MATHEMATICAL ENGINEERING
TECHNICAL REPORTS

Holonomic Gradient Method for Distribution Function of a Weighted Sum of Noncentral Chi-square Random Variables

Tamio KOYAMA and Akimichi TAKEMURA

METR 2015–17 April 2015

DEPARTMENT OF MATHEMATICAL INFORMATICS
GRADUATE SCHOOL OF INFORMATION SCIENCE AND TECHNOLOGY
THE UNIVERSITY OF TOKYO
BUNKYO-KU, TOKYO 113-8656, JAPAN

WWW page: http://www.keisu.t.u-tokyo.ac.jp/research/techrep/index.html

Holonomic gradient method for distribution function of a weighted sum of noncentral chi-square random variables

Tamio Koyama* and Akimichi Takemura*

April, 2015

Abstract

We apply the holonomic gradient method to compute the distribution function of a weighted sum of independent noncentral chi-square random variables. It is the distribution function of the squared length of a multivariate normal random vector. We treat this distribution as an integral of the normalizing constant of the Fisher-Bingham distribution on the unit sphere and make use of the partial differential equations for the Fisher-Bingham distribution.

Keywords and phrases: algebraic statistics, cumulative chi-square distribution, Fisher-Bingham distribution, goodness of fit

1 Introduction

The weighted sum of independent chi-square variables appears in many important problems in statistics. In the problems for testing against ordered alternatives, cumulative chi-square statistic (cf. [5], [11]) has a good power. For studying the power function of the cumulative chi-square statistic, we need to evaluate the distribution function a sum of weighted independent *noncentral* chi-square variables. Goodness of fit test statistics based on empirical cumulative distribution function, such as the Cramér-von Mises statistic or the Anderson-Darling statistic ([1]), are infinite sums of weighted independent chi-square variables. Chapter 4 of [4] gives a survey of these statistics. Under an alternative hypothesis the chi-square variables are noncentral. For studying the power function of these statistics we want to approximate the infinite sum by a finite sum of sufficiently many terms and compute the cumulative distribution of the finite sum.

An exact evaluation of the cumulative distribution function of a weighted sum of independent noncentral chi-square random variables was considered to be a difficult numerical problem (see [2]). Although the moment generating function is explicitly given, its Fourier inversion to evaluate the density function and the cumulative distribution function is difficult as extensively discussed in Chapter 6 of [14]. See [3] for the similar problems in other areas of applied mathematics.

Recently in [12] we proposed the holonomic gradient method (HGM) for calculating distribution functions and the maximum likelihood estimates using differential equations satisfied by a probability density function with respect to the parameters. Since then the method has been successfully used in many problems, including the computations related to the Fisher-Bingham

*Graduate School of Information Science and Technology, University of Tokyo

distribution on the unit sphere ([8], [6], [7], [13]). In this paper we utilize the results on HGM for the Fisher-Bingham distribution to evaluate the distribution function of a weighted sum of noncentral chi-square random variables.

Let X denote a d-dimensional random vector following the multivariate normal distribution $N(\mu, \Sigma)$. Consider the cumulative distribution function $G(r)$ of $\|X\|^2$:

$$G(r) = \int_{x_1^2+\cdots+x_d^2 \le r^2} \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right) dx. \tag{1}$$

We call $G(r)$ the ball probability with radius r. By rotation we can assume that $\Sigma = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_d^2)$ is a diagonal matrix without loss of generality. Hence $G(r)$ is the distribution function of a weighted sum of independent noncentral chi-square random variables, where weights are σ_i^2, $i = 1, \ldots, d$. Furthermore the conditional distribution of X given its length $r = \|X\|$ is the Fisher-Bingham distribution. This fact allows us to directly apply the results for the Fisher-Bingham distribution to the evaluation of the distribution of the weighted sum of independent noncentral chi-square random variables. As we show in Section 4 our method works very well, both in accuracy and speed.

The organization of this paper is as follows. In Section 2 we summarize known results on HGM for the Fisher-Bingham distribution and show how they can be used to evaluate the distribution of the a weighted sum of independent noncentral chi-square random variables. We also discuss the problem of initial values needed to use HGM. In Section 3 we present asymptotic results for the Fisher-Bingham integral and its derivatives for the case that the length of the multivariate normal vector diverges to infinity. This result is used to check the the numerical accuracy of our experiments in Section 4. We end the paper with some discussions in Section 5.

