Annales Univ. Sci. Budapest., Sect. Comp. 42 (2014) 109–125

EXPLICIT MULTIVARIATE BOUNDS OF CHEBYSHEV TYPE

Villő Csiszár (Budapest, Hungary)

Tamás Fegyverneki (Budapest, Hungary)

Tamás F. Móri (Budapest, Hungary)

Dedicated to Professor András Benczúr on the occasion of his 70th birthday

Communicated by László Lakatos

(Received June 1, 2014; accepted July 1, 2014)

Abstract. In the paper, Chebyshev type upper bounds are presented for the probability that a random vector X falls outside of an ellipsoid, in terms of the mean and variance of X. Our inequalities are better than those found in the literature.

1. Introduction

The classical Bienaymé–Chebyshev inequality says that

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq t) \leq \frac{\mathbb{V}\mathrm{ar}(X)}{t^2}$$

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

---

Key words and phrases: Chebyshev inequality, Selberg inequality, diagonally symmetric distribution

2010 Mathematics Subject Classification: 60E15

The research was supported by the Hungarian Scientific Research Fund under grant OTKA K108615.

https://doi.org/10.71352/ac.42.109

is known about X, it can be improved significantly. For instance, there exist sharper inequalities for bounded random variables [11], or random variables having unimodal distribution [13], 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, 4, 6, 7, 8, 9]. In a recent paper [12] 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 explicit, they can be efficiently computed using convex optimization.

If the distribution of the random vector X is known to be a member of a certain family, sharper bounds can be constructed. This is the case, for example, when the distribution is a scale mixture of the n-variate standard normal distribution, that is, $X = \mathbf{S}^{1/2}Z$, where Z is an n-variate standard normal vector, and $\mathbf{S}$ is a symmetric, positive semidefinite random matrix of size $n \times n$, independent of Z. In [3] a Chebyshev type inequality is proved for the probability $\mathbb{P}(X \notin \mathcal{C})$, where $\mathcal{C}$ is an ellipsoid containing the origin.

In the present paper, we do not suppose any particular property of the distribution except that $\mathbb{E}X = 0$ and $\mathbb{V}\mathrm{ar}(X) = \mathbf{\Sigma} > 0$, and we construct a relatively good upper bound for $\mathbb{P}(X \notin \mathcal{C})$. Our approach is somewhat similar to what Olkin and Pratt [9] applied to the probability content of higher dimensional rectangles (cuboids). Then we also have a look at diagonally symmetric distributions.

Our inequalities are presented in Section 2. Section 3 contains the proofs. In Section 4, we compare our inequalities with earlier results through some numerical examples.

2. Results

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

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

an ellipsoid with center a. We will suppose that $0 \in \mathcal{C}$, that is, $a^\top \mathbf{A}\, a < t^2$.

Let X be an n-variate random vector, $\mathbb{E}X = 0$, and $\mathbb{V}\mathrm{ar}(X) = \mathbb{E}(XX^\top) = \mathbf{\Sigma}$. Suppose that $\mathbf{\Sigma}$ is positive definite.

A well known simple estimation can be obtained in the following way. Let us apply the Markov inequality to the random variable $(X-a)^\top \mathbf{A}(X-a)$.

Then

$$\mathbb{P}(X \notin \mathcal{C}) \leq \frac{\mathbb{E}(X-a)^\top \mathbf{A}(X-a)}{t^2}. \tag{2.1}$$

Using the cyclic invariance of the trace function we can write

$$\begin{aligned}\mathbb{E}(X-a)^\top \mathbf{A}(X-a) &= \mathbb{E}\operatorname{tr}\big((X-a)^\top \mathbf{A}(X-a)\big)\\ &= \mathbb{E}\operatorname{tr}\big(\mathbf{A}(X-a)(X-a)^\top\big) = \operatorname{tr}\big(\mathbf{A}\,\mathbb{E}(X-a)(X-a)^\top\big)\\ &= \operatorname{tr}\big(\mathbf{A}\big(\boldsymbol{\Sigma}+aa^\top\big)\big) = \operatorname{tr}\mathbf{A}\boldsymbol{\Sigma} + \operatorname{tr}\big(\mathbf{A}aa^\top\big) = \operatorname{tr}\mathbf{A}\boldsymbol{\Sigma} + a^\top \mathbf{A}\,a.\end{aligned}$$

Plugging this back into (2.1) we have

$$\mathbb{P}(X \notin \mathcal{C}) \leq \frac{\operatorname{tr}\mathbf{A}\boldsymbol{\Sigma} + a^\top \mathbf{A}\,a}{t^2}. \tag{2.2}$$

If ellipsoid $\mathcal{C}$ *does not* contain the origin, the right-hand side of (2.2) is greater than 1. In fact, no upper estimate, apart from the trivial one, exists for $\mathbb{P}(X \notin \mathcal{C})$. In that case the complementary probability can be estimated in the following way [7, Corollary 5.1]

$$\mathbb{P}(X \in \mathcal{C}) \leq \sup_{z\in\mathcal{C}} \frac{1}{1+z^\top \boldsymbol{\Sigma}^{-1} z}.$$

Suppose the center of the ellipsoid coincides with the expectation of the random vector, i.e, $a = 0$. In that case, if the upper bound of (2.2) is less than 1, then there exists a distribution with mean 0 and covariance matrix $\boldsymbol{\Sigma}$ for which the bound is exact. This remains true in some cases where $a \neq 0$, but then ellipsoid $\mathcal{C}$ should satisfy certain conditions depending on $\boldsymbol{\Sigma}$. They are as follows.

