Random Sampling from the Watson Distribution

Sablica, Lukas; Hornik, Kurt; Leydold, Josef

DOI:
10.57938/6e5ffebc-7ad2-4225-9909-d8833529e808

Published: *01/01/2022*

Document Version:
Publisher's PDF, also known as Version of record

Random Sampling from the Watson Distribution

A Preprint

Lukas Sablica
Institute for Statistics and Mathematics
Vienna University of Economics and Business
Austria
https://www.wu.ac.at/en/statmath
ORCiD: 0000-0001-9166-4563
Lukas.Sablica@wu.ac.at

Kurt Hornik
Institute for Statistics and Mathematics
Vienna University of Economics and Business
Austria
https://www.wu.ac.at/en/statmath
ORCiD: 0000-0003-4198-9911
Kurt.Hornik@wu.ac.at

Josef Leydold
Institute for Statistics and Mathematics
Vienna University of Economics and Business
Austria
https://www.wu.ac.at/en/statmath
Jozef.Leydold@wu.ac.at

November 29, 2021

Abstract

In this paper we present and discuss two methods for efficient sampling from the Watson distribution. The first approach adapts the rejection sampling algorithm from Kent et al. (2018), which originally samples Bingham distribution using angular central Gaussian envelopes. We show that for the case of the Watson distribution, this allows for a closed form expression for the parameters that maximize the efficiency of the sampling procedure, which is then further investigated and bounded by derived asymptotic results. What is more, we present a sampling algorithm that removes the curse of dimensionality by a smart matrix inversion, which allows for fast runtimes even for complex problems with high dimension. The second method relates to Saw (1978), and simulates from a projected distribution using adaptive rejection sampling. Also for this sampling procedure, the derived algorithm offers fast sampling for large dimension. This is not the case for similar algorithms in the field, which usually require an expensive rotation of the sampled results using a QR-decomposition. Finally, we propose some simple generators for the trivial cases and compare the two main methods in a simulation study.

Keywords Sampling · Rejection sampling · Simulation · Watson distribution · Spherical distributions

1 Introduction

Directional data analysis has attracted a fair amount of attention over the past years. However, the topic of sampling has been mostly overshadowed by the estimation problems that these distributions bring. These problems have been recently answered for many distributions (see e.g. Hornik and Grün (2014) and Sra and Karp (2013)) making these distributions more accessible to the data modeling community, particularly in the field of mixture modeling.

This motivates also a need for an efficient simulation method to accompany these estimation results and to help to identify the uncertainty that comes with such an estimation. For example in the context of mixture modeling, the parametric bootstrap has been proposed to be used for the likelihood ratio test to assess the number of components and also to obtain standard errors for the parameters, see e.g. McLachlan (1987) and O'Hagan et al. (2019). Another example is Bayesian inference, where sampling from a directional distribution can be used to sample from a proposal distribution of unitary directional parameters in a Metropolis–Hastings algorithm or for sampling from a predictive distribution. Altogether this clearly shows the importance of fast sampling algorithms for directional distributions.

The Watson distribution (Watson, 1965) is used in the modeling of axially symmetric data (i.e., x and $-x$ are indistinguishable) which frequently occur for example in structural geology or rock magnetism areas. Its application in mixture modeling was discussed for example in Bijral et al. (2007) and Sra and Karp (2013).

Outside of the case where the Watson distribution simplifies to uniform distribution on a sphere, the current literature that considers the Watson distribution directly is, according to our best knowledge, equipped only with the cases where dimension is at most equal to 3, for example see Best and Fisher (1986) or Li and Wong (1993). Interestingly however, a slightly more general answer to the question was published by Kent et al. (2018), where the authors derived a BACG rejection sampling method for the Bingham distribution which generalizes the Watson one. As will be shown below, directly applying this theory to the Watson distribution not only provides a sampling algorithm, but also allows to calculate the optimal hyper-parameters analytically, which is not the case for the general Bingham distribution. Furthermore, the curse of dimensionality can be overcome and the numerical complexity of the matrix operations in the presented algorithm scales linearly in the dimension.

2 The Watson distribution

Following the notation in Mardia and Jupp (1999), let $\mathbb{S}^{d-1} = \{x \in \mathbb{R}^d, \|x\| = 1\}$ be a $(d-1)$-dimensional unit sphere with the projective hyperplane $\mathbb{P}^{d-1}$ such that the symmetric vectors $\pm x \in \mathbb{S}^{d-1}$ are equivalent. A random unit length vector in $\mathbb{R}^d$ has a Watson distribution with concentration parameter $\kappa \in \mathbb{R}$ and mean direction parameter $\mu \in \mathbb{S}^{d-1}$, if its density with respect to the uniform distribution on the unit sphere is

$$f(x|\kappa, \mu) = M\left(\frac{1}{2}, \frac{p}{2}, \kappa\right)^{-1} e^{\kappa (x'\mu)^2},$$

where

$$M(a, b, z) = {}_1F_1(a; b; z) = \sum_{n=0}^{\infty} \frac{(a)_n}{(b)_n} \frac{z^n}{n!}$$

is the confluent hypergeometric function of the first kind, also known as Kummer's function, solving the Kummer's differential equation $zw'' + (b - z)w' - aw = 0$.

We note that if $\kappa > 0$ the density has maxima at $\pm\mu$ and so is bipolar. On the other hand, for $\kappa < 0$ the distribution is concentrated around the great circle orthogonal to μ, and so the distribution is a symmetric girdle distribution. Finally, observe that for orthogonal Q s.t. $Q\mu = \mu$, it holds that $\mu'(Qx) = \mu' x$, implying the rotational symmetry about μ. For more details, see Mardia and Jupp (1999).

3 Simulating with an angular central Gaussian envelope

Kent et al. (2018) introduce a rejection algorithm for sampling from the Bingham distribution using angular central Gaussian (ACG) envelopes. With υ_d the uniform distribution on $\mathbb{S}^{d-1}$, the densities with respect to υ_d of the Bingham distribution with parameter A and the ACG with parameter Ω, are, respectively, proportional to $\exp(-x'Ax)$ and given by $f_{\mathrm{ACG}(\Omega)}(x) = |\Omega|^{1/2}(x'\Omega x)^{-d/2}$.

The basic idea is the following. Write $\lambda_1 \leq \cdots \leq \lambda_d$ for the eigenvalues of A, $A_0 = A - \lambda_1 I$, and consider $\Omega(b) = I + 2A_0/(b + 2\lambda_1)$ for $b > -2\lambda_1$. Let $x \in \mathbb{S}^{d-1}$. Then $x'\Omega(b)x = 1 + 2x'A_0x/(b + 2\lambda_1)$ and hence, using the elementary inequality $\log(u) \leq u - 1$ valid for $u > 0$,

$$\begin{aligned} d\log(x'\Omega(b)x) &= d\log\frac{d}{b+2\lambda_1} + d\log\frac{b+2\lambda_1+2x'A_0x}{d} \\ &\leq d\log\frac{d}{b+2\lambda_1} + d\left(\frac{b+2\lambda_1+2x'A_0x}{d} - 1\right) \\ &= d\log\frac{d}{b+2\lambda_1} + b + 2\lambda_1 + 2x'A_0x - d, \end{aligned}$$

so that

$$\begin{aligned}\frac{e^{-x'Ax}}{f_{\text{ACG}(\Omega(b))}(x)} &= |\Omega(b)|^{-1/2}\exp\left(-\lambda_1 - x'A_0x + \frac{d}{2}\log(x'\Omega(b)x)\right)\\ &\leq \exp\left(-\frac{1}{2}\sum_{i=1}^{d}\log\frac{b+2\lambda_i}{b+2\lambda_1} + \frac{d}{2}\log\frac{d}{b+2\lambda_1} + \frac{b-d}{2}\right)\\ &= (d/e)^{d/2}\exp\left(\frac{b}{2} - \frac{1}{2}\sum_{i=1}^{d}\log(b+2\lambda_i)\right).\end{aligned}$$

The first two derivatives of $h(b) = b - \sum_i \log(b+2\lambda_i)$ are given by $1 - \sum_i 1/(b+2\lambda_i)$ and $\sum_i 1/(b+2\lambda_i)^2 > 0$. Hence, as $\sum_i 1/(b+2\lambda_i)$ decreases from ∞ to 0 as b increases from $-2\lambda_1$ to ∞, h has a unique minimum on $(-2\lambda_1, \infty)$ attained at the unique solution of the equation $\sum_i 1/(b+2\lambda_i) = 1$.