2 Holonomic system and initial values

Let

$$\Sigma = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_d^2), \qquad \mu = (\mu_1, \ldots, \mu_d)^\top.$$

We define new parameters λ_i, τ_i, $i = 1, \ldots, d$, by

$$\lambda_i = -\frac{1}{2\sigma_i^2}, \qquad \tau_i = \frac{\mu_i}{\sigma_i^2}$$

and the Fisher-Bingham integral $f(\lambda, \tau, r)$ by

$$f(\lambda, \tau, r) = \int_{S^{d-1}(r)} \exp\left(\sum_{i=1}^d \lambda_i t_i^2 + \sum_{i=1}^d \tau_i t_i\right) dt, \tag{2}$$

where $S^{d-1}(r) = \left\{t \in \mathbf{R}^d \mid t_1^2 + \cdots + t_d^2 = r^2\right\}$ is the sphere of radius r and dt is the volume element of $S^{d-1}(r)$ so that

$$\int_{S^{d-1}(r)} dt = r^{d-1} S_{d-1}, \quad S_{d-1} = \mathrm{Vol}(S^{d-1}(1)) = \frac{2\pi^{d/2}}{\Gamma(d/2)}.$$

Then $G(r)$ in (1) is written as

$$G(r) = \frac{\prod_{i=1}^{d}\sqrt{-\lambda_i}}{\pi^{d/2}} \exp\left(\frac{1}{4}\sum_{i=1}^{d}\frac{\tau_i^2}{\lambda_i}\right)\int_0^r f(\lambda,\tau,s)ds. \tag{3}$$

We will numerically integrate the right-hand side of (3). We denote the partial differential operator with respect to λ by ∂_λ. For $t \in S^{d-1}(r)$, $(t_1^2+\cdots+t_d^2)/r^2 = 1$ and

$$\begin{aligned} f(\lambda,\tau,r) &= \int_{S^{d-1}(r)} \frac{1}{r^2}(t_1^2+\cdots+t_d^2)\exp\left(\sum_{i=1}^{d}\lambda_i t_i^2 + \sum_{i=1}^{d}\tau_i t_i\right)dt \\ &= \frac{1}{r^2}\left(\partial_{\lambda_1}+\cdots+\partial_{\lambda_d}\right) f(\lambda,\tau,r). \end{aligned} \tag{4}$$

By HGM we evaluate $\partial_{\lambda_i} f(\lambda,\tau,r)$, $i=1,\ldots,d$, and use (4) to compute $f(\lambda,\tau,r)$. In fact we also evaluate $\partial_{\tau_i} f(\lambda,\tau,r)$, $i=1,\ldots,d$.

Define a $2d$-dimensional vector of partial derivatives of $f(\lambda,\tau,r)$ by

$$F = \left(\partial_{\tau_1} f,\ldots,\partial_{\tau_d} f, \partial_{\lambda_1} f,\ldots,\partial_{\lambda_d} f\right)^\top. \tag{5}$$

Elements of F are called "standard monomials" in HGM. By Theorem 3 of [7] we have

$$\partial_r F = P_r F, \tag{6}$$

where the $2d\times 2d$ matrix $P_r = (p_{ij})$, called the Pfaffian matrix, is of the form

$$P_r = \frac{1}{r}\begin{pmatrix} 2r^2\lambda_1+1 & & \boldsymbol{O} & \tau_1 & \cdots & \tau_1 \\ & \ddots & & & \vdots & \\ \boldsymbol{O} & & 2r^2\lambda_d+1 & \tau_d & \cdots & \tau_d \\ r^2\tau_1 & & \boldsymbol{O} & 2r^2\lambda_1+2 & & \mathbf{1} \\ & \ddots & & & \ddots & \\ \boldsymbol{O} & & r^2\tau_d & \mathbf{1} & & 2r^2\lambda_d+2 \end{pmatrix}, \tag{7}$$

with $\boldsymbol{O}$ denoting an off-diagonal block of 0's and $\mathbf{1}$ denoting an off-diagonal block of 1's. The elements p_{ij} of P_r are expressed as

$$\begin{aligned} rp_{ij} &= (2\lambda_i r^2+1)\delta_{ij} + \sum_{k=1}^{d}\tau_i\delta_{j(k+d)} \quad (1\le i\le d), \\ rp_{(i+d)j} &= \tau_i r^2\delta_{ij} + (2\lambda_i r^2+2)\delta_{j(i+d)} + \sum_{k\neq i}\delta_{j(k+d)} \quad (1\le i\le d), \end{aligned}$$

for $1\le j\le 2d$, where δ_{ij} denotes Kronecker's delta. Given initial values for the elements of F at $r=r_0$, we can apply a standard ODE solver to (6) for numerically evaluating F.