Consider $\mathbf{B} = \mathbf{A}^{1/2}\big(\boldsymbol{\Sigma} + aa^\top\big)\mathbf{A}^{1/2}$. This is a symmetric positive definite matrix, and $\operatorname{tr}\big(t^{-2}\mathbf{B}\big)$ is just the right-hand side of (2.2). The spectral decomposition of the matrix $t^{-2}\mathbf{B}$ is of the form

$$t^{-2}\mathbf{B} = \sum_{j=1}^{n} \lambda_j z_j z_j^\top,$$

where $\{z_j : j = 1,\dots,n\}$ is an orthonormal basis consisting of eigenvectors, and λ_j, $j = 1,\dots,n$, are the corresponding eigenvalues (they are positive). Then the right-hand side of (2.2) is less than 1 if and only if

$$\lambda_1 + \cdots + \lambda_n < 1. \tag{2.3}$$

Let us express $t^{-1}\mathbf{A}^{1/2}a$ in this orthonormal basis,

$$t^{-1}\mathbf{A}^{1/2}a = \sum_{j=1}^{n} \alpha_j z_j.$$

Theorem 2.1. *Let* $\mathbf{\Sigma}$, $\mathbf{A}$, a *and* t *be given in such a way that they satisfy* (2.3) *and also condition*

$$|\alpha_j| \le \lambda_j, \quad j = 1, \ldots, n. \tag{2.4}$$

Then there exists a random vector X *with mean* 0 *and covariance matrix* $\mathbf{\Sigma}$ *for which* (2.2) *holds true with equality.*

Note that $\alpha_j^2 \le \lambda_j$ by definition, since

$$\lambda_j = z_j^\top \big(t^{-2}\mathbf{B}\big) z_j = t^{-2} z_j^\top \mathbf{\Sigma} z_j + \alpha_j^2 \ge \alpha_j^2.$$

Condition (2.4) is clearly satisfied for $a = 0$.

For $a \ne 0$ inequality (2.2) may be rather crude. The following theorems provide estimations improving (2.2) in certain cases.

Theorem 2.2.

$$\mathbb{P}(X \notin \mathcal{C}) \le \begin{cases} \dfrac{\operatorname{tr}\mathbf{A\Sigma}}{\operatorname{tr}\mathbf{A\Sigma} + \big(t - \sqrt{a^\top \mathbf{A}\, a}\,\big)^2}, & \textit{if } \operatorname{tr}\mathbf{A\Sigma} \le t\sqrt{a^\top \mathbf{A}\, a} - a^\top \mathbf{A}\, a, \\ \dfrac{\operatorname{tr}\mathbf{A\Sigma} + a^\top \mathbf{A}\, a}{t^2}, & \textit{if } \operatorname{tr}\mathbf{A\Sigma} > t\sqrt{a^\top \mathbf{A}\, a} - a^\top \mathbf{A}\, a. \end{cases} \tag{2.5}$$

This is a direct multivariate generalization of the univariate Selberg inequality [12, Sec. 3]. Note that

$$\frac{\operatorname{tr}\mathbf{A\Sigma} + a^\top \mathbf{A}\, a}{t^2} = \frac{\operatorname{tr}\mathbf{A\Sigma}}{\operatorname{tr}\mathbf{A\Sigma} + \big(t - \sqrt{a^\top \mathbf{A}\, a}\,\big)^2} + \frac{\big(\operatorname{tr}\mathbf{A\Sigma} - t\sqrt{a^\top \mathbf{A}\, a} + a^\top \mathbf{A}\, a\big)^2}{t^2\Big(\operatorname{tr}\mathbf{A\Sigma} + \big(t - \sqrt{a^\top \mathbf{A}\, a}\,\big)^2\Big)},$$

thus inequality (2.5) is sharper than (2.2) if $\operatorname{tr}\mathbf{A\Sigma} \le t\sqrt{a^\top \mathbf{A}\, a} - a^\top \mathbf{A}\, a$.

Theorem 2.3. *Let* $d \in \mathbb{R}^n$ *be fixed in such a way that* $d^\top \mathbf{A}\, d < t^2$, *and define*

$$\begin{aligned} u &= d^\top \mathbf{A\Sigma A}\, d + \big(d^\top \mathbf{A}\, d + a^\top \mathbf{A}\, d\big)^2, & v &= t^2 - d^\top \mathbf{A}\, d, \\ w &= \operatorname{tr}\mathbf{A\Sigma} + (d + a)^\top \mathbf{A}\,(d + a), & z &= u + vw. \end{aligned} \tag{2.6}$$

Then

$$\mathbb{P}(X \notin \mathcal{C}) \le \left(\frac{\sqrt{z} + \sqrt{u}}{v}\right)^2 = \frac{w}{v} \cdot \frac{\sqrt{z} + \sqrt{u}}{\sqrt{z} - \sqrt{u}}\,. \tag{2.7}$$

Particularly, for $d = 0$ estimation (2.7) gives (2.2). If $d = -a$, the quantities in (2.6) have the following simpler form

$$u = a^\top \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}\, a, \quad v = t^2 - a^\top \mathbf{A}\, a, \quad w = \operatorname{tr} \mathbf{A}\boldsymbol{\Sigma}, \quad z = u + vw. \tag{2.8}$$

We remark that inequality (2.7) with

$$d = \left(\frac{\operatorname{tr} \mathbf{A}\boldsymbol{\Sigma}}{t\sqrt{a^\top \mathbf{A}\, a} - a^\top \mathbf{A}\, a} - 1 \right) a$$

is even better than (2.5) if $\operatorname{tr} \mathbf{A}\boldsymbol{\Sigma} \leq t\sqrt{a^\top \mathbf{A}\, a} - a^\top \mathbf{A}\, a$. This will be seen easily from the proofs in Section 3. For more details see the end of Section 3.

Theorem 2.4. *If, in addition, the distribution of X is diagonally symmetric, that is, X and $-X$ have the same distribution, then*

$$\mathbb{P}(X \notin \mathcal{C}) \leq \frac{w}{v} \cdot \frac{\sqrt{z}}{\sqrt{z} - \sqrt{u}} - \frac{u}{v} \cdot \frac{1}{t\sqrt{a^\top \mathbf{A}\, a} + a^\top \mathbf{A}\, a}, \tag{2.9}$$

where u, v, w, z is given by (2.8).

This is obviously sharper than (2.7) with $d = -a$.

