Mathematical
Inequalities
& Applications
Volume 12, Number 4 (2009), 839–844

A BIENAYMÉ–CHEBYSHEV INEQUALITY FOR SCALE MIXTURES OF THE MULTIVARIATE NORMAL DISTRIBUTION

VILLŐ CSISZÁR AND TAMÁS F. MÓRI

Dedicated to Professor Josip Pečarić
on the occasion of his 60th birthday

Abstract. In this short note a Chebyshev type sharp upper bound is presented for the tail probability of scale mixtures of the zero mean multivariate normal distribution, only in terms of the variance. Similar estimation is proved for the probability content of an arbitrary ellipsoid containing the origin.

1. Introduction

The classical Bienaymé–Chebyshev inequality says that

$$P(|X - E(X)| \geqslant t) \leqslant \frac{\mathbf{Var}(X)}{t^2}$$

holds for every positive t. This simple inequality is valid for all univariate random variables X with finite variance, but in particular cases, when more is known about X, it can be improved significantly. For instance, there exist sharper inequalities for random variables having unimodal distribution [14], or finite higher moments, and also in the case where X is a sum of n i.i.d. random variables [1], etc. Efforts have also been made to construct similar inequalities for random vectors, see e.g. [2, 9, 10]. In a recent paper [13] sharp lower bounds are presented on the probability of a set defined by quadratic inequalities, given the first two moments of the distribution. Though the bounds are not explicite, they can be efficiently computed using convex optimization.

Let $Z = (Z_1, \dots, Z_n)^\top$ be an n-variate standard normal vector, and $\mathbf{S}$ a symmetric, positive semidefinite random matrix of size $n \times n$, independent of Z. Scale mixtures of the n-variate standard normal distribution are defined as distributions of random vectors X of the form $X = \mathbf{S}^{1/2} Z$.

Scale mixtures of multivariate normal distributions are widely applied in many fields, though in most cases $\mathbf{S}$ is a random scalar multiple of a fixed positive semidefinite matrix [3, 4, 6, 7, 8, 11]. If $\mathbf{S}$ is not of that simpler form, the mixture is no longer elliptically contoured: the equiprobability contours are star-convex from the origin, but not convex in general.

Mathematics subject classification (2000): 60E15.
Keywords and phrases: Chebyshev inequality, scale mixtures, multivariate normal distribution.
This research has been supported by the Hungarian National Foundation for Scientific Research, Grant No. K 67961.

Paper MIA-12-67

Clearly, $EX = 0$, and $\mathbf{Var}(X) = E\mathbf{S} = \Sigma$. We want to estimate the tail probability $P\big(\|X\| \geqslant t\big)$, for $t > 0$, similarly to the univariate case. One can apply Markov's inequality to $\|X\|^2$ to obtain

$$P\big(\|X\| \geqslant t\big) = P\big(\|X\|^2 \geqslant t^2\big) \leqslant \frac{E\|X\|^2}{t^2} = \frac{\operatorname{tr}\Sigma}{t^2}. \tag{1}$$

While (1) is valid for all random vectors with zero mean and finite variance, it may turn out to be weaker than optimal when X is known to possess a particular structure. Therefore, we are lookig for the smallest constant $C = C_n$ for which

$$P\big(\|X\| \geqslant t\big) \leqslant C \cdot \operatorname{tr}\Sigma \cdot t^{-2} \tag{2}$$

holds for every positive t, whenever the distribution of X is a scale mixture of the n-variate standard normal distribution.

The aim of the present note is twofold. First, we determine the value of C_n, then we extend inequality (2) to the probability content of ellipsoids containing the origin.

2. Results

First we characterize the optimal constant C_n.

THEOREM 1.

$$C_n = \max_{t>0} tP(V_n \geqslant t),$$

where the distribution of V_n is gamma, with parameters $n/2$, $n/2$ (see Table 1).

n	C_n	n	C_n	n	C_n	n	C_n
1	0.33143	6	0.45703	12	0.52803	30	0.62656
2	0.36788	7	0.47214	14	0.54457	40	0.65653
3	0.39649	8	0.48560	16	0.55899	50	0.67905
4	0.41998	9	0.49771	18	0.57174	100	0.74381
5	0.43985	10	0.50871	20	0.58314	1000	0.89215

Table 1. Value of the optimal constant in n dimensions.