The Watson distribution with concentration parameter κ and direction parameter $\mu \in \mathbb{S}^{d-1}$ has density with respect to υ_d given by

$$\frac{e^{\kappa(\mu'x)^2}}{M(1/2, d/2, \kappa)}$$

and thus is a special case of the Bingham distribution with $A = -\kappa\mu\mu'$, which has eigenvalues $-\kappa$ and 0 with respective multiplicities 1 and $d-1$ so that $\lambda_1 = \min(-\kappa, 0) = -\kappa_+$. Using the above, we have

$$\frac{e^{\kappa(\mu'x)^2}}{f_{\text{ACG}(\Omega(b))}(x)} \leq (d/e)^{d/2}\exp\left(\frac{b}{2} - \frac{1}{2}\left(\log(b-2\kappa) + (d-1)\log(b)\right)\right)$$

which is minimized for $b = b(\kappa)$ the unique solution of

$$\frac{1}{b-2\kappa} + \frac{d-1}{b} = 1$$

on $(2\kappa_+, \infty)$. The equation can be rewritten as

$$\begin{aligned}0 &= b(b-2\kappa) - (b + (d-1)(b-2\kappa))\\ &= b^2 - (2\kappa+d)b + 2\kappa(d-1)\end{aligned}$$

so that the unique solution on $(2\kappa_+, \infty)$ must be given by

$$b(\kappa) = (\kappa + d/2) + \sqrt{(\kappa+d/2)^2 - 2\kappa(d-1)}. \tag{1}$$

As an aside, we note that the discriminant is

$$\kappa^2 + d\kappa + d^2/4 - 2d\kappa + 2\kappa = \kappa^2 + (2-d)\kappa + d^2/4$$

which as a function of κ has roots

$$d/2 - 1 \pm \sqrt{(d/2-1)^2 - d^2/4} = 1 - d/2 \pm \sqrt{1-d},$$

so that for $d > 1$ the discriminant is indeed positive.

Alternatively, we could have proceeded "directly" as in Kent et al. (2018). If $\kappa \leq 0$, we have $\lambda_1 = \cdots = \lambda_{d-1} = 0$ and $\lambda_d = -\kappa$, so that $\Omega(b)$ has eigenvalues 1 ($d-1$ times) and $1 - 2\kappa/b$, and the optimal bound is

$$\begin{aligned}&e^{-(d-b)/2}(d/b)^{d/2}(1-2\kappa/b)^{-1/2}\\ &= (d/e)^{d/2}\exp\left(\frac{b}{2} - \frac{1}{2}(\log(1-2\kappa/b) + d\log(b))\right)\\ &= (d/e)^{d/2}\exp\left(\frac{b}{2} - \frac{1}{2}(\log(b-2\kappa) + (d-1)\log(b))\right),\end{aligned}$$

where b solves $(d-1)/b + 1/(b-2\kappa) = 1$ on $(0, \infty)$. If $\kappa \geq 0$, we have $\lambda_1 = -\kappa$ and $\lambda_2 = \cdots = \lambda_d = 0$ and hence need to consider $A_0 = A + \kappa I$ instead, which has eigenvalues $\lambda_1 = 0$ and $\lambda_2 = \cdots = \lambda_d = \kappa$, so that $\Omega(b)$ has eigenvalues 1 and $1 + 2\kappa/b$ ($d-1$ times), and the optimal bound is

$$\begin{aligned}&e^{\kappa}e^{-(d-b)/2}(d/b)^{d/2}(1+2\kappa/b)^{-(d-1)/2}\\ &= (d/e)^{d/2}\exp\left(\kappa + \frac{b}{2} - \frac{1}{2}(d\log(b) + (d-1)\log(1+2\kappa/b))\right)\\ &= (d/e)^{d/2}\exp\left(\frac{b+2\kappa}{2} - \frac{1}{2}(\log(b) + (d-1)\log(b+2\kappa)),\right)\end{aligned}$$

where b solves $1/b + (d-1)/(b+2\kappa) = 1$ on $(0, \infty)$. Replacing b by $b - 2\kappa$, the bound becomes

$$(d/e)^{d/2} \exp\left(\frac{b}{2} - \frac{1}{2}(\log(b-2\kappa) + (d-1)\log(b))\right),$$

where b solves $1/(b-2\kappa) + (d-1)/b = 1$ on $(-2\kappa, \infty)$.

Write $\mathcal{M}_d(\kappa)$ for the optimal rejection constant achieved for sampling from the Watson distribution with concentration parameter κ with ACG envelopes, i.e.,

$$\mathcal{M}_d(\kappa) = \frac{(d/e)^{d/2}}{M(1/2, d/2, \kappa)} \exp\left(\frac{b(\kappa)}{2} - \frac{1}{2}(\log(b(\kappa)-2\kappa) + (d-1)\log(b(\kappa)))\right). \tag{2}$$

The rejection 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).

Theorem 1. *The rejection constant $\mathcal{M}_d(\kappa)$ is decreasing and increasing on $(-\infty, 0)$ and $(0, \infty)$ respectively. Hence in particular,*

$$1 \le M_d(\kappa) \le M_d(-\infty) \text{ for all } \kappa \le 0,$$
$$1 \le M_d(\kappa) \le M_d(\infty) \text{ for all } \kappa \ge 0.$$

Proof. Write

$$\log(\mathcal{M}_d(\kappa)) = \frac{d}{2}\log\left(\frac{d}{e}\right) + \left(\frac{b(\kappa)}{2} - \frac{1}{2}\log(b(\kappa)-2\kappa) - \frac{1}{2}(d-1)\log(b(\kappa))\right) - \log\left(M\left(\frac{1}{2}, \frac{d}{2}, \kappa\right)\right)$$

for the logarithm of the rejection constant. The derivative is then equal to

$$\begin{aligned} \frac{\mathrm{d}\log(\mathcal{M}_d(\kappa))}{\mathrm{d}\kappa} &= \frac{b'(\kappa)}{2} - \frac{1}{2}\frac{b'(\kappa)-2}{b(\kappa)-2\kappa} - \frac{1}{2}(d-1)\frac{b'(\kappa)}{b(\kappa)} - \tilde{g}\left(\frac{1}{2}, \frac{d}{2}, \kappa\right) \\ &= \frac{b'(\kappa)}{2}\left(1 - \frac{1}{b(\kappa)-2\kappa} - \frac{(d-1)}{b(\kappa)}\right) + \frac{1}{b(\kappa)-2\kappa} - \tilde{g}\left(\frac{1}{2}, \frac{d}{2}, \kappa\right), \end{aligned}$$

where

$$\tilde{g}(a, b, z) = \frac{\mathrm{d}\log(M(a, b, z))}{\mathrm{d}z} = \frac{a}{b}\frac{M(a+1, b+1, z)}{M(a, b, z)}.$$

The terms in the bracket can be simplified as

$$\begin{aligned} 1 - \frac{1}{b(\kappa)-2\kappa} - \frac{(d-1)}{b(\kappa)} &= 1 - \frac{b(\kappa) - (d-1)(b(\kappa)-2\kappa)}{(b(\kappa)-2\kappa)\,b(\kappa)} \\ &= 1 - \frac{\frac{d^2}{2} + 2\kappa + d\left(-\kappa + \sqrt{\kappa^2 + (2-d)\kappa + d^2/4}\right)}{\frac{d^2}{2} + 2\kappa + d\left(-\kappa + \sqrt{\kappa^2 + (2-d)\kappa + d^2/4}\right)} = 1 - 1 = 0, \end{aligned}$$

and hence

$$\frac{\mathrm{d}\log(\mathcal{M}_d(\kappa))}{\mathrm{d}\kappa} = \frac{1}{b(\kappa)-2\kappa} - \tilde{g}\left(\frac{1}{2}, \frac{d}{2}, \kappa\right).$$