3. Proofs

Proof of Theorem 2.1. Let

$$X = \begin{cases} a + t\mathbf{A}^{-1/2} z_j & \text{with probability } \frac{1}{2}(\lambda_j - \alpha_j),\ j = 1, \ldots, n, \\ a - t\mathbf{A}^{-1/2} z_j & \text{with probability } \frac{1}{2}(\lambda_j + \alpha_j),\ j = 1, \ldots, n, \\ a & \text{with probability } 1 - (\lambda_1 + \cdots + \lambda_n). \end{cases}$$

Then

$$\mathbb{E}X = a + \sum_{j=1}^n t\mathbf{A}^{-1/2} z_j \left(\frac{\lambda_j - \alpha_j}{2} - \frac{\lambda_j + \alpha_j}{2} \right) = a - t\mathbf{A}^{-1/2} \sum_{j=1}^n \alpha_j z_j = 0.$$

Moreover,

$$\begin{aligned} \mathbb{V}\mathrm{ar}(X) &= \mathbb{V}\mathrm{ar}(X - a) = \mathbb{E}(X - a)(X - a)^\top - aa^\top \\ &= \sum_{j=1}^n \lambda_j t^2 \mathbf{A}^{-1/2} z_j z_j^\top \mathbf{A}^{-1/2} - aa^\top = \mathbf{A}^{-1/2}\mathbf{B}\mathbf{A}^{-1/2} - aa^\top = \boldsymbol{\Sigma}. \end{aligned}$$

In addition, if $X \neq 0$, then $X^\top \mathbf{A}\, X = t^2$, thus

$$\mathbb{P}(X \notin \mathcal{C}) = \mathbb{P}(X \neq 0) = \lambda_1 + \cdots + \lambda_n = \frac{\operatorname{tr} \mathbf{A\Sigma} + a^\top \mathbf{A}\, a}{t^2}\,.$$

■

Let us turn to the proofs of the upper bounds. In what follows we do not only justify our estimations, but also show why they are relatively good.

First, by introducing $\tilde{a} = \mathbf{A}^{1/2} a$, $\tilde{d} = \mathbf{A}^{1/2} d$, and $\tilde{X} = \mathbf{A}^{1/2} X$ we can reduce the problem to the special case of $\mathbf{A} = \mathbf{I}$. Indeed, $X \in \mathcal{C}$ if and only if $\tilde{X} \in \tilde{\mathcal{C}}$, where $\tilde{\mathcal{C}}$ is the open sphere of center $\tilde{a}$ and radius t. In the conditions we now have $|\tilde{a}|^2 = a^\top \mathbf{A}\, a < t^2$, and similarly, $|\tilde{d}|^2 = d^\top \mathbf{A}\, d < t^2$. In addition, $\tilde{\mathbf{\Sigma}} = \mathbb{V}\mathrm{ar}\, \tilde{X} = \mathbf{A}^{1/2}\mathbf{\Sigma}\mathbf{A}^{1/2}$, thus $\operatorname{tr} \tilde{\mathbf{\Sigma}} = \operatorname{tr} \mathbf{A}^{1/2}\mathbf{\Sigma}\mathbf{A}^{1/2} = \operatorname{tr} \mathbf{A\Sigma}$. Thus, in the sequel we suppose $\mathbf{A} = \mathbf{I}$, and omit the tilde. Definition (2.6) also becomes simpler

$$\begin{aligned} u &= d^\top \mathbf{\Sigma}\, d + \big(|d|^2 + a^\top d\big)^2, & v &= t^2 - |d|^2, \\ w &= \operatorname{tr} \mathbf{\Sigma} + |d + a|^2, & z &= u + vw. \end{aligned} \tag{3.1}$$

Introduce the following notation. For a given symmetric positive definite matrix $\mathbf{B}$ of size $n \times n$, vector $b \in \mathbb{R}^n$, and positive number t let

$$\mathcal{E}(\mathbf{B}, b, t) = \{x \in \mathbb{R}^n : (x - b)^\top \mathbf{B}\,(x - b) < t^2\}.$$

This is an ellipsoid of center b.

This time $\mathcal{C} = \mathcal{E}(\mathbf{I}, a, t) = \{x \in \mathbb{R}^n : |x - a| < t\}$. Our aim is to estimate the probability $\mathbb{P}(X \notin \mathcal{C})$. Let $d = b - a$, that is, $b = d + a$. Suppose that $\mathcal{E}(\mathbf{B}, b, s) \subset \mathcal{C}$, that is, $|d| < t$ and

$$s^2 \leq \min\{(z - d)^\top \mathbf{B}\,(z - d) : |z| = t\}. \tag{3.2}$$

Then, by (2.2),

$$\mathbb{P}(X \notin \mathcal{C}) \leq \mathbb{P}\big(X \notin \mathcal{E}(\mathbf{B}, b, s)\big) \leq \frac{\operatorname{tr} \mathbf{B\Sigma} + b^\top \mathbf{B}\, b}{s^2}. \tag{3.3}$$

With $\mathbf{B}$ and b fixed, the right-hand side of (3.3) attains its minimum when s takes on its largest possible value. Thus, in (3.2) equality must hold. Hence

$$\mathbb{P}(X \notin \mathcal{C}) \leq \min\left\{\frac{\operatorname{tr} \mathbf{B\Sigma} + b^\top \mathbf{B}\, b}{\min\{(z - d)^\top \mathbf{B}\,(z - d) : |z| = t\}} : \mathbf{B} > \mathbf{0},\ |d| < t\right\}. \tag{3.4}$$

It appears quite hard to compute the right-hand side of (3.4). To make calculations easier, instead of arbitrary inscribed ellipsoids we will only consider those of a particular form. Let

$$\mathbf{B} = \mathbf{I} + \lambda d d^\top,$$

then $\mathcal{E}(\mathbf{B}, b, s)$ is a spheroid (cylindrically symmetric ellipsoid) with rotation axis going through the center of $\mathcal{C}$. Making this choice, let us apply the notation $\mathcal{E}(\lambda, d, s)$ instead of $\mathcal{E}(\mathbf{B}, b, s)$, and let $U(t;\lambda, d)$ stand for the fraction on the right-hand side of (3.4). Its numerator is of the form

$$\operatorname{tr}\mathbf{B}\boldsymbol{\Sigma} + b^\top \mathbf{B}\, b = \operatorname{tr}\boldsymbol{\Sigma} + \lambda d^\top \boldsymbol{\Sigma}\, d + |d+a|^2 + \lambda\big(|d|^2 + a^\top d\big)^2,$$

hence we obtain

$$U(t;\lambda, d) = \frac{\operatorname{tr}\boldsymbol{\Sigma} + \lambda d^\top \boldsymbol{\Sigma}\, d + |d+a|^2 + \lambda\big(|d|^2 + a^\top d\big)^2}{\min\{(z-d)^\top(\mathbf{I} + \lambda d d^\top)(z-d) : |z| = t\}}. \tag{3.5}$$