For the initial values at a small $r=r_0>0$, we can use the following series expansion of the Fisher-Bingham integral ([9]):

$$f(\lambda,\tau,r) = r^{d-1}S_{d-1}\times\sum_{\alpha,\beta\in\mathbb{N}_0^d} r^{2|\alpha+\beta|}\frac{(d-2)!!\prod_{i=1}^{d}(2\alpha_i+2\beta_i-1)!!}{(d-2+2|\alpha|+2|\beta|)!!\alpha!(2\beta)!}\lambda^\alpha\tau^{2\beta}, \tag{8}$$

where $\mathbb{N}_0 = \{0, 1, 2, \dots\}$ and for a multi-index $\alpha \in \mathbb{N}_0^d$ we define

$$\alpha! = \prod_{i=1}^{d} \alpha_i!, \quad \alpha!! = \prod_{i=1}^{d} \alpha_i!! \text{ and } |\alpha| = \sum_{i=1}^{d} \alpha_i.$$

By term differentiation of this series we can evaluate derivatives of $f(\lambda, \tau, r)$. For computing the initial values, we apply the following approximation:

$$\frac{\partial f}{\partial \tau_i} = S_{d-1} r^{d+1} \tau_i + O(r^{d+2}) \quad (i = 1, \dots, d),$$
$$\frac{\partial f}{\partial \lambda_i} = S_{d-1} r^{d+1} + O(r^{d+2}) \quad (i = 1, \dots, d).$$

By this approximation, we reduce the computational time for the initial values. However the accuracy of the result does not decrease at all as we will show in Section 4.

As $r \to \infty$, the absolute values of $f(\lambda, \tau, r)$ and its derivatives become exponentially small, as we analyze the behavior in the next section. Hence we also consider the following vector

$$Q = \exp(-r^2\lambda_1 - r|\tau_1|) \left(\frac{1}{r}\partial_{\tau_1} f, \partial_{\tau_2} f, \dots, \partial_{\tau_d} f, \frac{1}{r^2}\partial_{\lambda_1} f, \partial_{\lambda_2} f, \dots, \partial_{\lambda_d} f \right)^\top. \tag{9}$$

Then from (6) it is easy to obtain $\partial_r Q$ as

$$\partial_r Q = \left(D^{-1}\partial_r D - (2r\lambda_1 + |\tau_1|) I_{2d} + D P_r D^{-1} \right) Q, \tag{10}$$

where I_{2d} is the identity matrix with size $2d$ and

$$D = \mathrm{diag}(\frac{1}{r}, 1, \dots, 1, \frac{1}{r^2}, 1, \dots, 1).$$

The equation (9) is a refinement of the equation (21) in [7]. By Proposition 3.1 in the next section, each element of Q converges to some non-zero value when r goes to the infinity. This prevents the adaptive Runge-Kutta method from slowing down.

3 Laplace approximation close to the infinity

In our implementation of HGM, we start from a small $r = r_0 > 0$ and numerically integrate F in (5) up to $r = 1$ and then integrate Q in (9) toward $r = \infty$. In order to assess the accuracy of Q for large r, we derive the asymptotic values of the elements of Q by the Laplace method. The Laplace approximation for the Fisher-Bingham integral was given in [10]. However here we also need approximations for its derivatives. Hence we give the approximations and a sketch of their proofs.

We first consider the case of single largest λ_1. We state the following result.

Proposition 3.1. *Suppose* $0 > \lambda_1 > \lambda_2 \geq \cdots \geq \lambda_d$. *Then, as* $r \to \infty$,

$$f(\lambda, \tau, r) = \frac{\pi^{(d-1)/2}}{\prod_{i=2}^{d}(\lambda_1 - \lambda_i)^{1/2}}(e^{r\tau_1} + e^{-r\tau_1}) \exp\left(r^2\lambda_1 - \sum_{i=2}^{d} \frac{\tau_i^2}{4(\lambda_i - \lambda_1)}\right)(1 + o(1)). \tag{11}$$

$$\partial_{\lambda_1} f(\lambda, \tau, r) = r^2 f(\lambda_1, \tau_1, r)(1 + o(1)), \tag{12}$$

$$\partial_{x_j} f(\lambda, \tau, r) = \left\{\left(\frac{\tau_j}{2(\lambda_j - \lambda_1)}\right)^2 + \frac{1}{2(\lambda_1 - \lambda_j)}\right\} f(\lambda, \tau, r)(1 + o(1)), \quad (j = 2, \ldots, d) \tag{13}$$

$$\partial_{y_1} f(\lambda, \tau, r) = r\frac{e^{r\tau_1} - e^{-r\tau_1}}{e^{r\tau_1} + e^{-r\tau_1}} f(\lambda, \tau, r)(1 + o(1)), \tag{14}$$

$$\partial_{y_j} f(\lambda, \tau, r) = \frac{\tau_j}{2(\lambda_1 - \lambda_j)} f(\lambda, \tau, r)(1 + o(1)), \quad (j = 2, \ldots, d). \tag{15}$$

Note that for $\tau_1 > 0$, in (11) $e^{-r\tau_1}$ is exponentially smaller than $e^{r\tau_1}$ and it can be omitted. However we leave $e^{-r\tau_1}$ there for consistency with the case of $\tau_1 = 0$. Also we found that leaving $e^{-r\tau_1}$ in (11) greatly improves the approximation.