Proof. Let us first suppose that $\mathbf{S}$ is a constant (i.e., non-random) matrix. Let $\mathbf{S} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$ be the spectral decomposition of $\mathbf{S}$, where $\mathbf{\Lambda} = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ is a diagonal matrix with the eigenvalues of $\mathbf{S}$ in the main diagonal, and $\mathbf{U}$ is an orthonormal matrix (i.e., $\mathbf{U}^{-1} = \mathbf{U}^\top$). Then

$$\|X\|^2 = Z^\top \mathbf{S} Z = Z^\top \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top Z = \lambda_1 W_1^2 + \ldots + \lambda_n W_n^2,$$

where the distribution of $W = (W_1, \ldots, W_n)^\top = \mathbf{U}^\top Z$ is n-variate standard normal, that is, $W_1, \ldots, W_n$ are i.i.d. $N(0,1)$ random variables. Let

$$C_n = \max\Big\{ tP\big(\lambda_1 W_1^2 + \ldots + \lambda_n W_n^2 \geqslant t\big) : \\ \lambda_1, \ldots, \lambda_n \geqslant 0,\ \lambda_1 + \ldots + \lambda_n = 1,\ t > 0 \Big\}, \tag{3}$$

then

$$P\big(\|X\| \geqslant t\big) = P\big(\|X\|^2 \geqslant t^2\big) = P\big(\lambda_1 W_1^2 + \ldots + \lambda_n W_n^2 \geqslant t^2\big)$$
$$\leqslant C_n \frac{\lambda_1 + \ldots + \lambda_n}{t^2} = C_n \frac{\operatorname{tr}\mathbf{S}}{t^2}.$$

When $\mathbf{S}$ is a random matrix, the same inequality holds for the *conditional* probability:

$$P\big(\|\mathbf{S}^{1/2}Z\| \geqslant t \mid \mathbf{S}\big) \leqslant C_n \frac{\operatorname{tr}\mathbf{S}}{t^2}.$$

By taking expectations on both sides, and remembering that trace and expectation can be interchanged, we obtain

$$P(\|X\| \geqslant t) \leqslant C_n \frac{\operatorname{tr}\Sigma}{t^2}. \tag{4}$$

Thus, the optimal constant C_n in (4) is given by (3).

Let t be fixed, and $\lambda_1, \ldots, \lambda_n$ vary, in such a way that $P\big(\lambda_1 W_1^2 + \ldots + \lambda_n W_n^2 \geqslant t\big)$ is maximal. By Theorem 1 of Székely and Bakirov [12] it follows that all non-zero coefficients λ_i must be equal. Hence

$$C_n = \max_{i \leqslant n} \max_{t>0} tP(V_i \geqslant t),$$

where the distribution of V_i is gamma with parameters $i/2$, $i/2$.

Since $EV_i = 1$, $\max\limits_{t>0} tP(V_i \geqslant t)$ is equal to the minimal value of constant κ with which

$$P\big(|X| \geqslant x\big) \leqslant \kappa E|X|/x$$

holds for every positive x and arbitrary symmetric random variable X of the form $|X| = \sigma V_i$, where the random variable $\sigma > 0$ is independent of V_i (see Remark 2 of [5]). In terms of [5], this κ is the Chebyshev constant associated with $\alpha = 1$ in the family of scale mixtures of the two-sided gamma distribution with parameter $\nu = \frac{i}{2} - 1$ (Example 4 of [5]). By Theorem 4 of [5], the Chebyshev constant is an increasing function of the parameter ν, that is, of i. The proof is completed. □

The probability on the left-hand side of (1) can also be written in the form $P(X \notin \mathscr{M})$, where $\mathscr{M} \subset \mathbb{R}^n$ is the open ball with center 0 and radius t. Let us extend inequality (1) to arbitrary ellipsoids containing the origin.

THEOREM 2. *Let $\mathbf{A}$ be a positive definite $n \times n$ matrix, $a \in \mathbb{R}^n$, and $t > 0$. Define*

$$\mathscr{M} = \{x \in \mathbb{R}^n : (x-a)^\top \mathbf{A}\,(x-a) < t^2\};$$

an ellipsoid with center a. Suppose that $0 \in \mathscr{M}$, that is, $a^\top \mathbf{A}\, a < t^2$. Let the distribution of the random vector X be a scale mixture of the n-variate standard normal distribution, with variance Σ. Then

$$P(X \notin \mathscr{M}) \leqslant C_n \operatorname{tr}(\Sigma\mathbf{A})\, t^{-2}, \quad \textit{if } a = 0, \tag{5}$$

$$P(X \notin \mathscr{M}) \leqslant \frac{C_n}{t^2 - a^\top \mathbf{A}\, a}\left(\operatorname{tr}(\Sigma\mathbf{A})\, \frac{t}{t - \sqrt{a^\top \mathbf{A}\, a}} - \frac{a^\top \mathbf{A}\Sigma\mathbf{A}^\top a}{a^\top \mathbf{A}\, a + t\sqrt{a^\top \mathbf{A}\, a}}\right), \quad \textit{if } a \neq 0, \tag{6}$$

where C_n is given in Theorem 1.