Firstly, let us find the minimum in (3.5). Suppose $\lambda \neq 0$, and let $y = d^\top z$. Then

$$\begin{aligned}(z-d)^\top(\mathbf{I} + \lambda dd^\top)(z-d) &= |z|^2 - 2y + |d|^2 + \lambda y^2 - 2\lambda|d|^2 y + \lambda|d|^4 \\ &= \lambda\Big(y - \frac{1}{\lambda}\big(1+\lambda|d|^2\big)\Big)^2 + t^2 - \frac{1}{\lambda}\,\big(1+\lambda|d|^2\big).\end{aligned} \tag{3.6}$$

Suppose $\lambda > 0$. It is easy to see that the minimum is attained at $y = \dfrac{1}{\lambda}\big(1 + \lambda|d|^2\big)$, if this value is admissible, that is,

$$\frac{1}{\lambda}\big(1+\lambda|d|^2\big) \le |d|t. \tag{3.7}$$

In that case $s^2 = t^2 - \dfrac{1}{\lambda}\big(1+\lambda|d|^2\big)$, thus

$$U(t;\lambda, d) = \frac{\operatorname{tr}\boldsymbol{\Sigma} + \lambda d^\top \boldsymbol{\Sigma}\, d + |d+a|^2 + \lambda\big(|d|^2 + a^\top d\big)^2}{t^2 - \big(1/\lambda + |d|^2\big)}. \tag{3.8}$$

If (3.7) does not hold, the minimum is attained at the admissible point being the closest to the minimum point of the quadratic polynomial, namely, at $y = |d|t$, and the minimum is $(t-|d|)^2\big(1+\lambda|d|^2\big)$. Hence,

$$U(t;\lambda, d) = \frac{\operatorname{tr}\boldsymbol{\Sigma} + \lambda d^\top \boldsymbol{\Sigma}\, d + |d+a|^2 + \lambda\big(|d|^2 + a^\top d\big)^2}{(t-|d|)^2\big(1+\lambda|d|^2\big)}. \tag{3.9}$$

If $\lambda < 0$, then the positive definiteness of $\mathbf{B}$ requires $d^\top \mathbf{B}\, d = |d|^2 + \lambda|d|^4 = |d|^2\big(1+\lambda|d|^2\big) > 0$. In that case (3.6) implies that the minimum is at the upper end of the domain of admissibility, $y = |d|t$. Thus, formula (3.9) remains true.

Finally, if $\lambda = 0$, the denominator of $U(t;\lambda, d)$ becomes simpler: it equals $\min\{|z-d|^2 : |z| = t\}$. Clearly, $|z-d|^2 \ge (|z| - |d|)^2 = (t-|d|)^2$, and this lower bound can be attained. Hence the minimum is $(t-|d|)^2$, thus (3.9) is still valid:

$$U(t;0,d) = \frac{\operatorname{tr}\boldsymbol{\Sigma} + |d+a|^2}{(t-|d|)^2}. \tag{3.10}$$

In what follows we first determine $\min\{U(t;0,d) : |d| < t\}$, that will imply Theorem 2.2, then also $\min\{U(t;\lambda,d) : 1+\lambda|d|^2 > 0\}$ with d fixed, justifying Theorem 2.3. Unfortunately, we cannot derive an explicit form of $\min\limits_{\lambda,\, d} U(t;\lambda,d)$.

Lemma 3.1.

$$\min\{U(t;0,d) : |d| < t\} = \begin{cases} \dfrac{\operatorname{tr}\boldsymbol{\Sigma}}{\operatorname{tr}\boldsymbol{\Sigma} + (t-|a|)^2}, & \textit{if } \operatorname{tr}\boldsymbol{\Sigma} \le |a|(t-|a|), \\ \dfrac{\operatorname{tr}\boldsymbol{\Sigma} + |a|^2}{t^2}, & \textit{if } \operatorname{tr}\boldsymbol{\Sigma} > |a|(t-|a|). \end{cases}$$

Proof. Having $|d|$ fixed we find $|d+a|^2$ minimal for $d = -\dfrac{|d|}{|a|}\,a$, thus $|d+a|^2 = (|d|-|a|)^2$. Let $y = |d|$, then (3.10) implies that

$$U(t;0,d) = \frac{\operatorname{tr}\boldsymbol{\Sigma} + (y-|a|)^2}{(t-y)^2}.$$

From that we obtain

$$\begin{aligned} \frac{d}{dy}\log U(t;0,d) &= \frac{2(y-|a|)}{\operatorname{tr}\boldsymbol{\Sigma} + (y-|a|)^2} + \frac{2}{t-y} \\ &= \frac{2\big(\operatorname{tr}\boldsymbol{\Sigma} + (t-|a|)(y-|a|)\big)}{\big(\operatorname{tr}\boldsymbol{\Sigma} + (y-|a|)^2\big)(t-y)}. \end{aligned}$$

Consequently, if $\operatorname{tr}\boldsymbol{\Sigma} \le |a|(t-|a|)$, then the minimum is attained at

$$y = |a| - \frac{\operatorname{tr}\boldsymbol{\Sigma}}{t-|a|}, \tag{3.11}$$

and its value is

$$\frac{\operatorname{tr}\boldsymbol{\Sigma}}{\operatorname{tr}\boldsymbol{\Sigma} + (t-|a|)^2}. \tag{3.12}$$

If, on the other hand, $\operatorname{tr}\boldsymbol{\Sigma} > |a|(t-|a|)$, then $\frac{d}{dy}\log U(t;0,d) > 0$, thus $U(t;0,d)$ is an increasing function of y. Hence the minimum is attained at $y=0$, and it equals the bound of (2.2), namely, $\dfrac{\operatorname{tr}\boldsymbol{\Sigma} + |a|^2}{t^2}$. ∎

Lemma 3.2. *With the notations of* (3.1) *we have*

$$\min\{U(t;\lambda,d) : 1+\lambda|d|^2>0\,\} = \left(\frac{\sqrt{z}+\sqrt{u}}{v}\right)^2 = \frac{w}{v}\cdot\frac{\sqrt{z}+\sqrt{u}}{\sqrt{z}-\sqrt{u}}\,. \tag{3.13}$$