Sablica and Hornik (electronically published on November 5, 2021) define a bound $l_{a,b}^{(0)}(z)$ for the function $g(a, b, z)$ given by the form

$$l_{a,b}^{(0)}(z) = \frac{2a}{\sqrt{z^2 + 2(2a-b)z + b^2} - z + b}.$$

The bound satisfies

$$l_{a,b}^{(0)}(z) \begin{cases} < \tilde{g}(a, b, z) & \text{if} \quad z < 0, \\ = \tilde{g}(a, b, z) & \text{if} \quad z = 0, \\ > \tilde{g}(a, b, z) & \text{if} \quad z > 0, \end{cases}$$

i.e., is a lower bound on the negative reals and an upper bound on the positive ones. Thus plugging in the values of interest gives

$$l_{\frac{1}{2},\frac{d}{2}}^{(0)}(\kappa) = \frac{1}{\sqrt{\kappa^2 + (2-d)\kappa + \frac{d^2}{4}} - \kappa + \frac{d}{2}} = \frac{1}{b(\kappa)-2\kappa}$$

so that for $\kappa < 0$

$$\frac{\mathrm{d}\log(\mathcal{M}_d(\kappa))}{\mathrm{d}\kappa} = \frac{1}{b(\kappa)-2\kappa} - \tilde{g}\left(\frac{1}{2},\frac{d}{2},\kappa\right) < \frac{1}{b(\kappa)-2\kappa} - l^{(0)}_{\frac{1}{2},\frac{d}{2}}(\kappa) = 0,$$

and for $\kappa > 0$

$$\frac{\mathrm{d}\log(\mathcal{M}_d(\kappa))}{\mathrm{d}\kappa} = \frac{1}{b(\kappa)-2\kappa} - \tilde{g}\left(\frac{1}{2},\frac{d}{2},\kappa\right) > \frac{1}{b(\kappa)-2\kappa} - l^{(0)}_{\frac{1}{2},\frac{d}{2}}(\kappa) = 0.$$

□

Lemma 1. *Let $b(\kappa)$ be the unique solution of*

$$\frac{1}{b-2\kappa} + \frac{d-1}{b} = 1$$

on $(2\kappa_+,\infty)$. Then as $\kappa\to-\infty$, $b(\kappa)\to d-1$, and for $\kappa\to\infty$, $b(\kappa)=2\kappa+1+o(1)$.

Proof. Write

$$F(b,\kappa) = \frac{1}{b-2\kappa} + \frac{d-1}{b} - 1$$

so that $F(b(\kappa),\kappa)=0$. Clearly,

$$F_b(b,\kappa) = \frac{\partial F}{\partial b}(b,\kappa) = -\frac{1}{(b-2\kappa)^2} - \frac{d-1}{b^2} < 0$$

and

$$F_\kappa(b,\kappa) = \frac{\partial F}{\partial \kappa}(b,\kappa) = \frac{2}{(b-2\kappa)^2} > 0.$$

By the implicit function theorem,

$$b'(\kappa) = -\frac{F_\kappa(b(\kappa),\kappa)}{F_b(b(\kappa),\kappa)} > 0.$$

Clearly, $b(0)=d$. For $|\kappa|\to\infty$,

$$\begin{aligned} b(\kappa) &= \kappa + d/2 + |\kappa|\sqrt{1+\frac{2-d}{\kappa}+\frac{d^2}{4\kappa^2}} \\ &= \kappa + d/2 + |\kappa|\left(1+\frac{2-d}{2\kappa}+o(1/\kappa)\right) \\ &= \kappa + |\kappa| + d/2 + \operatorname{sgn}(\kappa)(1-d/2) + o(1). \end{aligned}$$

Hence for $\kappa\to-\infty$, $b(\kappa)\to d-1$, and for $\kappa\to\infty$, $b(\kappa)=2\kappa+1+o(1)$. □

Theorem 2. *Let $\mathcal{M}_d(\kappa)$ be the rejection constant* (2)*. Then $\mathcal{M}_d(0)=1$,*

$$M_d(-\infty) = \lim_{\kappa\to-\infty}\mathcal{M}_d(\kappa) = \frac{1}{\sqrt{2e}}\frac{\Gamma((d-1)/2)}{(d-1)^{(d-1)/2}}\frac{d^{d/2}}{\Gamma(d/2)},$$

$$M_d(\infty) = \lim_{\kappa\to\infty}\mathcal{M}_d(\kappa) = \sqrt{2\pi e}\frac{(d/2)^{d/2}e^{-d/2}}{\Gamma(d/2)},$$

and as $d\to\infty$

$$M_d(-\infty) \approx \sqrt{\frac{d}{d-1}}, \qquad \textit{and} \qquad M_d(\infty) \approx \sqrt{\frac{de}{2}}.$$

Proof. For $\kappa=0$ we have $b(0)=d$ so that with $M(1/2,d/2,0)=1$,

$$\mathcal{M}_d(0) = (d/e)^{d/2}\exp\left(\frac{d}{2}-\frac{1}{2}(\log(d)+(d-1)\log(d))\right) = 1.$$

Using (DLMF, Eq. 13.2.39, Eq. 13.7.1) for $\kappa \to -\infty$ (and d fixed),

$$\begin{aligned} M(1/2, d/2, \kappa) &= e^{\kappa} M((d-1)/2, d/2, -\kappa) \\ &\approx e^{\kappa} \frac{\Gamma(d/2)}{\Gamma((d-1)/2)} e^{-\kappa} (-\kappa)^{-1/2} \\ &= \frac{\Gamma(d/2)}{\Gamma((d-1)/2)} |\kappa|^{-1/2} \end{aligned}$$

so that as $b(\kappa) \to d-1$,

$$\begin{aligned} \mathcal{M}_d(\kappa) &\approx \frac{\Gamma((d-1)/2)}{\Gamma(d/2)} |\kappa|^{1/2} (d/e)^{d/2} \\ &\quad \times \exp\left(\frac{d-1}{2} - \frac{1}{2}(\log(2|\kappa| + d - 1) + (d-1)\log(d-1))\right) \\ &\to \frac{\Gamma((d-1)/2)}{\Gamma(d/2)} (d/e)^{d/2} \exp\left(\frac{d-1}{2} - \frac{\log(2)}{2} - \frac{d-1}{2}\log(d-1)\right) \\ &= \frac{\Gamma((d-1)/2)}{\Gamma(d/2)} \frac{d^{d/2}}{(d-1)^{(d-1)/2}} \exp\left(-\frac{\log(2e)}{2}\right) \\ &= \frac{1}{\sqrt{2e}} \frac{\Gamma((d-1)/2)}{(d-1)^{(d-1)/2}} \frac{d^{d/2}}{\Gamma(d/2)}. \end{aligned}$$

I.e.,

$$\lim_{\kappa \to -\infty} \mathcal{M}_d(\kappa) = \mathcal{M}_d(-\infty) = \frac{1}{\sqrt{2e}} \frac{\Gamma((d-1)/2)}{(d-1)^{(d-1)/2}} \frac{d^{d/2}}{\Gamma(d/2)}.$$

Using (DLMF, Eq. 15.11.1),

$$\log(\Gamma(x)) \approx (x - 1/2)\log(x) - x + \log(2\pi)/2,$$

so that for large d,

$$\begin{aligned} \log(\mathcal{M}_d(-\infty)) &\approx -\frac{\log(2e)}{2} + \frac{d-2}{2}\log\frac{d-1}{2} - \frac{d-1}{2} - \frac{d-1}{2}\log(d-1) \\ &\quad -\frac{d-1}{2}\log\frac{d}{2} + \frac{d}{2} + \frac{d}{2}\log(d) \\ &= -\frac{1}{2}\log(d-1) + \frac{1}{2}\log(d) + \log(2)\left(-\frac{1}{2} - \frac{d-2}{2} + \frac{d-1}{2}\right) \\ &= \frac{1}{2}\log\frac{d}{d-1} \end{aligned}$$

so that as $d \to \infty$,

$$\mathcal{M}_d(-\infty) \approx \sqrt{\frac{d}{d-1}} \to 1.$$

Using (DLMF, Eq. 13.7.1), for $\kappa \to \infty$ (and d fixed)

$$M(1/2, d/2, \kappa) \approx \frac{\Gamma(d/2)}{\Gamma(1/2)} e^{\kappa} \kappa^{(1-d)/2}$$

so that as $b(\kappa) = 2\kappa + 1 + o(1)$,

$$\begin{aligned} \mathcal{M}_d(\kappa) &\approx \frac{\Gamma(1/2)}{\Gamma(d/2)} (d/e)^{d/2} e^{-\kappa} \kappa^{(d-1)/2} \exp\left(\kappa + \frac{1}{2} - \frac{d-1}{2}\log(2\kappa+1)\right) \\ &= \frac{\Gamma(1/2)}{\Gamma(d/2)} (d/e)^{d/2} \exp\left(\frac{1}{2} - \frac{d-1}{2}\log\frac{2\kappa+1}{\kappa}\right) \\ &\to \frac{\Gamma(1/2)}{\Gamma(d/2)} (d/e)^{d/2} \exp\left(\frac{1}{2} - \frac{d-1}{2}\log(2)\right) \\ &= \sqrt{2\pi e}\frac{(d/2)^{d/2} e^{-d/2}}{\Gamma(d/2)}. \end{aligned}$$