We now give a rough proof of Proposition 3.1. Replacing t_i by rt_i and integrating over $S^{d-1}(1)$ can write

$$f(\lambda, \tau, r) = r^{d-1} \int_{S^{d-1}(1)} \exp\left(r^2 \sum_{i=1}^{d} \lambda_i t_i^2 + r \sum_{i=1}^{d} \tau_i t_i\right) dt, \tag{16}$$

$$\partial_{\lambda_j} f(\lambda, \tau, r) = r^{d+1} \int_{S^{d-1}(1)} t_j^2 \exp\left(r^2 \sum_{i=1}^{d} \lambda_i t_i^2 + r \sum_{i=1}^{d} \tau_i t_i\right) dt, \tag{17}$$

$$\partial_{\tau_j} f(\lambda, \tau, r) = r^{d} \int_{S^{d-1}(1)} t_j \exp\left(r^2 \sum_{i=1}^{d} \lambda_i t_i^2 + r \sum_{i=1}^{d} \tau_i t_i\right) dt. \tag{18}$$

For very large r

$$r^2(\lambda_1 t_1^2 + \lambda_2 t_2^2 + \cdots + \lambda_d t_d^2), \qquad 1 = t_1^2 + \cdots + t_d^2, \tag{19}$$

takes its maximum value at two points $t_1 = \pm 1, t_2 = \cdots = t_d = 0$. The main contributions to (16)–(18) come from neighborhoods of these two points $(\pm 1, 0, \ldots, 0)$. The contribution from the complement of these two neighborhoods are exponentially small as $r \to \infty$. In fact, we also have to consider the effect of $r\sum_{i=1}^{d} \tau_i t_i$. But it is of the order $O(r)$, whereas (19) is of the order $O(r^2)$. Hence $r\sum_{i=1}^{d} \tau_i t_i$ only perturbs the maximizing values $(\pm 1, 0, \ldots, 0)$ by the term of the order $O(1/r)$. Based on these considerations write

$$t_1^2 = 1 - t_2^2 - \cdots - t_d^2, \qquad t_1 = \pm\sqrt{1 - t_2^2 - \cdots - t_d^2} \doteq \pm\big(1 - \frac{1}{2}(t_2^2 + \cdots + t_d^2)\big),$$

where $|t_2|, \ldots, |t_d|$ are small. As shown below, $|t_i|$, $i = 2, \ldots, d$, are of the order $O(1/r)$. We now

consider the neighborhood of $(1, 0, \ldots, 0)$. By completing the squares we have

$$\begin{aligned}
&r^2 \sum_{i=1}^{d} \lambda_i t_i^2 + r \sum_{i=1}^{d} \tau_i t_i \\
&\quad = r^2\lambda_1 + r\tau_1 + r^2 \sum_{i=2}^{d} \big((\lambda_i - \lambda_1 - \frac{\tau_1}{2r}) t_i^2 + \frac{\tau_i}{r} t_i \big) + o(1) \\
&\quad = r^2\lambda_1 + r\tau_1 + \sum_{i=2}^{d} \big[(\lambda_i - \lambda_1 - \frac{\tau_1}{2r})(rt_i + \frac{\tau_i}{2(\lambda_i - \lambda_1 - \frac{\tau_1}{2r})})^2 - \frac{\tau_j^2}{4(\lambda_i - \lambda_1 - \frac{\tau_1}{2r})} \big] + o(1) \quad (20) \\
&\quad = r^2\lambda_1 + r\tau_1 + \sum_{i=2}^{d} \big[(\lambda_i - x_1)(rt_i + \frac{\tau_i}{2(\lambda_i - \lambda_1)})^2 - \frac{\tau_j^2}{4(\lambda_i - \lambda_1)} \big] + o(1).
\end{aligned}$$

Furthermore around $(1, 0, \ldots, 0)$ the volume element dt of the unit sphere $S^{d-1}(1)$ is approximately equal to the Lebesgue measure $dt_2 \ldots dt_d$, with the error of the order $t_2^2 + \cdots + t_d^2$. Hence by the change of variables

$$u_i = rt_i, \quad i = 2, \ldots, d,$$

the contribution to $f(\lambda, \tau, r)$ from the neighborhood of $(1, 0, \ldots, 0)$ is evaluated as

$$\begin{aligned}
&\exp(r^2\lambda_1 + r\tau_1) \int_{\mathbb{R}^{d-1}} \exp((\lambda_i - \lambda_1)(u_i + \frac{\tau_i}{2(\lambda_i - \lambda_1)})^2 - \frac{\tau_j^2}{4(\lambda_i - \lambda_1)}) du_1 \ldots du_d \\
&\quad = \exp\left(r^2\lambda_1 + r\tau_1 - \sum_{i=2}^{d} \frac{\tau_i^2}{4(\lambda_i - \lambda_1)} \right) \frac{\pi^{(d-1)/2}}{\prod_{i=2}^{d} (\lambda_1 - \lambda_i)^{1/2}}. \quad (21)
\end{aligned}$$