Proof. Firstly, let us focus on values of λ for which (3.7) holds with the opposite inequality, that is,

$$\frac{1}{\lambda}\big(1+\lambda|d|^2\big)\geq |d|t.$$

Then, by (3.9),

$$\begin{aligned}
U(t;\lambda,d) &= \frac{\operatorname{tr}\boldsymbol{\Sigma}+\lambda d^\top\boldsymbol{\Sigma}\,d+|d+a|^2+\lambda\big(|d|^2+a^\top d\big)^2}{(t-|d|)^2\big(1+\lambda|d|^2\big)}\\
&= \frac{\operatorname{tr}\boldsymbol{\Sigma}+\lambda d^\top\boldsymbol{\Sigma}\,d+|d+a|^2+\dfrac{1+\lambda|d|^2}{|d|^2}\big(|d|^2+a^\top d\big)^2-\dfrac{\big(|d|^2+a^\top d\big)^2}{|d|^2}}{(t-|d|)^2\big(1+\lambda|d|^2\big)}\\
&= \frac{\operatorname{tr}\boldsymbol{\Sigma}+\lambda d^\top\boldsymbol{\Sigma}\,d+|d+a|^2-|d|^2-2a^\top d-\dfrac{(a^\top d)^2}{|d|^2}}{(t-|d|)^2\big(1+\lambda|d|^2\big)}+\frac{\big(|d|^2+a^\top d\big)^2}{|d|^2(t-|d|)^2}\\
&= \frac{1}{(t-|d|)^2}\cdot\frac{\operatorname{tr}\boldsymbol{\Sigma}+\lambda d^\top\boldsymbol{\Sigma}\,d+|a|^2-(a^\top d)^2|d|^{-2}}{1+\lambda|d|^2}+\frac{\big(|d|^2+a^\top d\big)^2}{|d|^2(t-|d|)^2}.
\end{aligned}$$

It suffices to deal only with the term containing λ.

$$\begin{aligned}
&\frac{d}{d\lambda}\log\frac{\operatorname{tr}\boldsymbol{\Sigma}+\lambda d^\top\boldsymbol{\Sigma}\,d+|a|^2-(a^\top d)^2|d|^{-2}}{1+\lambda|d|^2}\\
&\qquad= \frac{d^\top\boldsymbol{\Sigma}\,d}{\operatorname{tr}\boldsymbol{\Sigma}+\lambda d^\top\boldsymbol{\Sigma}\,d+|a|^2-(a^\top d)^2|d|^{-2}}-\frac{|d|^2}{1+\lambda|d|^2}\\
&\qquad\qquad= \frac{d^\top\boldsymbol{\Sigma}\,d-|d|^2\operatorname{tr}\boldsymbol{\Sigma}-\big(|a|^2|d|^2-(a^\top d)^2\big)}{\big(\operatorname{tr}\boldsymbol{\Sigma}+\lambda d^\top\boldsymbol{\Sigma}\,d+|a|^2-(a^\top d)^2|d|^{-2}\big)\big(1+\lambda|d|^2\big)}\leq 0,
\end{aligned}$$

hence $U(t;\lambda,d)$ is minimal if and only if equality holds in (3.7).

Let us turn to the case where (3.7) is valid. Using the notations of (3.1) we obtain (3.8) in the following form

$$U(t;\lambda,d)=\frac{w+u\lambda}{v-1/\lambda}.$$

Note that condition (3.7) requires

$$0<\frac{1}{\lambda}\leq |d|(t-|d|)=\frac{|d|}{t+|d|}\,v. \tag{3.14}$$

Let us differentiate it with respect to λ.

$$\frac{d}{d\lambda}\log U(t;\lambda,d)=\frac{u}{w+u\lambda}-\frac{1}{\lambda(v\,\lambda-1)}=\frac{uv\lambda^2-2u\lambda-w}{(w+u\lambda)\lambda(v\lambda-1)}. \tag{3.15}$$

The positive root of the numerator is

$$\lambda=\frac{u+\sqrt{u^2+uvw}}{uv}=\frac{\sqrt{z}+\sqrt{u}}{v\sqrt{u}}. \tag{3.16}$$

If this λ satisfies (3.14), then it provides the minimum of $U(t;\lambda,d)$. It is easy to see that (3.14) equivalent to $|d|^2z-t^2u\ge 0$. The latter obviously holds now, because $|d|^2z-t^2u=v\big(|d|^2\operatorname{tr}\mathbf{\Sigma}-d^\top\mathbf{\Sigma}\,d+|a|^2|d|^2-(a^\top d)^2\big)\ge 0$.

From (3.15) we have $w=uv\lambda^2-2u\lambda$, hence the minimum is equal to

$$\frac{\lambda(w+u\lambda)}{v\lambda-1}=u\lambda^2=\left(\frac{\sqrt{z}+\sqrt{u}}{v}\right)^2,$$

as stated. ■

Proof of Theorem 2.4. Consider an arbitrary ellipsoid with the origin as its center, and bisect it with a hyperplane going through the origin. Then, by the diagonal symmetry, both halves have the same probability of containing X. Basing on this observation we cut $\mathcal{C}$ by the hyperplane orthogonal to the radius through the origin, and inscribe a half of a spheroid into both parts in such a way that each of those halves maximizes the right-hand side of (2.2).