I.e.,

$$\lim_{\kappa\to\infty}\mathcal{M}_d(\kappa)=\mathcal{M}_d(\infty)=\sqrt{2\pi e}\frac{(d/2)^{d/2}e^{-d/2}}{\Gamma(d/2)}.$$

Expanding this at $d=\infty$ gives

$$\mathcal{M}_d(\infty)=\sqrt{\frac{de}{2}}-\frac{1}{6}\sqrt{\frac{e}{2}}\frac{1}{\sqrt{d}}+\frac{1}{72}\sqrt{\frac{e}{2}}\left(\frac{1}{d}\right)^{3/2}+O\left(\left(\frac{1}{d}\right)^{5/2}\right)$$

or again using (DLMF, Eq. 5.11.1), for large d

$$\begin{aligned}\log(\mathcal{M}_d(\infty)) &\approx \frac{\log(2\pi e)}{2}+\frac{d}{2}\log\frac{d}{2}-\frac{d}{2}-\frac{d-1}{2}\log\frac{d}{2}+\frac{d}{2}-\frac{\log(2\pi)}{2}\\ &= \frac{1}{2}+\frac{1}{2}\log\frac{d}{2}\end{aligned}$$

so that as $d\to\infty$,

$$\mathcal{M}_d(\infty)\approx\sqrt{\frac{de}{2}}.$$

□

Let $f^*_{\mathrm{ACG}(\Omega)}(x)=(x'\Omega x)^{-d/2}$ be the unnormalized density of angular central Gaussian distribution. From the previous calculations it easily follows that

$$\frac{e^{-x'Ax}}{f^*_{\mathrm{ACG}(\Omega(b))}(x)}\le\exp\left(\frac{b-d}{2}\right)\left(\frac{d}{b+2\lambda_1}\right)^{d/2},$$

with the optimal $b=b(\kappa)$ given by (1). Recall that if $Y\sim N_p(0,\Sigma)$ then $Y/\|Y\|\sim ACG(\Omega)$ with $\Omega=\Sigma^{-1}$. Since $A=-\kappa\mu\mu'$ and Ω is of the form

$$\Omega=I+\frac{2}{b-2\kappa_+}(-\kappa\mu\mu'+\kappa_+I)=\frac{b}{b-2\kappa_+}\left(I-\frac{2\kappa}{b}\mu\mu'\right),$$

we can rewrite the ratio as

$$\frac{e^{\kappa(\mu'x)^2}}{\left(\frac{b-2\kappa(\mu'x)^2}{d}\right)^{-\frac{d}{2}}}\le\exp\left(\frac{b-d}{2}\right),$$

which is for a candidate vector x essentially a function of $(\mu'x)^2$.

To sample from ACG distribution, note that $X\sim\mathrm{ACG}(\Omega)$ has the same distribution as $Y\sim\mathrm{ACG}(c\Omega)$ for some $c>0$. Hence without any loss of generality we can consider only $\Omega=I-\frac{2\kappa}{b}\mu\mu'$.

Proposition 1. *Let Ω be a rank-1 updated identity matrix of the form $\Omega=I-\beta_0\mu\mu'$ with μ of length 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 $\beta_2=-1+1/\sqrt{1-\beta_0}$.

Proof. Trivial. □

Thus the candidate draw x from the appropriate ACG distribution, is just a normalized $y=\Omega^{-1/2}z$, where $z\sim N(0,I)$ and $\Omega^{-1/2}=I+\beta_2\mu\mu'$ with $\beta_2=-1+\sqrt{b/b-2\kappa}$. The essential dot-product is then given by

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

Putting everything together we end up with Algorithm 1, which requires per iteration $O(d)$ computations and by the Theorem 2 needs in average at worst $O(\sqrt{d})$ such iterations.

Algorithm 1 Generator for Watson distribution with parameters κ and μ using ACG envelope

```
b ← b(κ)                                                        ▷ eq. 1
β₂ ← −1 + sqrt(b/(b−2κ))
β₁ ← β₂ (β₂ + 2)
repeat                                                          ▷ acceptance-rejection loop
    Sample U ~ U(0,1)
    Sample Z_i ~ N(0,1) for i = 1,2,...,d
    z ← (Z_1, Z_2, ..., Z_d)
    κ(μ'x)² ← κ((μ'z)² + β₁(μ'z)²)/(z'z + β₁(μ'z)²)
until log(U) ≤ κ(μ'x)² + (d/2) log((b−2κ(μ'x)²)/d) − (b−d)/2
return x ← (z + β₂μ'zμ)/sqrt(z'z + β₁(μ'z)²)
```

$b \leftarrow b(\kappa)$ ▷ eq. 1

$\beta_2 \leftarrow -1 + \sqrt{\frac{b}{b-2\kappa}}$

$\beta_1 \leftarrow \beta_2\,(\beta_2 + 2)$

repeat ▷ acceptance-rejection loop

  Sample $U \sim U(0,1)$

  Sample $Z_i \sim N(0,1)$ for $i = 1, 2, \ldots, d$

  $z \leftarrow (Z_1, Z_2, \ldots, Z_d)$

  $\kappa(\mu' x)^2 \leftarrow \kappa((\mu' z)^2 + \beta_1(\mu' z)^2)/(z'z + \beta_1(\mu' z)^2)$

until $\log(U) \leq \kappa(\mu' x)^2 + \frac{d}{2}\log\left(\frac{b-2\kappa(\mu' x)^2}{d}\right) - \frac{b-d}{2}$

return $x \leftarrow (z + \beta_2 \mu' z \mu)/\sqrt{z'z + \beta_1(\mu' z)^2}$

4 Simulating with adaptive rejection sampling from the projected Saw distribution

Let $\|\cdot\|$ denote the Euclidean norm. For $d \geq 2$, Saw (1978) introduced the family of distributions on $\mathbb{S}^{d-1}$ with densities with respect to the uniform distribution υ_d on $\mathbb{S}^{d-1}$ given by

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

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

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

$\lambda \geq 0$ is a concentration parameter, $u \in \mathbb{S}^{d-1}$ is a direction parameter, and g is a function from $\mathbb{R}$ to $[0, \infty)$ controlling the shape of the distribution. These distributions, which we will denote by $\mathrm{Saw}_d(g, u, \lambda)$, are well-defined provided that $0 < c_{g;\lambda,\gamma} < \infty$.

If $\lambda > 0$ and g is positive and increasing, $\mathrm{Saw}_d(g, u, \lambda)$ has a unique mode at the direction parameter u. This includes the case where $g = \exp$, giving the well-known von Mises-Fisher distribution (e.g., Mardia and Jupp, 1999, p. 168). If $\lambda > 0$ and g is symmetric about zero, positive and increasing on the positive reals, $\mathrm{Saw}_d(g, u, \lambda)$ has its modes at $\pm u$. One example for this case is $g(t) = \exp(t^2)$, corresponding to the Watson distribution (e.g., Mardia and Jupp, 1999, p. 181, using $\sqrt{\kappa} = \lambda$ as concentration parameter). Note that the distribution for negative κ can be obtained with $g(t) = \exp(-t^2)$ and $\sqrt{-\kappa} = \lambda$.

Theorem 3 (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 use $X' = \left(\sqrt{1-W^2}Y', W\right)$, with $W \sim \mathrm{pSaw}_d(g, \lambda)$ drawn independently from $Y \sim \upsilon_{d-1}$ (which can most conveniently be accomplished by $Y = Z/\|Z\|$ with the elements of Z i.i.d. standard normal).

