. (1984). *Brownian Motion and Martingales in Analysis*. Wadsworth Advanced Books & Software California. MR0750829

Eddelbuettel, D. and François, R. (2011). Rcpp: Seamless R and C++ integration. *Journal of Statistical Software* **40** 1–18.

Gilks, W. R. and Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **41** 337–348.

Golzy, M. and Markatou, M. (2020). Poisson kernel-based clustering on the sphere: Convergence properties, identifiability, and a method of sampling. *Journal of Computational and Graphical Statistics* **0** 1–13. MR4191241

Hörmann, W., Leydold, J. and Derflinger, G. (2004). *Automatic Nonuniform Random Variate Generation*. Springer-Verlag, Berlin Heidelberg. MR2143197

Hornik, K. and Gr ün, B. (2014). movMF: An R package for fitting mixtures of von Mises-Fisher distributions. *Journal of Statistical Software, Articles* **58** 1–31.

Hornik, K. and Grün, B. (2013). Amos-type bounds for modified Bessel function ratios. *Journal of Mathematical Analysis and Applications* **408** 91–101. MR3079949

Kato, S. and Jones, M. (2013). An extended family of circular distributions related to wrapped Cauchy distributions via Brownian motion. *Bernoulli* **19** 154–171. MR3019490

Kato, S. and McCullagh, P. (2020). Some properties of a Cauchy family on the sphere derived from the Möbius transformations. *Bernoulli* **26**. MR4140543

Kent, J. T., Ganeiber, A. M. and Mardia, K. V. (2018). A new unified approach for the simulation of a wide class of directional distributions. *Journal of Computational and Graphical Statistics* **27** 291–301. MR3816265

Leydold, J., Botts, C. and Hörmann, W. (2019). Tinflex: A Universal Non-Uniform Random Number Generator R package version 1.5.

Mardia, K. V. and Jupp, P. E. (2009). *Directional Statistics* **494**. John Wiley & Sons. MR1828667

McLachlan, G. J. (1987). On bootstrapping the likelihood ratio test stastistic for the number of components in a normal mixture. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **36** 318–324.

O'Hagan, A., Murphy, T. B., Scrucca, L. and Gormley, I. C. (2019). Investigation of parameter uncertainty in clustering using a Gaussian mixture model via jackknife, bootstrap and weighted likelihood bootstrap. *Computational Statistics* **34** 1779–1813. MR4030442

Saw, J. G. (1978). A family of distributions on the m-sphere and some hypothesis tests. *Biometrika* **65** 69–73. MR0497510

R Core Team (2018). R: A Language and Environment for Statistical Computing R Foundation for Statistical Computing, Vienna, Austria.

Ulrich, G. (1984). Computer generation of distributions on the m-sphere. *Journal of the Royal Statistical Society Series C (Applied Statistics)* **33** 158–163. MR0767583