The smaller part of $\mathcal{C}$ is the one that does not contain the center. By inversion with respect to the origin it is transformed into a subset of the bigger part. Thus the image of the inscribed half ellipsoid lies entirely in $\mathcal{C}$. Among all inscribed ellipsoids of the form $\mathcal{E}(\lambda,-a,s)$ the right-hand side of (2.2) is maximized by (3.16), giving the following estimation.

$$\mathbb{P}(X\notin\mathcal{C},\ a^\top X<0)\le\frac{w}{2v}\cdot\frac{\sqrt{z}+\sqrt{u}}{\sqrt{z}-\sqrt{u}}. \tag{3.17}$$

Let us turn to the bigger part. We want to minimize $\operatorname{tr}\mathbf{B}\mathbf{\Sigma}/s^2$ among all ellipsoids $\mathcal{E}(\lambda,-a,s)$ for which

$$\mathcal{E}(\lambda,-a,s)\cap\{z\in\mathbb{R}^n: a^\top z\ge 0\}\subset\mathcal{C}. \tag{3.18}$$

Note that for every point z of such an ellipsoid either z or $-z$ falls in $\mathcal{C}$, hence

$$|z|\le t+|a|. \tag{3.19}$$

We will show that the minimum is attained for

$$\lambda_1 = -\frac{2}{|a|(t+|a|)}, \quad s_1^2 = t^2 - |a|^2. \tag{3.20}$$

First we check that $\mathbf{B} = \mathbf{I} + \lambda_1 aa^\top$ is positive definite. Let $z \in \mathbb{R}^n$ be different from 0, then

$$z^\top \mathbf{B}\, z = |z|^2 - \frac{2(a^\top z)^2}{|a|(t+|a|)} \geq |z|^2 - \frac{2|a|^2|z|^2}{|a|(t+|a|)} = |z|^2\, \frac{t-|a|}{t+|a|} > 0. \tag{3.21}$$

Next we show that (3.18) holds for λ_1 and s_1. Indeed, let $z^\top \mathbf{B}\, z < t^2 - |a|^2$ and $a^\top z \geq 0$. Then from (3.21) we get that $|z| < t + |a|$, thus $a^\top z \leq |a|(t+|a|)$, whence

$$\begin{aligned} |z-a|^2 = |z|^2 + |a|^2 - 2a^\top z &= z^\top \mathbf{B}\, z + |a|^2 - 2a^\top z - \lambda_1 (a^\top z)^2 \\ &< t^2 - 2a^\top z \left(1 - \frac{a^\top z}{|a|(t+|a|)}\right) \leq t^2. \end{aligned}$$

Finally, suppose that (3.18) is satisfied for some λ_2 and s_2. We are going to prove that $\mathcal{E}(\lambda_2, -a, s_2) \subset \mathcal{E}(\lambda_1, -a, s_1)$, that is,

$$\mathcal{E}(\mathbf{I} + \lambda_2 aa^\top, 0, s_2) \subset \mathcal{E}(\mathbf{I} + \lambda_1 aa^\top, 0, s_1).$$

This will be sufficient, because it is not hard to see in general that $\mathcal{E}(\mathbf{B}_2, 0, s_2) \subset \mathcal{E}(\mathbf{B}_1, 0, s_1)$ implies $\operatorname{tr} \mathbf{B}_2 \mathbf{\Sigma}/s_2^2 \geq \operatorname{tr} \mathbf{B}_1 \mathbf{\Sigma}/s_1^2$. Indeed, in that case

$$\mathbb{P}\left(\frac{X^\top \mathbf{B}_1\, X}{s_1^2} \geq y\right) \leq \mathbb{P}\left(\frac{X^\top \mathbf{B}_2\, X}{s_2^2} \geq y\right)$$

holds for every positive y. Integrating on both sides with respect to y from zero to infinity we obtain

$$\frac{\operatorname{tr} \mathbf{B}_1 \mathbf{\Sigma}}{s_1^2} = \mathbb{E}\left(\frac{X^\top \mathbf{B}_1\, X}{s_1^2}\right) \leq \mathbb{E}\left(\frac{X^\top \mathbf{B}_2\, X}{s_2^2}\right) = \frac{\operatorname{tr} \mathbf{B}_2 \mathbf{\Sigma}}{s_2^2}.$$

Firstly, suppose $\lambda_2 \leq \lambda_1$. Let $z_1 = (t+|a|)a/|a|$. It is a boundary point of $\mathcal{C}$, thus $s_2^2 \leq z_1^\top (\mathbf{I} + \lambda_2 aa^\top) z_1 = (t+|a|)^2(1 + \lambda_2 |a|^2)$. Therefore, if $z \in \mathcal{E}(\lambda_2, -a, s_2)$, then by (3.19) we have

$$\begin{aligned} z^\top(\mathbf{I} + \lambda_1 aa^\top)z = z^\top(\mathbf{I} + \lambda_2 aa^\top)z + (\lambda_1 - \lambda_2)(a^\top z)^2 &< s_2^2 + (\lambda_1 - \lambda_2)|a|^2|z|^2 \\ &\leq (t+|a|)^2(1+\lambda_2|a|^2) + (\lambda_1 - \lambda_2)|a|(t+|a|) \\ &= (t+|a|)^2(1+\lambda_1|a|^2) = t^2 - |a|^2. \end{aligned}$$

Secondly, suppose that $\lambda_2 > \lambda_1$. Pick a boundary point z_1 of $\mathcal{C}$ such that $a^\top z_1 = 0$. Then $s_2^2 \le z_1^\top(\mathbf{I} + \lambda_2 aa^\top)z_1 = |z_1|^2 = |z_1 - a|^2 - |a|^2 = t^2 - |a|^2$. Hence, if $z \in \mathcal{E}(\lambda_2, -a, s_2)$, then

$$z^\top(\mathbf{I} + \lambda_1 aa^\top)z \le z^\top(\mathbf{I} + \lambda_2 aa^\top)z < s_2^2 \le t^2 - |a|^2.$$

From all these we conclude that

$$\mathbb{P}(X \notin \mathcal{C},\ a^\top X \ge 0) \le \frac{1}{2} \cdot \frac{\operatorname{tr} \mathbf{\Sigma} + \lambda_1 a^\top \mathbf{\Sigma}\, a}{t^2 - |a|^2} = \frac{w + \lambda_1 u}{2v} = \frac{w}{2v} - \frac{u}{v} \cdot \frac{1}{|a|(t + |a|)}. \tag{3.22}$$

Now the proof can be completed by combining (3.17) with (3.22). ∎

Finally, from Lemma 3.1 and its proof we know that the upper bound of Theorem 2.2 is obtained as $\min_{|d|<t} U(t; 0, d)$. Moreover, if $\operatorname{tr} \mathbf{\Sigma} \le |a|(t - |a|)$, this minimum is attained at

$$d = \left(\frac{\operatorname{tr} \mathbf{\Sigma}}{|a|(t - |a|)} - 1 \right) a.$$

On the other hand, the upper bound of Theorem 2.3 is equal to $\min_\lambda U(t; \lambda, d)$, by Lemma 3.2. Hence inequality (2.7) with d as above is better than (2.5) if $\operatorname{tr} \mathbf{\Sigma} \le |a|(t - |a|)$. This proves our remark preceding Theorem 2.4.