Similarly by changing the sign of τ_1 we can evaluate the contribution from the neighborhood of $(-1, 0, \ldots, 0)$ as

$$\exp\left(r^2\lambda_1 - r\tau_1 - \sum_{i=2}^{d} \frac{\tau_i^2}{4(\lambda_i - \lambda_1)} \right) \frac{\pi^{(d-1)/2}}{\prod_{i=2}^{d} (\lambda_1 - \lambda_i)^{1/2}}. \quad (22)$$

Adding (21) and (22) we obtain (11).

For $\partial_{\lambda_1} f(\lambda, \tau, r)$ and $\partial_{\tau_1} f(\lambda, \tau, r)$, we can just put $t_1 = \pm 1$ in (17) and (18). Adding contributions from two neighborhoods we obtain (12) and (14).

For $\partial_{x_i} f(\lambda, \tau, r)$ and $\partial_{\tau_i} f(\lambda, \tau, r)$, $j \geq 2$, we write

$$\begin{aligned}
t_j = \frac{u_j}{r} &= \frac{1}{r} \left(u_j + \frac{\tau_j}{2(\lambda_j - \lambda_1)} - \frac{\tau_j}{2(\lambda_j - \lambda_1)} \right), \\
t_j^2 &= \frac{1}{r^2} \left(u_j + \frac{\tau_j}{2(\lambda_j - \lambda_1)} - \frac{\tau_j}{2(\lambda_j - \lambda_1)} \right)^2
\end{aligned}$$

and take the expectation with respect to a normal density. Then we obtain (13) and (15). Although we did not give a detailed analysis of the remainder terms, we can show that the relative errors in (12)–(15) are of the order $O(1/r)$. This completes the proof of Proposition 3.1.

A generalization of Proposition 3.1 to the case that $\lambda_1 = \cdots = \lambda_m > \lambda_{m+1} \geq \cdots \geq \lambda_d$ is given in Appendix.

4 Numerical experiments

In this section we describe our numerical experiments on the performance of HGM. The programs and the raw data of our numerical experiments are obtained at

`http://github.com/tkoyama-may10/-ball-probability/`

In our experiments we compute the initial values of $r^{-(d+1)}F$ at $r = r_0 = 1.0\times10^{-6}$ by truncating the series in (8). The reason for multiplying F by $r^{-(d+1)}$ is that the values of elements of F are too small at r_0 for floating point numbers. Then up to $r = 1$, we solve the differential equation (6) numerically. In our implementation, we utilize explicit embedded Runge-Kutta Prince-Dormand (8, 9) method and we set the accuracy to 1.0×10^{-6}. In order to prevent the elements of F becoming too large, we re-scale the elements of F several times. Then at $r = 1$ we switch to Q in (9) and solve (10).

Our first experiment is for $d = 3$ and the following parameter values

$$\begin{aligned} &\sigma_1 = 3.0, && \sigma_2 = 2.0, && \sigma_3 = 1.0, \\ &\mu_1 = 0.01, && \mu_2 = 0.02, && \mu_3 = 0.03, \end{aligned} \tag{23}$$

i.e.,

$$\begin{aligned} &\lambda_1 = -0.0555556, && \lambda_2 = -0.125, && \lambda_3 = -0.5, \\ &\tau_1 = 0.00111111, && \tau_2 = 0.005, && \tau_3 = 0.03. \end{aligned}$$

By HGM we compute $G(r)$. We show its graph in Figure 1.

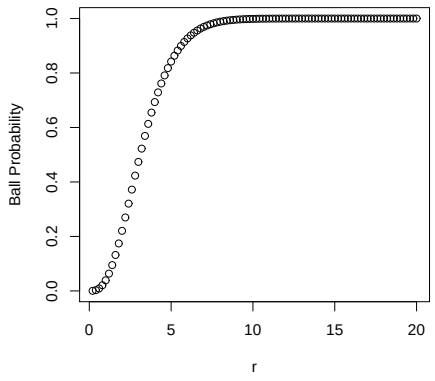


Figure 1: CDF $G(r)$ for the first experiment

For this example, we also check the accuracy by computing the ratios of $f(\lambda, \tau, r)$ and the elements of F to their asymptotic expressions in Proposition 3.1. The left figure of Figure 2 shows the ratio of $f(\lambda, \tau, r)$ to its asymptotic expression and the right figure shows the ratios of elements of F to their asymptotic expressions. We see that the numerical integration involved in HGM, starting from a small r_0, is remarkably accurate, so that the ratios numerically converge to 1 as $r \to \infty$.