Proof. Let us first suppose that $a = 0$. Clearly, $X \notin \mathscr{M}$ if and only if $\|Y\| \geqslant t$, where $Y = \mathbf{A}^{1/2}X = (\mathbf{SA})^{1/2}Z$. Thus Y is a scale transform of the n-variate normal distribution again. Since $E(\mathbf{SA}) = \Sigma\mathbf{A}$, inequality (5) immediately follows from (4).

In the general case we first show that $\mathscr{M} \supset \mathscr{M}_1 \cup \mathscr{M}_2$, where

$$\mathscr{M}_1 = \Big\{x \in \mathbb{R}^n : a^\top \mathbf{A} x < 0,\ x^\top \mathbf{A} x < \big(t - \sqrt{a^\top \mathbf{A} a}\big)^2\Big\},$$
$$\mathscr{M}_2 = \Big\{x \in \mathbb{R}^n : a^\top \mathbf{A} x \geqslant 0,\ x^\top (\mathbf{A} - \mathbf{B}) x < t^2 - a^\top \mathbf{A} a\Big\},$$

with $\mathbf{B} = \dfrac{2\mathbf{A} a a^\top \mathbf{A}}{a^\top \mathbf{A} a + t\sqrt{a^\top \mathbf{A} a}}$.

$\mathscr{M}_1$ and $\mathscr{M}_2$ are obviously disjoint. Both are halves of ellipsoids centered at the origin. The first one is homothetic to $\mathscr{M}$. In order to show what the second one is, let us check that $\mathbf{A} - \mathbf{B}$ is positive definite. Let $x \in \mathbb{R}^n$ be different from 0. Using that $a \in \mathscr{M}$ we have

$$x^\top \mathbf{B} x = \frac{2\big(a^\top \mathbf{A} x\big)^2}{a^\top \mathbf{A} a + t\sqrt{a^\top \mathbf{A} a}} < \frac{\big(a^\top \mathbf{A} x\big)^2}{a^\top \mathbf{A} a} \leqslant x^\top \mathbf{A} x.$$

Suppose $x \in \mathscr{M}_1$. Then by using the triangle inequality we get

$$(x-a)^\top \mathbf{A}(x-a) \leqslant \Big(\sqrt{x^\top \mathbf{A} x} + \sqrt{a^\top \mathbf{A} a}\Big)^2 < t^2.$$

On the other hand, let $x \in \mathscr{M}_2$. Then

$$(x-a)^\top \mathbf{A}(x-a) = x^\top \mathbf{A} x - 2a^\top \mathbf{A} x + a^\top \mathbf{A} a < t^2 + x^\top \mathbf{B} x - 2a^\top \mathbf{A} x,$$

therefore it suffices to show that $x^\top \mathbf{B} x \leqslant 2a^\top \mathbf{A} x$, that is,

$$(a^\top \mathbf{A} x)^2 \leqslant (a^\top \mathbf{A} x)\big(a^\top \mathbf{A} a + t\sqrt{a^\top \mathbf{A} a}\big). \tag{7}$$

Let us start from the inequality

$$\big(a^\top (\mathbf{A}-\mathbf{B}) x\big)^2 \leqslant \big(a^\top (\mathbf{A}-\mathbf{B}) a\big)\big(x^\top (\mathbf{A}-\mathbf{B}) x\big) \leqslant \big(a^\top (\mathbf{A}-\mathbf{B}) a\big)\big(t^2 - a^\top \mathbf{A} a\big). \tag{8}$$

Here

$$a^\top (\mathbf{A}-\mathbf{B}) x = \big(a^\top \mathbf{A} x\big)\,\frac{t - \sqrt{a^\top \mathbf{A} a}}{t + \sqrt{a^\top \mathbf{A} a}}. \tag{9}$$

Let us apply (9) to both sides of (8). Then we get

$$\big(a^\top \mathbf{A} x\big)^2 \leqslant \frac{t + \sqrt{a^\top \mathbf{A} a}}{t - \sqrt{a^\top \mathbf{A} a}}\,\big(a^\top \mathbf{A} a\big)\big(t^2 - a^\top \mathbf{A} a\big) = \Big(a^\top \mathbf{A} a + t\sqrt{a^\top \mathbf{A} a}\Big)^2, \tag{10}$$

which is equivalent to (7). Thus $\mathscr{M}_2 \subset \mathscr{M}$.