4. Comparison with earlier results

In [9], Olkin and Pratt constructed an upper bound for the probability

$$\mathbb{P}\left(|Y_i| \ge k_i \ \text{ for some } i,\ 1 \le i \le n\right)$$

by considering an ellipsoid $\{x \in \mathbb{R}^n : x^\top \mathbf{A} x \le 1\}$ contained by the hypercube $[-1, +1]^n$, then minimizing $\operatorname{tr} \mathbf{A\Sigma}$, where $X_i = Y_i/k_i$ and $\mathbf{\Sigma} = \mathbb{V}\mathrm{ar}(X)$.

They showed that the minimum is attained at $\mathbf{A} = \mathbf{B}^{-1}$, where $\mathbf{B}$ is the unique positive definite matrix with diagonal elements 1, such that $\mathbf{A\Sigma A} = \mathbf{\Lambda}$ is diagonal. They stated that $\mathbf{B}$ can not be obtained from $\mathbf{\Sigma}$ by standard matrix operations except in special cases. Later Meaux et al. [8] suggested a method of solving such a matrix equation numerically by giving lower and upper bounds for the elements of $\mathbf{B}$ and $\mathbf{\Lambda}$, and then using a generalized bisection method.

The approach of Birnbaum and Marshall [2] was similar, but they considered the case when the expectation of the random vector did not coincide with the center of the ellipsoid.

These methods resulted in the upper bound that arises when the Markov inequality is applied to $(X-a)^\top \mathbf{A}(X-a)$; that is, in (2.2).

Jensen [4] generalized that to multiple ellipsoids.

Thus, in comparable cases, our upper bounds are strictly better than theirs, provided $\mathrm{tr}\mathbf{A\Sigma} \le t\sqrt{a^\top \mathbf{A} a} - a^\top \mathbf{A} a$.

In the following numerical examples upper bounds are labelled with the theorems they come from, and the result of the former approach is referred to as ‘earlier’. In Theorem 2.3 we choose $d = \big(\mathrm{tr}\mathbf{A\Sigma}/(t\sqrt{a^\top \mathbf{A}a}+a^\top \mathbf{A}a)-1\big)a$, and in Theorem 2.4, $d=-a$. The exact probabilities are computed by simulations implemented in R, with sample size 10^7 at least.

Example 4.1. Set

$$\mathbf{A}_0 = \begin{bmatrix} 0.6 & -0.3 & 0 \\ -0.3 & 0.7 & -0.1 \\ 0 & -0.1 & 0.8 \end{bmatrix}, \quad a = \begin{bmatrix} 1 \\ 1/2 \\ 2 \end{bmatrix}, \quad \mathbf{\Sigma} = \begin{bmatrix} 1 & 1/4 & 1/4 \\ 1/4 & 2 & 1/2 \\ 1/4 & 1/2 & 3 \end{bmatrix}.$$

Let the random vectors X and Y be Gaussian $N(0,\mathbf{\Sigma})$, and $t_\nu(0, \frac{\nu-2}{\nu}\mathbf{\Sigma})$, resp. (For the definition of the multivariate t-distribution see [5]). Thus the variances of the random vectors X and Y are the same. As before, we consider the ellipsoid $\mathcal{C} = \{x \in \mathbb{R}^3 : (x-a)^\top \mathbf{A}(x-a) < t^2\}$.

In Table 1 we compare earlier bounds with the ones provided by Theorems 2.2–2.4 under three different choices of parameters t, $\mathbf{A}$, and ν. Although our

Table 1. Comparison of bounds in the case of multivariate Gaussian and multivariate t-distribution.

t	6	4.5	1.5
$\mathbf{A}$	$\mathbf{A}_0$	$\mathbf{A}_0$	$0.1\,\mathbf{A}_0$
ν	3	3	4
$\mathbb{P}(X \notin \mathcal{C})$	0.0045	0.0565	0.0398
$\mathbb{P}(Y \notin \mathcal{C})$	0.0174	0.0485	0.0452
Earlier	0.2118	0.3765	0.3389
Theorem 2.2	0.1952	0.3739	0.3336
Theorem 2.3	0.1815	0.3701	0.3241
Theorem 2.4	0.1499	0.3220	0.2773*

new estimations are better than the widely known bound of (2.2), they are still far from the exact probabilities. This is no surprise, since a general estimation, valid without strict conditions on the distribution, is necessarily rather crude, especially for distributions with very light tails.

Example 4.2. This is a variant of Example 4.1 in 5 dimensions. Again, X is zero mean Gaussian, Y is multivariate t, and they have the same covariance matrix $\mathbf{\Sigma}$. This time let

$$\mathbf{A}_0 = \begin{bmatrix} 0.80 & 0.12 & -0.08 & 0.08 & -0.1 \\ 0.12 & 0.73 & -0.1 & 0 & 0.13 \\ -0.08 & -0.1 & 0.56 & 0.06 & 0.01 \\ 0.08 & 0 & 0.06 & 0.31 & -0.05 \\ -0.1 & 0.13 & 0.01 & -0.05 & 0.86 \end{bmatrix}, \quad a = \begin{bmatrix} -0.8 \\ 0.5 \\ 0 \\ 0.5 \\ 0.5 \end{bmatrix},$$

$$\mathbf{\Sigma} = \begin{bmatrix} 0.71 & 0.01 & 0.02 & -0.01 & -0.01 \\ 0.01 & 0.72 & 0 & -0.06 & -0.03 \\ 0.02 & 0 & 0.66 & 0.14 & 0.08 \\ -0.01 & -0.06 & 0.14 & 0.34 & -0.23 \\ -0.01 & -0.03 & 0.08 & -0.23 & 0.61 \end{bmatrix}.$$

Table 2 contains the exact probabilities and the earlier bounds together with those provided by our Theorems 2.2–2.4. The overall picture is quite similar to what we have seen in Table 1, but this time Theorem 2.2 does not significantly improve the "Earlier" estimate $(\operatorname{tr}\mathbf{A\Sigma} + a^\top \mathbf{A} a)/t^2$.