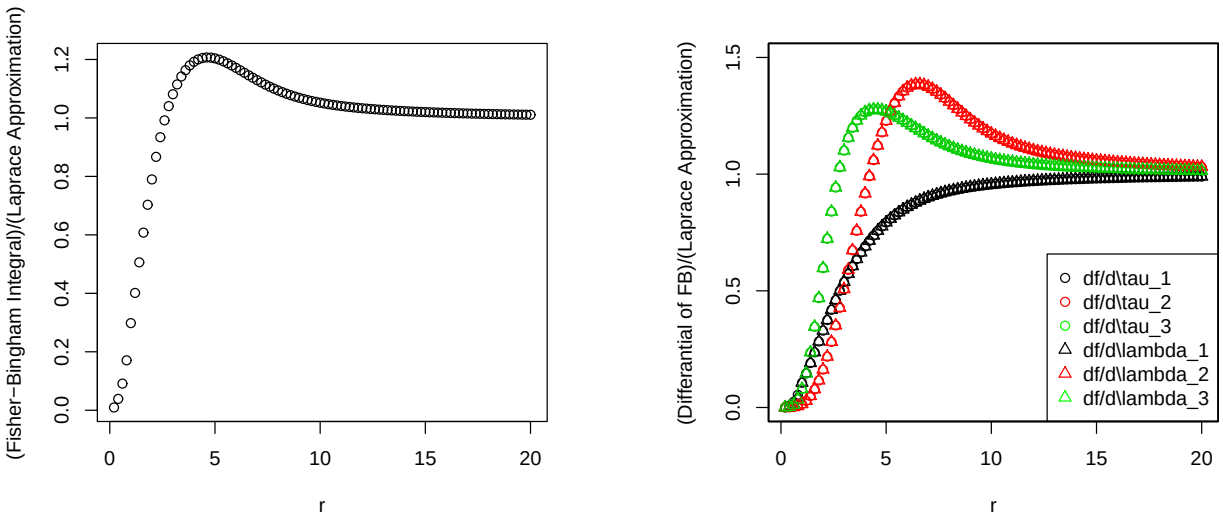


Figure 2: Ratios to the Laplace approximations

In our second example we consider diagonal matrices $\Sigma^{(1)}$ and $\Sigma^{(2)}$ with diagonal elements

$$(\sigma_i^2)^{(1)} = \frac{d+1}{k(k+1)} \quad (1 \le k \le d), \tag{24}$$

and

$$(\sigma_i^2)^{(2)} = \frac{2(d+2)(d+3)}{k(k+1)(k+2)(k+3)} \quad (1 \le k \le d), \tag{25}$$

respectively. These weights are considered for cumulative chi-square statistics in [5]. Let

$$\mu^{(1)} = 0,$$
$$\mu^{(2)} = \begin{pmatrix} 0 & 0.01 & 0.02 & \cdots & 0.01 \times (d-1) \end{pmatrix}^\top.$$

For each dimension d, we computed the probability $P\left(1.0 \times 10^{-6} \le \sum_{i=1}^d X_i^2 < 40.0\right)$ and measured the computational times in seconds. We considered the following four patterns of parameters:

$$(\Sigma^{(1)}, \mu^{(1)}), \qquad (\Sigma^{(1)}, \mu^{(2)}),$$
$$(\Sigma^{(2)}, \mu^{(1)}), \qquad (\Sigma^{(2)}, \mu^{(2)}).$$

The experimental results are shown in Table 1.

As our third example we consider how our method works for large dimension. Corresponding to the asymptotic null distribution of Anderson-Darling statistic, which is an infinite sum of weighted χ^2 variables, consider the weights

$$\sigma_k^2 = \frac{1}{k(k+1)}, \quad \mu_k = 0 \quad (1 \le k \le d).$$

Here we truncate the infinite series at d. We computed the probability and measured its computational time. We fixed the radius as $r = 20.0$. The results on the computational time are shown in Table 2 and its figure. Even for $d = 100$, our method is accurate and fast enough to be practical. This is a remarkable progress since the implementation of HGM in [7] can compute only up to dimension $d = 8$.