Rewriting the problem to the Watson distribution, to sample from the projected Saw distribution $\mathrm{pSaw}_d(\exp(x^2), \sqrt{\kappa})$ and $\mathrm{pSaw}_d(\exp(-x^2), \sqrt{-\kappa})$ we want to sample from a density proportional to

$$f_{d,\kappa}(x) = \begin{cases} \exp\left(\kappa x^2\right)\left(1-x^2\right)^{(d-3)/2} & \text{for } -1 < x < 1, \\ 0 & \text{otherwise,} \end{cases} \qquad \text{for } d \geq 2 \text{ and } \kappa \in \mathbb{R}.$$

In our application κ can be quite large so that $\exp\left(\kappa x^2\right)$ might overflow when implemented in floating point arithmetic. Due to the symmetry of $f_{d,\kappa}(x)$ we restrict our analysis to domain $x \in [0, 1)$. The log-density on [0,1) is then (up to an additive constant)

$$l_{d,\kappa}(x) = \begin{cases} \kappa x^2 + \frac{d-3}{2}\log\left(1-x^2\right) & \text{for } 0 \le x < 1, \\ -\infty & \text{otherwise,} \end{cases} \quad \text{for } d \ge 2 \text{ and } \kappa \in \mathbb{R}. \tag{3}$$

For $d = 3$, density f is obviously strictly increasing on [0,1] and log-convex when $\kappa > 0$, and decreasing and log-concave if $\kappa \le 0$. For the general case when $d \ge 2$ we need more properties of l.

Theorem 4. *Let $l_{d,\kappa}(x)$ be the marginal log-density* (3)*, the following then holds:*

(1) $l_{d,\kappa}(0) = 0$*,*

(2) $\lim_{x\to 1^-} l_{d,\kappa}(x) = \begin{cases} \infty & \text{for } d = 2, \\ \kappa & \text{for } d = 3, \\ -\infty & \text{for } d > 3. \end{cases}$

(3) For $d \ge 3$*,* $l_{d,\kappa}(x)$ *has a maximum in [0,1] located in*

$$m = \begin{cases} 1 & \text{for } d = 3 \text{ and } \kappa > 0, \\ 0 & \text{for } d \ge 3 \text{ and } \kappa \le \frac{d-3}{2}, \\ \sqrt{1 - \frac{d-3}{2\kappa}} & \text{for } d > 3 \text{ and } \kappa \ge \frac{d-3}{2}. \end{cases}$$

which is unique whenever $d > 3$ *or* $\kappa \ne 0$*.*

(4) Let $d \ge 3$ *and*

$$w_0 = \begin{cases} 1 & \text{for } d = 3 \text{ and } \kappa > 0, \\ 0 & \text{for } d \ge 3 \text{ and } \kappa \le \frac{d-3}{2}, \\ \sqrt{\left(1 + \frac{d-3}{4\kappa}\right) - \sqrt{\left(1 + \frac{d-3}{4\kappa}\right)^2 - \left(1 - \frac{d-3}{2\kappa}\right)}} & \text{for } d > 3 \text{ and } \kappa \ge \frac{d-3}{2}. \end{cases} \tag{4}$$

Then $l_{d,\kappa}$ *is convex in* $[0, w_0]$ *and concave in* $[w_0, 1)$*.*

Proof. (1) and (2) are trivial. Similarly for (3) when $d = 3$. Otherwise observe that $l'_{d,\kappa}(x) = 2\kappa x - (d-3)\frac{x}{1-x^2}$ vanishes if and only if $x = 0$ or $x^2 = 1 - \frac{d-3}{2\kappa}$ and $\kappa \ne 0$. So there is a unique stationary point in $(0, 1)$ given by $x = \sqrt{1 - \frac{d-3}{2\kappa}}$ if and only if $\kappa > \frac{d-3}{2}$ and $d > 3$. Thus the result follows.

For $\kappa \le 0$, (4) follows from the fact that both κx^2 and $\log\left(1 - x^2\right)$ are concave. For $d = 3$. $l_{3,\kappa}(x) = \kappa x^2$ and the result follows as well. For case $d > 3$ and $\kappa > 0$ observe that $l''_{d,\kappa}(x) = 2\kappa - (d-3)\frac{1+x^2}{(1-x^2)^2}$ vanishes if and only if $\left(1 - x^2\right)^2 - \frac{d-3}{2\kappa}\left(1 + x^2\right) = 0$. This quartic equation can be solved by substituting $z = x^2$ and solving $z^2 - \left(2 + \frac{d-3}{2\kappa}\right) z + \left(1 - \frac{d-3}{2\kappa}\right) = 0$ which yields

$$z = \left(1 + \frac{d-3}{4\kappa}\right) \pm \sqrt{\left(1 + \frac{d-3}{4\kappa}\right)^2 - \left(1 - \frac{d-3}{2\kappa}\right)}.$$

However, as $z \in [0, 1)$ there is no feasible solution when $\kappa < \frac{d-3}{2}$ and we obtain the unique solution

$$z_0 = \left(1 + \frac{d-3}{4\kappa}\right) - \sqrt{\left(1 + \frac{d-3}{4\kappa}\right)^2 - \left(1 - \frac{d-3}{2\kappa}\right)}$$

when $\kappa \ge \frac{d-3}{2}$. As we then also have $l''_{d,\kappa}(0) = 2\kappa - (d-3) > 0$ the result follows □

Corollary 1. *Let $l_{d,\kappa}(x)$ be the marginal log-density* (3)*, the following then holds:*

(1) $l_{2,\kappa}(x)$ *has a minimum in [0,1] located in*

$$m = \begin{cases} 0 & \text{for } \kappa \ge -\frac{1}{2}, \\ \sqrt{1 + \frac{1}{2\kappa}} & \text{for } \kappa \le -\frac{1}{2}. \end{cases}$$

which is unique whenever $\kappa \ne 0$*.*

(2) Let

$$w_0 = \begin{cases} 0 & \text{for } \kappa \geq -\frac{1}{2}, \\ \sqrt{\left(1-\frac{1}{4\kappa}\right) - \sqrt{\left(1-\frac{1}{4\kappa}\right)^2 - \left(1+\frac{1}{2\kappa}\right)}} & \text{for } \kappa \leq -\frac{1}{2}. \end{cases}$$

Then $l_{2,\kappa}(x)$ *is concave in* $[0, w_0]$ *and convex in* $[w_0, 1)$ *.*

Proof. The results follow by observing that $l_{2,\kappa}(x) = -l_{4,-\kappa}(x)$ and from the results of Theorem 4. □

Theorem 4 together with corollary 1 immediately implies that log-density $l_{d,\kappa}$ is either concave or has at most one inflection point on its support [0,1) . For the log-concave part adaptive rejection sampling (Gilks and Wild, 1992) is a suitable method. There tangents and secants to the log-density can be used to construct hat and squeeze for an acceptance-rejection algorithm automatically. It is obvious that this idea is also suitable for the log-convex part as well with the roles of tangents and secants exchanged. Botts et al. (2013) have proposed such an algorithm where even the exact position of the inflection point is not required. In fact the information that there is at most one inflection point is sufficient. However, the stochastic method for finding construction points for the tangents is replaced by deterministic method called derandomized adaptive rejection sampling in (Hörmann et al., 2013). The method is implemented in CRAN package Tinflex (Leydold et al., 2019). It is important that the log-density $l_{d,\kappa}$ has to be normalized such that it avoids numerical overflow when computing $f(x) = \exp(l_{d,\kappa}(x))$. In our experiments we use $g(x) - g(m)$ where m is the extremum from previous theorem and corollary. Notice that some rough estimate of $g(m)$ is sufficient. In our computational experiments the method works fine even for (very) large values of κ. An advantage of this method is that the rejection constant can be adjusted to some preset value. In particular it can be made close to one at the cost of a more expensive setup. Then the marginal sampling time becomes very fast. Notice, however, that this black box method requires some time consuming setup and thus might not be appropriate in the varying parameter case when only one random variate is requested for a given pair of parameters (d, κ).

4.1 Transformed Density Rejection

Transformed density rejection is a generalization of adaptive rejection sampling by Hörmann (1995) where the logarithm is replaced by some strictly monotonically increasing transformation T. Thus hat and squeezes are constructed by tangents and secants of the transformed density $T \circ f$. Hörmann (1995) also has proposed the Box-Cox-like transformations

$$T_c(x) = \begin{cases} \log(x) & \text{for } c = 0, \\ \operatorname{sgn}(c) x^c & \text{otherwise.} \end{cases}$$

These have the appealing property the any T_c-concave density is also T_{c_1} concave for every $c_1 < c$. For details we refer to Hörmann et al. (2013). There the notion of local concavity was introduced to find proper parameters for this family of distributions. It is defined as the maximal c where f is T_c-concave at x (or more precisely where $(T(f(x)))'' = 0)$. It is given by

$$\operatorname{lc}(x) = 1 - \frac{f''(x) f(x)}{\left(f'(x)\right)^2}.$$

Obviously f is T_c-concave if and only if $\operatorname{lc}(x) \geq c$ for all x in the support of f and the inflection points of $T_c \circ f$ coincide with the roots of $\operatorname{lc}(x) - c$.