Table 2. Comparison of bounds in the case of multivariate Gaussian and multivariate t-distribution.

t	4	3.5	3.2	5
$\mathbf{A}$	$\mathbf{A}_0$	$\mathbf{A}_0$	$\mathbf{A}_0$	$2.5\,\mathbf{A}_0$
ν	3	3	3	3
$\mathbb{P}(X \notin \mathcal{C})$	0.0003	0.0025	0.0082	0.0095
$\mathbb{P}(Y \notin \mathcal{C})$	0.0164	0.0250	0.0332	0.0343
Earlier	0.1921	0.2509	0.30014	0.307340
Theorem 2.2	0.1884	0.2498	0.30011	0.307339
Theorem 2.3	0.1712	0.2304	0.2813	0.2888
Theorem 2.4	0.1543	0.2092	0.2580	0.2654

Example 4.3. When $\mathcal{C}$ is not centered at the origin, the upper bound can be far from the exact probability, but it is sharper when applied to fat tailed distributions. In this example, we also consider a fat tailed distribution with generalized Pareto marginals.

The generalized Pareto distribution $GPD(\mu, \sigma, \xi)$ [10] with location parameter $\mu \in \mathbb{R}$, scale parameter $\sigma > 0$ and shape parameter $\xi \in \mathbb{R}$ is defined by its distribution function

$$F(x) = \begin{cases} 1 - \left(1 + \left(\frac{\xi(x-\mu)}{\sigma}\right)\right)^{-1/\xi} & \text{for } \xi \neq 0, \\ 1 - \exp\left(-\frac{x-\mu}{\sigma}\right) & \text{for } \xi = 0. \end{cases}$$

where x is to satisfy $1 + \xi(x-\mu)/\sigma \geq 0$ and $x \geq \mu$.

Let U_1 and U_2 be independent random variables of $GPD(3, 1, -10)$ and $GPD(1, 0.2, 0.05)$ distribution, resp. Define $V^\top = [U_1, \frac{1}{2}U_1 + \frac{1}{2}U_2]$, and $W = V - \mathbb{E}V$.

Then the covariance matrix of the random vector W is

$$\mathbf{\Sigma} = \begin{bmatrix} 0.0003935458 & 0.0001967729 \\ 0.0001967729 & 0.0124098669 \end{bmatrix}.$$

In addition, let X be zero mean bivariate Gaussian and Y bivariate t as above, Y with $\nu = 3$, such that they have covariance matrix $\mathbf{\Sigma}$.

Consider the ellipsoid $\mathcal{C}$ given by

$$\mathbf{A} = \begin{bmatrix} 0.6 & -0.3 \\ -0.3 & 0.7 \end{bmatrix}, \quad a = \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \quad t = 1.$$

Table 3. Comparison of bounds in the case of bivariate Gaussian, t and Pareto distributions.

$\mathbb{P}(X \notin \mathcal{C})$	$\mathbb{P}(Y \notin \mathcal{C})$	$\mathbb{P}(W \notin \mathcal{C})$	Earlier	Theorem 2.2	Theorem 2.3
0.0040	0.0099	0.0255	0.7088	0.2481	0.1382

Under this parametrization, Theorem 2.4 would give the upper bound 0.0825, but it can not be applied to the Pareto distribution. It can be seen that for W the relative errors of the upper bounds given by Theorems 2.2–2.3 are much less than that of the "Earlier" estimate.

5. Summary

We have presented Chebyshev type upper bounds for the probability that a random vector falls outside an ellipsoid. We have supposed nothing about the distribution apart from the mean and variance, and, in only one of the theorems, diagonal symmetry. Therefore our results may be applied in nonparametric statistics in models where no particular properties, such as normality, unimodality or light tails can be assumed. Since mean and variance can be efficiently estimated from the sample, our inequalities can be applied to constructing (asymptotically) conservative critical regions in certain hypothesis testing problems.

We do not require the ellipsoid to be centered at the mean of the random vector. In that excentric case our bounds are better than those presented earlier. We have also compared our bounds to earlier ones through numerical examples. We found that significant improvement on earlier bounds can be achieved especially when the the random vector has a heavy tailed distribution.

References

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

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

[3] **Csiszár, V. and T.F. Móri**, A Bienaymé–Chebyshev inequality for scale mixtures of the multivariate normal distribution, *Math. Inequal. Appl.*, **12** (2009), 839–844.

[4] **Jensen, D.R.**, Inequalities for joint distributions of quadratic forms, *SIAM J. Appl. Math.,* **42** (1982), 297–301.

[5] **Kotz, S. and S. Nadarajah**, *Multivariate t Distributions and Their Applications*, Cambridge University Press, Cambridge, 2004.

[6] **Marshall, A. and I. Olkin**, Multivariate Chebyshev inequalities, *Ann. Math. Statist.,* **31** (1960), 1001–1014.

[7] **Mallows, C.L. and D. Richter**, Inequalities of Chebyshev type involving conditional expectations, *Ann. Math. Statist.,* **40** (1969), 1922–1932.

[8] **Meaux, L.M., J.W. Seaman and T.L. Boullion**, Calculation of multivariate Chebyshev-type inequalities, *Comp. Math. Appl.,* **20** (1990), 55–60.

[9] **Olkin, I. and J.W. Pratt**, A multivariate Tchebycheff inequality, *Ann. Math. Statist.,* **29** (1958), 226–234.

[10] **Pickands, J.**, Statistical inference using extreme order statistics, *Annals of Statistics*, **3** (1975), 119–131.

[11] **Skibinsky, M.**, The minimum probability of an interval when the mean and the variance are known, *Ann. Inst. Stat. Math.,* **32** (1980), 377–385.

[12] **Vandenberghe, L., S. Boyd and K. Comanor**, Generalized Chebyshev bounds via semidefinite programming, *SIAM Rev.*, **49** (2007), 52–64 (electronic).

[13] **Vysochanskiĭ, D.F. and Y.I. Petunin**, Justification of the three sigma rule for unimodal distributions, *Theory Probab. Math. Statist.,* **21** (1980), 25–36.

Villő Csiszár
Tamás Fegyverneki
Tamás F. Móri
Department of Probability Theory and Statistics, Faculty of Science
Eötvös Loránd University
Budapest, Hungary
villo@ludens.elte.hu
tfegyi@math.elte.hu
mori@math.elte.hu