5 Summary and discussion

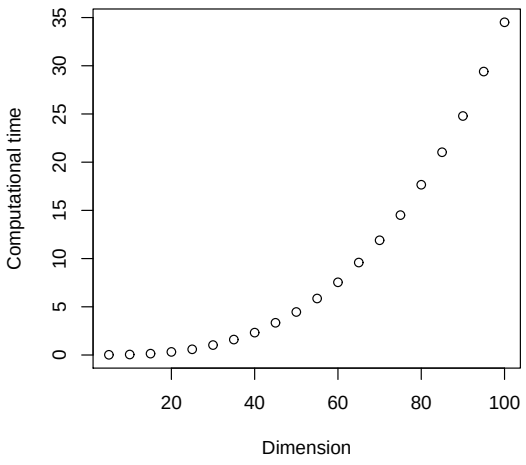
In this paper we applied HGM method for computing distribution function of a weighted sum of independent noncentral chi-square random variables. We found that our method is numerically

Table 1: Accuracy and computational times for $\Sigma^{(1)}$ and $\Sigma^{(2)}$

dimension	$\Sigma^{(1)}$				$\Sigma^{(2)}$			
	$\mu = 0$		$\mu \neq 0$		$\mu = 0$		$\mu \neq 0$	
	p	times(s)	p	times(s)	p	times(s)	p	times(s)
10	1.000000	0.02	1.000000	0.03	1.000000	0.11	1.000000	0.10
11	1.000000	0.03	1.000000	0.04	1.000000	0.12	1.000000	0.13
12	1.000000	0.04	1.000000	0.04	1.000000	0.16	1.000000	0.16
13	1.000000	0.05	1.000000	0.04	1.000000	0.19	1.000000	0.19
14	1.000000	0.04	1.000000	0.05	1.000000	0.24	1.000000	0.24
15	1.000000	0.05	1.000000	0.06	1.000000	0.28	1.000000	0.30
16	1.000000	0.06	1.000000	0.08	1.000000	0.34	1.000000	0.35
17	1.000000	0.06	1.000000	0.07	1.000000	0.41	1.000000	0.43
18	1.000000	0.09	1.000000	0.09	1.000000	0.48	1.000000	0.49
19	1.000000	0.08	1.000000	0.10	1.000000	0.56	1.000000	0.58
20	1.000000	0.09	1.000000	0.10	1.000000	0.64	1.000000	0.66

Table 2: Computational times for Anderson-Darling statistic

dim	p	time(s)
30	1.000000	1.03
35	1.000000	1.60
40	1.000000	2.32
45	1.000000	3.34
50	1.000000	4.46
55	1.000000	5.86
60	1.000000	7.54
65	1.000000	9.59
70	1.000000	11.90
75	1.000000	14.51
80	1.000000	17.66
85	1.000000	21.03
90	1.000000	24.79
95	1.000000	29.40
100	1.000000	34.51



Graph of computational times

both accurate and fast, after we implemented the following ideas. First, during the application of Runge-Kutta method, we re-scaled the vector F in (5) as needed to keep its elements within the precision for floating point numbers. Also we divided the interval for integration into $(0, 1]$ and $[1, \infty)$ and switched from F to Q in (9) in view of the asymptotic values for Q. Our experience in this paper shows that re-scaling of the standard monomials is important in numerical implementation of HGM.

In our implementation, the numerical integration starts from a small $r = r_0 > 0$ and the integration proceeds to $r = \infty$. On the other hand, we have asymptotic results for large r in Section 3. Then we might consider reversing the direction of integration and start with initial values at very large r. We may call the former the "forward integration" and the latter the "backward integration". However we found that the backward integration is not numerically stable. Hence the asymptotic values can not be used as initial values. In this paper we used the asymptotic values just for checking the accuracy HGM in the forward direction.

It is an interesting question, whether the asymptotic values can be used to adjust the values of the forward integration. We may look at the difference between F by forward HGM and its asymptotic value for very large r and use the difference to adjust F at intermediate values of r. However it is not clear how this adjustment can be implemented.

A A general form of Proposition 3.1

In Proposition 3.1 we assumed $\lambda_1 > \lambda_2$. In this appendix we state the following proposition for the general case without a proof.

Proposition A.1. *Assume that*

$$0 > \lambda_1 = \cdots = \lambda_m > \lambda_{m+1} \geq \cdots \geq \lambda_d.$$

If $0 = \tau_1 = \cdots = \tau_m$, *then as* $r \to \infty$,

$$f(\lambda, \tau, r) = r^{m-1} S_{m-1} \exp\left(r^2 \lambda_1 - \sum_{i=m+1}^{d} \frac{\tau_i^2}{4(\lambda_i - \lambda_1)} \right) \frac{\pi^{(d-m)/2}}{\prod_{i=m}^{d} (\lambda_1 - \lambda_i)^{1/2}} (1 + o(1)),$$

$$\partial_{\lambda_j} f(\lambda, \tau, r) = \frac{r^2}{m} f(\lambda, \tau, r)(1 + o(1)), \quad j \leq m,$$

$$\partial_{\tau_j} f(\lambda, \tau, r) = 0, \quad j \leq m,$$

$$\partial_{\tau_j} f(\lambda, \tau, r) = -\frac{\tau_j}{2(\lambda_j - \lambda_1)} f(\lambda, \tau, r)(1 + o(1)), \quad j > m,$$

$$\partial_{\lambda_j} f(\lambda, \tau, r) = \left(\frac{1}{2(\lambda_1 - \lambda_j)} + \frac{\tau_j^2}{4(\lambda_j - \lambda_1)^2} \right) f(\lambda, \tau, r)(1 + o(1)), \quad j > m.$$