Since scale transforms of the multivariate normal distributions are diagonally symmetric (that is, X and $-X$ have the same distribution), we have

$$P(X \in \mathscr{M}_1) = \frac{1}{2} P\left(X^\top \mathbf{A} X < \left(t - \sqrt{a^\top \mathbf{A} a}\right)^2\right) \geqslant \frac{1}{2}\left(1 - C_n \cdot \frac{\operatorname{tr}(\Sigma \mathbf{A})}{\left(t - \sqrt{a^\top \mathbf{A} a}\right)^2}\right),$$

and similarly,

$$P(X \in \mathscr{M}_2) = \frac{1}{2} P\left(X^\top (\mathbf{A} - \mathbf{B}) X < t^2 - a^\top \mathbf{A} a\right) \geqslant \frac{1}{2}\left(1 - C_n \cdot \frac{\operatorname{tr}(\Sigma(\mathbf{A} - \mathbf{B}))}{t^2 - a^\top \mathbf{A} a}\right).$$

Using the cyclic invariance of the trace operator we can write

$$\operatorname{tr}(\Sigma \mathbf{B}) = \frac{2 \operatorname{tr}(\Sigma \mathbf{A} a a^\top \mathbf{A})}{a^\top \mathbf{A} a + t\sqrt{a^\top \mathbf{A} a}} = \frac{2 a^\top \mathbf{A} \Sigma \mathbf{A} a}{a^\top \mathbf{A} a + t\sqrt{a^\top \mathbf{A} a}}.$$

Hence,

$$P(X \in \mathscr{M}) \geqslant 1 - C_n \operatorname{tr}(\Sigma \mathbf{A}) \frac{1}{2}\left(\frac{1}{t^2 - a^\top \mathbf{A} a} + \frac{1}{\left(t - \sqrt{a^\top \mathbf{A} a}\right)^2}\right) + C_n \frac{a^\top \mathbf{A} \Sigma \mathbf{A}^\top a}{a^\top \mathbf{A} a + t\sqrt{a^\top \mathbf{A} a}} \cdot \frac{1}{t^2 - a^\top \mathbf{A} a},$$

which is tantamount to (6). □

REFERENCES

[1] P. J. BICKEL AND A. M. KRIEGER, *Extensions of Chebyshev's inequality with applications*, Probab. Math. Statist., **13** (1992), 293–410.

[2] Z. W. BIRNBAUM AND A. W. MARSHALL, *Some multivariate Chebyshev inequalities with extensions to continuous parameter processes*, Ann. Math. Statist., **32** (1961), 687–703.

[3] S. CAMBANIS, S. B. FOTOPOULOS AND L. HE, *On the conditional variance for scale mixtures of normal distributions*, J. Multivar. Anal., **74** (2000), 163–192.

[4] Z. CHEN AND L. KUO, *Discrete choice models based on the scale mixture of multivariate normal distributions*, Sankhyā Ser. B, **64** (2002), 192–213.

[5] V. CSISZÁR, T. F. MÓRI, AND G.J. SZÉKELY, *Chebyshev-type inequalities for scale mixtures*, Statist. Probab. Lett., **71** (2005), 323-335.

[6] T. ELTOFT, T. KIM AND T. W. LEE, *Multivariate Scale Mixture of Gaussians Modeling*, In: Independent Component Analysis and Blind Signal Separation, Lecture Notes in Computer Science **3889** (2006), 799–806.

[7] S. B. FOTOPOULOS, AND V. K. JANDHYALA, *Bessel inequalities with applications to conditional log-returns under GIG scale mixtures of normal vectors*, Statist. Probab. Lett., **66** (2004), 117–125.

[8] B. GRIGELIONIS, *On Pólya mixtures of multivariate Gaussian distributions*, Statist. Probab. Lett., **78** (2008), 1459–1465.

[9] A. MARSHALL AND I. OLKIN, *Multivariate Chebyshev inequalities*, Ann. Math. Statist., **31** (1960), 1001–1014.

[10] I. OLKIN AND J. W. PRATT, *A multivariate Tchebycheff inequality*, Ann. Math. Statist., **29** (1958), 226–234.

[11] R. SHIMIZU, *Expansion of the scale mixture of the multivariate normal distribution with error bound evaluated in the L_1-norm*, J. Multivar. Anal., **53** (1995), 126–138.

[12] G. J. SZÉKELY AND N. K. BAKIROV, *Extremal probabilities for Gaussian quadratic forms*, Probab. Theory Related Fields, **126** (2003), 184–202.
[13] L. VANDENBERGHE, S. BOYD AND K. COMANOR, *Generalized Chebyshev bounds via semidefinite programming*, SIAM Rev., **49** (2007), 52–64 (electronic).
[14] D. F. VYSOCHANSKIĬ AND Y. I. PETUNIN, *Justification of the three sigma rule for unimodal distributions*, Theory Probab. Math. Statist., **21** (1980), 25–36.

(Received October 31, 2008)

Villő Csiszár
Department of Probability Theory and Statistics
Loránd Eötvös University
Pázmány P. s. 1/C
H-1117 Budapest
Hungary
e-mail: villo@ludens.elte.hu

Tamás F. Móri
Department of Probability Theory and Statistics
Loránd Eötvös University
Pázmány P. s. 1/C
H-1117 Budapest
Hungary
e-mail: moritamas@ludens.elte.hu