Proposition 2. *Let* $b = \frac{d-3}{2}$. *Then*

$$\operatorname{lc}(x) = -\frac{2\kappa\left(1-x^2\right)^2 - 2b\left(1-x^2\right) - 4bx^2}{x^2\left[2\kappa\left(1-x^2\right) - 2b\right]^2}.$$

Proof. Let $q(x) = 2\kappa x - 2hx\left(1-x^2\right)^{-1}$. Then straightforward computation gives

$$\begin{aligned} f(x) &= \exp\left(\kappa x^2\right)\left(1-x^2\right)^b \\ f'(x) &= f(x)\left[2\kappa x - 2bx\left(1-x^2\right)^{-1}\right] = f(x) q(x) \\ f''(x) &= f'(x) q(x) + f(x) q'(x) = f(x)\left(q^2(x) + q'(x)\right) \end{aligned}$$

and thus

$$
\begin{aligned}
\operatorname{lc}(x) &= \frac{\left(f'(x)\right)^2 - f''(x)f(x)}{\left(f'(x)\right)^2} \\
&= \frac{f^2(x)q^2(x) - f(x)\left(q^2(x) + q'(x)\right)f(x)}{(f(x)q(x))^2} = -\frac{q'(x)}{q^2(x)} = \left(\frac{1}{q(x)}\right)' \\
&= -\frac{2\kappa - 2b\left(1-x^2\right)^{-1} - 4bx^2\left(1-x^2\right)^{-2}}{\left[2\kappa x - 2bx\left(1-x^2\right)^{-1}\right]^2} \\
&= -\frac{2\kappa\left(1-x^2\right)^2 - 2b\left(1-x^2\right) - 4bx^2}{x^2\left[2\kappa\left(1-x^2\right) - 2b\right]^2}
\end{aligned}
$$

as claimed. □

Proposition 3. *The following holds for* $\operatorname{lc}(x)$

(1)

$$
\lim_{x\to 0+} \operatorname{lc}(x) = \begin{cases} \infty & \text{for } \kappa < \frac{d-3}{2}, \\ 0 & \text{for } \kappa = \frac{d-3}{2}, \\ -\infty & \text{for } \kappa > \frac{d-3}{2}. \end{cases}
$$

(2)

$$
\lim_{x\to 1-} \operatorname{lc}(x) = \begin{cases} -\infty & \text{for } d = 3, \\ \frac{2}{d-3} & \text{for } d \neq 3. \end{cases}
$$

(3) $\operatorname{lc}(x) \geq 0$ *for all* $x \in (0,1)$, $d \geq 3$ *and* $\kappa \leq \frac{d-3}{2}$.

(4) For $d > 3$ *and* $\kappa > \frac{d-3}{2}$, $\operatorname{lc}(x)$ *has a unique root in (0,1) given by equation 4.*

(5) For $\kappa > \frac{d-3}{2}$ *and* $d > 3$, $\operatorname{lc}(x) - c$ *has a unique root in (0,1) for every* $c \leq 0$.

Proof. (1) The numerator of $\operatorname{lc}(x)$ at $x = 0$ is positive if $\kappa < b = \frac{d-3}{2}$ and negative if $\kappa > b$. Thus the result follows.

(2) For $d = 3$ we find $\operatorname{lc}(x) = -\frac{1}{x^2(1-x^2)}$. For $d \neq 3$ we have $\operatorname{lc}(1) = \frac{2}{d-3}$.

(3) and (4) $\operatorname{lc}(x)$ vanishes if and only if its numerator vanishes. This reduces to the same quartic equation as in the proof of Theorem 4.

(5) By (1) and (2) there is at least one. Let $U(c) = \{x \in (0,1) : \operatorname{lc}(x) \geq c\}$. Then $U(c_1) \supseteq U(c_2)$ whenever $c_1 \leq c_2$. Observe that $x \in U(c)$ if and only if $-\frac{2x\left(1-z^2\right)^2 - 2b\left(1-x^2\right) - 4bx^2}{x^2)2\kappa(1-x^2) - 26^2} \geq c$. Which is equivalent to $2\kappa(1-z)^2 - 2b(1-z) - 4bz + cz[2\kappa(1-z) - 2b]^2 \leq 0$, where $z = x^2$. The l.h.s. of this inequality is a cubic polynomial for $c < 0$ which is positive and strictly convex near $z = 0$. For $c = 0$, this l.h.s. reduces to convex quadratic polynomial with exactly one root in (0,1] and one in $(1, \infty)$. Hence if there is more than one root of the l.h.s. in (0,1] for some $c < 0$ then all its roots must be in (0,1] . By the monotonicity of $U(c)$ this cannot happen. Thus the proposition follows. □

Proposition has two immediate consequences: First it is not possible to find a $c < 0$ such that our density f is T_c-concave whenever $\kappa > \frac{d-3}{2}$ and $d \geq 3$. Second, for every $c \leq 0, T_c \circ f$ has at most one inflexion point in $(0,1)$ and we always can apply method Tinflex. Using $c = -1/2$ results in an algorithm with more constructions points than with $c = 0$ (and thus in a more expensive setup) but has a faster marginal runtime.

This method is summarized in Algorithm 2 and in additionally to the univariate sample from Tinflex needs to draw d times from standard normal distribution per iteration. Note that a similar algorithm as presented in Hornik and Grün (2014), which needs only $d-1$ samples from standard normal distribution, can be applied also in the case of Watson distribution, however the algorithm presented below tends to offer much smaller runtime for d large due to more efficient postprocessing.

Algorithm 2 Generator for Watson distribution with parameters κ and μ using Tinflex package

Sample $w \sim \mathrm{pSaw}_d(\exp(x^2), \sqrt{\kappa})$ or $w \sim \mathrm{pSaw}_d(\exp(-x^2), \sqrt{-\kappa})$ for $0 \le x < 1$ ▷ using Tinflex
Sample $S \sim \{-1, 1\}$ ▷ sample sign
$W \leftarrow Sw$
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
$Y \leftarrow Y/\sqrt{Y'Y}$
$X \leftarrow W\mu + \sqrt{1 - W^2}Y$
return X

4.2 Simple generators for $d = 3$

4.2.1 Case $d = 3$ and $\kappa > 0$

By Theorem 4 density $f_{d,\kappa}$ is log-convex on [0,1] when $\kappa > 0$. Thus we use the secant and the tangent at κ of $l_{d,\kappa}$ to construct hat function and squeeze, respectively, for an acceptance rejection algorithm. These are given by

$$h_{3,\kappa}(x) = \begin{cases} \exp(\kappa x) & \text{for } 0 \le x \le 1, \\ 0 & \text{otherwise,} \end{cases} \tag{5}$$

$$s_{3,\kappa}(x) = \begin{cases} \exp(2\kappa x - \kappa) & \text{for } 1/2 \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$h_{3,\kappa}(x) \ge f_{3,\kappa}(x) \ge s_{3,\kappa}(x).$$

For the sampling algorithm we have to draw a random variate X from a density proportional to $h(x)$. Observe that $X = 1 - E/\kappa$ where E follows a truncated standard exponential distribution with domain $[0, \kappa]$ which can generated using the inversion method and $H^{-1}(u) = -\log\left(1 - u\left(1 - e^{-\kappa}\right)\right)$. Algorithm 3 compiles the details for the resulting acceptance-rejection algorithm.

Algorithm 3 Generator for $d = 3$ and $\kappa > 0$

$k \leftarrow 1 - e^{-\kappa}$ ▷ setup
repeat ▷ acceptance-rejection loop
    Sample $U, V \sim U(0, 1)$
    $X \leftarrow 1 + \log(1 - Uk)/\kappa$
until $\log(V) \le \kappa X(X - 1)$
return X

The rejection constant is given by the ratio of the integral of the hat function and the integral of the (multiple of the) density (the normalization constant for density f cancels out):

$$\alpha = \int_{\mathbb{R}} h_{d,\kappa}(x)dx \Big/ \int_{\mathbb{R}} f_{d,\kappa}(x)dx.$$

The rejection constant can be estimated by the ratio of the integral of the hat function and the integral of a squeeze function

$$\alpha \le \rho = \int_{\mathbb{R}} h_{d,\kappa}(x)dx \Big/ \int_{\mathbb{R}} s_{d,\kappa}(x)dx.$$

For Algorithm 3 (case $d = 3$) we easily find

$$\alpha \le \rho = \frac{\int_\Omega h_{d,\kappa}(x)dx}{\int_{\mathbb{R}} s_{d,\kappa}(x)dx} = \frac{\int_0^1 \exp(\kappa x)dx}{\int_{1/2}^1 \exp(2\kappa x - \kappa)dx}$$
$$= \frac{\frac{1}{\kappa}\left(e^{\kappa} - 1\right)}{e^{-\kappa}\frac{1}{2\kappa}\left(e^{2\kappa} - e^{\kappa}\right)} = 2.$$

That is, the rejection constant is uniformly bounded by 2 for all $\kappa > 0$.