If $(\tau_1, \ldots, \tau_m) \neq (0, \ldots, 0)$, *define* $\gamma = (\tau_1^2 + \cdots + \tau_m^2)^{1/2}$. *Then, as* $r \to \infty$,

$$f(\lambda, \tau, r) = \exp\left(r^2 \lambda_1 + r\gamma - \sum_{i=m+1}^{d} \frac{\tau_i^2}{4(\lambda_i - \lambda_1)} \right) \left(\frac{2r}{\gamma} \right)^{(m-1)/2} \frac{\pi^{(d-1)/2}}{\prod_{i=m}^{d} (\lambda_1 - \lambda_i)^{1/2}} (1 + o(1)),$$

$$\partial_{\tau_j} f(\lambda, \tau, r) = r \frac{\tau_j}{\gamma} f(\lambda, \tau, r)(1 + o(1)), \quad \tau_j \neq 0, \ j \leq m,$$

$$\partial_{\lambda_j} f(\lambda, \tau, r) = r^2 \frac{\tau_j^2}{\gamma^2} f(\lambda, \tau, r)(1 + o(1)), \quad \tau_j \neq 0, \ j \leq m,$$

$$\partial_{\tau_j} f(\lambda,\tau,r) = 0, \quad \tau_j = 0,\ j \le m,$$
$$\partial_{\lambda_j} f(\lambda,\tau,r) = \frac{r}{\gamma} f(\lambda,\tau,r)(1+o(1)), \quad \tau_j = 0,\ j \le m,$$
$$\partial_{\tau_j} f(\lambda,\tau,r) = -\frac{\tau_j}{2(\lambda_j-\lambda_1)} f(\lambda,\tau,r)(1+o(1)), \quad j > m,$$
$$\partial_{\lambda_j} f(\lambda,\tau,r) = \left(\frac{1}{2(\lambda_1-\lambda_j)} + \frac{\tau_j^2}{4(\lambda_j-\lambda_1)^2}\right) f(\lambda,\tau,r)(1+o(1)), \quad j > m.$$

References

[1] T. W. Anderson and D. A. Darling. Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *Ann. Math. Statistics*, 23:193–212, 1952.

[2] A. Castaño-Martínez and F. López-Blázquez. Distribution of a sum of weighted noncentral chi-square variables. *Test*, 14(2):397–415, 2005.

[3] A. M. Cohen. *Numerical Methods for Laplace Transform Inversion*, volume 5 of *Numerical Methods and Algorithms*. Springer, New York, 2007.

[4] R. B. D'Agostino and M. A. Stephens, editors. *Goodness-of-Fit Techniques*, volume 68 of *Statistics: Textbooks and Monographs*. Marcel Dekker Inc., New York, 1986.

[5] C. Hirotsu. Cumulative chi-squared statistic as a tool for testing goodness of fit. *Biometrika*, 73(1):165–173, 1986.

[6] T. Koyama. A holonomic ideal which annihilates the Fisher-Bingham integral. *Funkcial. Ekvac.*, 56(1):51–61, 2013.

[7] T. Koyama, H. Nakayama, K. Nishiyama, and N. Takayama. Holonomic gradient descent for the Fisher-Bingham distribution on the d-dimensional sphere. *Computational Statistics*, 29(3-4):661–683, 2014.

[8] T. Koyama, H. Nakayama, K. Nishiyama, and N. Takayama. The holonomic rank of the Fisher-Bingham system of differential equations. *J. Pure Appl. Algebra*, 218(11):2060–2071, 2014.

[9] A. Kume and S. G. Walker. On the Fisher-Bingham distribution. *Stat. Comput.*, 19(2):167–172, 2009.

[10] A. Kume and A. T. A. Wood. Saddlepoint approximations for the Bingham and Fisher-Bingham normalising constants. *Biometrika*, 92(2):465–476, 2005.

[11] V. N. Nair. On testing against ordered alternatives in analysis of variance models. *Biometrika*, 73(2):493–499, 1986.

[12] H. Nakayama, K. Nishiyama, M. Noro, K. Ohara, T. Sei, N. Takayama, and A. Takemura. Holonomic gradient descent and its application to the Fisher-Bingham integral. *Advances in Applied Mathematics*, 47:639–658, 2011.

[13] T. Sei and A. Kume. Calculating the normalising constant of the Bingham distribution on the sphere using the holonomic gradient method. *Stat. Comput.*, 25(2):321–332, 2015.

[14] K. Tanaka. *Time Series Analysis: Nonstationary and Noninvertible Distribution Theory.* Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, 1996.