We note that Best and Fisher (1986) already suggested hat (5). However, they only provide examples for the rejection constant using values of Dawson's integral and did not discuss numerical problems with their algorithm at all. Li and

Wong (1993) suggest an enveloping function which is slightly superior to (5). Although the algorithm has a slightly larger sampling efficiency, from their experiments it follows that its actual CPU time is not always superior to the algorithm presented here.

4.2.2 Case $d = 3$ and $\kappa < -1$

By Theorem 4 density $f_{d,\kappa}$ is log-concave with mode 0. The Ratio-of-Uniforms method has been proposed by Kinderman and Monahan (1977) and offers allows for quite simple sampling algorithm. It is based on the following principle.

Proposition 4. *Let f be a multiple of some density function. If (U, V) is uniformly distributed in*

$$\mathcal{A}_f = \left\{(u, v) : 0 < v^2 \leq f(u/v)\right\}$$

then $X = U/V$ has probability density function proportional to f.

Its proof is based on the observation that map

$$\phi : \mathcal{A}_f \to \mathcal{G}_f, (U, V) \mapsto (X, Y) = \left(U/V, V^2\right),$$

where

$$\mathcal{G}_f = \{(x, y) : 0 < y \leq f(x)\}$$

has constant Jacobian determinant. Its relation to the transformed density rejection is based on the following property.

Proposition 5 (Leydold (2000))**.** *Region $\mathcal{A}_f$ is convex if and only if f is $T_{-\frac{1}{2}}$-concave.*

In particular it follows that $\mathcal{A}_f$ when f is log-concave. Sampling uniformly from $\mathcal{A}_f$ can be done by rejection from the minimal bounding rectangle given by $[u_-, u_+ \mid \times [0, v_+]$. It is given by the extremal u and v coordinates of points the $(u, v) = \phi^{-1}(x, f(x)) = (x\sqrt{f(x)}, \sqrt{f(x)}))$. In the case of Watson distribution with $\kappa < 0$ we find

$$\begin{aligned} v_+ &= \sqrt{f(m)} = \sqrt{f(0)} = 1, \\ u_+ &= \max_{x\in[-1,1]} x\sqrt{f(x)} = \max_{x\in[-1,1]} x \exp\left(\kappa x^2/2\right) = \begin{cases} 1/\sqrt{-\kappa e} & \text{for } \kappa \leq -1, \\ \exp(\kappa/2) & \text{for } -1 < \kappa \leq 0, \end{cases} \\ u_- &= -u_+. \end{aligned}$$

The case distinction for u_+ is necessary as the stationary point for $x\sqrt{f(x)}$ is given by $x_0^2 = -1/\kappa$, which is out of domain when $\kappa > -1$. Algorithm 4 compiles the details for the resulting ratio-of-uniforms algorithm.

Algorithm 4 Generator for $d = 3$ and $\kappa \leq -1$

```
1: u+ ← 1/√(−κe)                                  ▷ setup
2: repeat                                          ▷ acceptance-rejection loop
3:    Sample U ∼ U(−u+, u+); V ∼ U(0, 1)
4:    X ← U/V
5: until log(V) ≤ κX²/2
6: Sample S ∼ {−1, 1}                              ▷ sample sign
7: return SX
```

The rejection constant for Algorithm 4 is uniformly bounded by 2 for all $\kappa \leq 2$. This follows from the fact that region $\mathcal{A}_f$ is convex (Leydold, 2000) and thus contains the quadrangle with vertices $(0, 0)$, $(0, 1)$, and the two extremal points with u -coordinates $\pm u_+$. We want to remark that the rejection constant can be improved if the minimal bounding rectangle in Algorithm 2 is replaced by a bounding trapezoid where the left and right edge is created by tangents to $\mathcal{A}_f$ constructed at the points x_0 with $f(x_0) = 1/4$. This is the optimal hat for the ratio-of-uniforms method with bounding quadrangles, see Hörmann et al. (2013) for details.

4.2.3 Case $d = 3$ and $-1 \leq \kappa \leq 0$

In this case we simply can use a constant hat, see Algorithm 5.

Using theory from the ratio-of-uniforms method it follows that the rejection constant is less than 2 for all $\kappa \in [-1, 0]$ as $f(1) = 1/e > 1/4$. We also remark that we can use Algorithm 4 where setup step 1 is replaced by $u_m \leftarrow \exp(\kappa/2)$. However, Algorithm 5 has a smaller rejection constant.

Algorithm 5 Generator for $d = 3$ and $-1 \leq \kappa \leq 0$

```
repeat                                          ▷ acceptance-rejection loop
    Sample X, V ~ U(0, 1)
until log(V) ≤ κX^2
Sample S ~ {−1, 1}                              ▷ sample sign
return SX
```

Li and Wong (1993) proposed for $\kappa < 0$ a sampling technique based on simulation from the truncated normal distribution. From our experiments, the CPU time of TE algorithm (Li and Wong, 1993) tends to be slightly smaller as from the presented ratio-of-uniforms methods and other methods in the literature, however its numerical stability appears questionable for small values of κ where the exponential functions could easily underflow.

We further note that a simple generator involving ratio-of-uniforms can be also applied to other sets of parameters, however, without further numerical techniques can be only used in some cases. Thus in our opinion a general approach using Tinflex package (Leydold et al., 2019) is a preferred choice for the other cases.

5 Simulation study

In the following we compare sampling times of both algorithms (ACG-envelope and Tinflex) for different sample sizes n. Since Tinflex requires some time for the setup, it is expected to be relatively slow for small n compared to the ACG-sampler, however thanks to the univariate form of the marginal distribution and bounded rejection constant it should dominate for n large enough and d small enough. All experiments were measured 300 times from which an average was calculated. The measurements were all done in R (R Core Team, 2018) on operating system Ubuntu 16.04 LTS and compiled using GNU Compiler Collection. The ACG algorithm (Algorithm 1) was written in C++ and the results were integrated into R using Rcpp (Eddelbuettel and François, 2011). Tinflex package for Algorithm 2 is mainly written in C and the postprocessing of returned values was calculated in R. The following tables and plots summarize the results.

$d =$	3	5	10	20	50	100	200	1000
$\kappa = -100$	0.02	0.02	0.02	0.03	0.06	0.09	0.16	0.81
-50	0.02	0.02	0.02	0.03	0.05	0.09	0.16	0.78
-10	0.02	0.02	0.02	0.03	0.05	0.09	0.16	0.79
-1	0.02	0.02	0.02	0.03	0.05	0.09	0.16	0.78
1	0.02	0.02	0.02	0.03	0.05	0.09	0.16	0.78
10	0.02	0.03	0.03	0.03	0.05	0.09	0.16	0.79
50	0.02	0.03	0.04	0.07	0.13	0.10	0.17	0.78
100	0.02	0.03	0.04	0.08	0.19	0.33	0.18	0.78

$d =$	3	5	10	20	50	100	200	1000
$\kappa = -100$	0.10	0.05	0.05	0.06	0.09	0.13	0.22	1.00
-50	0.05	0.05	0.05	0.06	0.09	0.13	0.22	0.94
-10	0.05	0.05	0.06	0.06	0.09	0.13	0.22	0.93
-1	0.05	0.05	0.05	0.06	0.09	0.13	0.22	0.95
1	0.05	0.05	0.05	0.06	0.09	0.13	0.22	0.95
10	0.05	0.05	0.06	0.06	0.09	0.13	0.22	0.93
50	0.05	0.05	0.06	0.07	0.09	0.13	0.22	0.95
100	0.05	0.06	0.06	0.07	0.10	0.13	0.22	0.95

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 κ (as columns).

$d =$	3	5	10	20	50	100	200	1000
$\kappa = -100$	0.83	0.85	1.18	1.92	3.93	7.52	15.10	83.21
-50	0.74	0.86	1.19	1.89	3.94	7.53	14.85	82.19
-10	0.76	0.83	1.15	1.99	3.92	7.55	14.75	84.33
-1	0.65	0.79	1.12	1.86	3.94	7.59	14.86	82.75
1	0.65	0.81	1.12	1.85	3.91	8.07	14.80	82.92
10	0.91	1.25	1.78	2.06	4.42	7.54	15.32	82.06
50	0.98	1.46	2.67	5.55	11.97	8.60	14.78	83.79
100	0.98	1.50	2.77	6.15	17.32	31.56	17.46	82.05

$d =$	3	5	10	20	50	100	200	1000
$\kappa = -100$	0.59	0.69	1.17	2.21	4.49	8.81	18.61	107.16
-50	0.51	0.70	1.18	1.99	4.48	8.97	18.61	109.31
-10	0.54	0.70	1.16	2.21	4.50	9.31	18.97	107.58
-1	0.48	0.70	1.13	1.97	4.51	9.09	18.60	106.69
1	0.51	0.70	1.16	2.00	4.51	8.96	18.68	109.32
10	0.52	0.72	1.22	2.01	4.57	8.86	18.68	110.27
50	0.52	0.71	1.21	2.01	4.60	8.88	18.42	106.11
100	0.52	0.70	1.15	2.06	4.64	9.40	18.64	109.00

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 κ (as columns).

We note that in the cases where the measured runtimes decrease as d is increased are caused by the steep decay of the rejection constant in those areas, which tends to have in some cases higher impact as extra complexity caused by the higher dimension. The measured times were further compared relatively and visualized in the following graphics. The results show a very nice balance between the samplers, where one algorithm excels in the areas of parameter space where the second has its limitations defined by the pure design of the algorithm.

Figure 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 $d > 3$ and $\kappa \geq (d-3)/2$ (i.e., 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.

Figure 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 $d > 3$ and $\kappa \geq (d-3)/2$ (i.e., 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.

6 Conclusions

In this paper we presented and analyzed two methods for sampling from the Watson distribution.

Firstly we introduced a generator based on the rejection sampling algorithm from Kent et al. (2018). We showed that by adapting this method to the Watson distribution, a closed form expression for the parameter that maximize the efficiency of the sampling procedure is obtained. This allows to further investigate the efficiency of the procedure which resulted in the asymptotic expressions for the rejection constant given by

$$\lim_{\kappa \to -\infty} \mathcal{M}_d(\kappa) \approx \sqrt{\frac{d}{d-1}}, \qquad \text{and} \qquad \lim_{\kappa \to \infty} \mathcal{M}_d(\kappa) \approx \sqrt{\frac{de}{2}},$$

for d large enough. Moreover, we showed that the rejection constant is decreasing and increasing on $(-\infty, 0)$ and $(0, \infty)$ respectively, and hence bounded by the asymptotic results for any κ. Thus, these results can be seen as worst case scenarios and a general bounds for the efficiency of the algorithm.

The second algorithm uses 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 one inflection point. This allows to use the Tinflex algorithm (Leydold et al., 2019), which exactly requires such a setup. Furthermore, we investigated the application of transformed density rejection and proposed some simple generators for the trivial cases where $d = 3$. What is more, both main algorithms offer fast runtimes also for high dimensions by avoiding the curse of dimensionality that is present in many similar sampling schemes in the literature.

Finally, both algorithms were compared in a simulation study for different set of parameters.

References

DJ Best and NI Fisher. Goodness-of-fit and discordancy tests for samples from the Watson distribution on the sphere. *Australian Journal of Statistics*, 28(1):13–31, 1986.

Avleen S. Bijral, Markus Breitenbach, and Greg Grudic. Mixture of watson distributions: A generative model for hyperspherical embeddings. volume 2 of *Proceedings of Machine Learning Research*, pages 35–42, San Juan, Puerto Rico, March 2007. URL `http://proceedings.mlr.press/v2/bijral07a/bijral07a.pdf`.

Carsten Botts, Wolfgang Hörmann, and Josef Leydold. Transformed density rejection with inflection points. *Statistics and Computing*, 23(2):251–260, 2013.

DLMF. *NIST Digital Library of Mathematical Functions*. http://dlmf.nist.gov/, Release 1.0.19 of 2018-06-22. URL `http://dlmf.nist.gov/`. 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.

Dirk Eddelbuettel and Romain François. Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, 40(8): 1–18, 2011. doi: 10.18637/jss.v040.i08. URL `https://www.jstatsoft.org/v40/i08/`.

Walter R Gilks and Pascal Wild. Adaptive rejection sampling for Gibbs sampling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 41(2):337–348, 1992.

Wolfgang Hörmann. A rejection technique for sampling from t-concave distributions. *ACM Transactions on Mathematical Software (TOMS)*, 21(2):182–193, 1995.

Wolfgang Hörmann, Josef Leydold, and Gerhard Derflinger. *Automatic nonuniform random variate generation*. Springer Science & Business Media, 2013.

Kurt Hornik and Bettina Grün. movMF: An R package for fitting mixtures of von Mises-Fisher distributions. *Journal of Statistical Software, Articles*, 58(10):1–31, 2014. ISSN 1548-7660. doi: 10.18637/jss.v058.i10. URL `https://www.jstatsoft.org/v058/i10`.

John T. Kent, Asaad M. Ganeiber, and Kanti V. Mardia. A new unified approach for the simulation of a wide class of directional distributions. *Journal of Computational and Graphical Statistics*, 27(2):291–301, 2018. doi: 10.1080/10618600.2017.1390468. URL `https://doi.org/10.1080/10618600.2017.1390468`.

Albert J Kinderman and John F Monahan. Computer generation of random variables using the ratio of uniform deviates. *ACM Transactions on Mathematical Software (TOMS)*, 3(3):257–260, 1977.

Josef Leydold. Automatic sampling with the ratio-of-uniforms method. *ACM Transactions on Mathematical Software (TOMS)*, 26(1):78–98, 2000.

Josef Leydold, Carsten Botts, and Wolfgang Hörmann. *Tinflex: A Universal Non-Uniform Random Number Generator*, 2019. URL `https://CRAN.R-project.org/package=Tinflex`. R package version 1.5.

Kim-Hung Li and Carl Ka-Fai Wong. Random sampling from the Watson distribution. *Communications in Statistics - Simulation and Computation*, 22(4):997–1009, 1993. doi: 10.1080/03610919308813139. URL `https://doi.org/10.1080/03610919308813139`.

Kanti V. Mardia and Peter E. Jupp. *Directional Statistics*. Probability and Statistics. Wiley, 1999. ISBN 978-0-471-95333-3.

G. J. McLachlan. 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(3):318–324, 1987. ISSN 00359254, 14679876. URL `http://www.jstor.org/stable/2347790`.

Adrian O'Hagan, Thomas Brendan Murphy, Luca Scrucca, and Isobel Claire Gormley. Investigation of parameter uncertainty in clustering using a gaussian mixture model via jackknife, bootstrap and weighted likelihood bootstrap. *Computational Statistics*, 34(4):1779–1813, 2019.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018. URL `https://www.R-project.org/`.

Lukas Sablica and Kurt Hornik. On bounds for Kummer's function ratio. *Mathematics of Computation*, electronically published on November 5, 2021. doi: https://doi.org/10.1090/mcom/3690. (to appear in print).

John G. Saw. A family of distributions on the m-sphere and some hypothesis tests. *Biometrika*, 65(1):69–73, 1978. doi: 10.2307/2335278.

Suvrit Sra and Dmitrii Karp. The multivariate Watson distribution: Maximum-likelihood estimation and other aspects. *Journal of Multivariate Analysis*, 114:256–269, February 2013. ISSN 0047-259X. doi: 10.1016/j.jmva.2012.08.010. URL `http://dx.doi.org/10.1016/j.jmva.2012.08.010`.

Gary Ulrich. Generation of distributions on the m-sphere. *Journal of the Royal Statistical Society Series C (Applied Statistics)*, 33(2):158–163, 1984. doi: 10.2307/2347441.

G. S. Watson. Equatorial distributions on a sphere. *Biometrika*, 52(1/2):193–201, 1965. ISSN 00063444. URL `http://www.jstor.org/stable/2333824`.