Probability in the Engineering and Informational Sciences, **21**, 2007, 289–300. Printed in the U.S.A.
DOI: 10.1017/S0269964807070179

A CHEBYSHEV INEQUALITY FOR MULTIVARIATE NORMAL DISTRIBUTION

DAVAADORJIN MONHOR
Faculty of Geoinformatics
University of West Hungary
H-8002 Szekesfehervar, Hungary
E-mail: monhor@ella.hu

Chebyshev inequality estimates the probability for exceeding the deviation of a random variable from its mathematical expectation in terms of the variance of the random variable. In modern probability theory, the Chebyshev inequality is the most frequently used tool for proving different convergence processes; for example, it plays a fundamental role in proofs of various forms of laws of large numbers. The mathematical expression of the bound on the probability in the Chebyshev inequality is very simple and can be modified easily for different kinds of sequence of random variables (e.g., for the case of sums of independent random variables). This fact lies behind these frequent applications. In this setting, the Chebyshev inequality has pure theoretical "applications" in probability theory and its role is to provide "a guarantee" of convergence but not to give a bound on concrete probability content.

In the present article we consider the Chebyshev inequality as a probability bound that is essential for the translation from its conventional theoretical applications to the practical setting if easy-to-compute multivariate generalizations are derived.

Such an inequality for the random vectors having multivariate Normal distribution is proved. The new inequality gives a lower bound in terms of variances on the probability that the random vector in question falls into an Euclidean ball with center at mean vector. The need and importance of consideration of this kind of multivariate Chebyshev inequality stemmed from several problems in engineering and informational sciences (Hassibi and Boyd [9], Jeng [10], Jeng and Woods [11], Molina, Katseggelos, Mateos, Hermoso, and Segall [13]). Jeng [10] derived an inequality that gives an upper bound for the probability in question. The simultaneous application of the established multivariate Chebyshev inequality and Jeng's inequality is useful in practical problems by providing lower and upper bounds on the probability content.

© 2007 Cambridge University Press 0269-9648/07 $25.00

https://doi.org/10.1017/S0269964807070179 Published online by Cambridge University Press

The inequality is attractive by its being easy to compute and its similarity to the original Chebyshev inequality, in contrast to well-known complicated multivariate Chebyshev inequalities. The present article also gives some insights into the very origin of the Chebyshev inequality, which makes the article self-contained.

1. INTRODUCTION: BACKGROUND AND MOTIVATION

The consideration of various sequences of random variables and proving or disproving their convergence are one of the main streamlines of modern theoretical probability. In the establishment of the convergence of these sequences, the one-dimensional Chebyshev inequality is an invaluable tool and most frequently used technique. The Chebyshev inequality estimates the probability for exceeding the deviation of a random variable from its mathematical expectation in terms of the variance of the random variable. The mathematical expression of the bound on the probability in the Chebyshev inequality is very simple and can be modified easily for different kinds of sequence of random variables (e.g., for the case of sums of independent random variables). This fact lies behind these frequent applications. In this setting, the Chebyshev inequality has pure theoretical "applications" in probability theory and its role is to provide "a guarantee" of convergence but not to give a bound on concrete probability content.

As is well known, the computing of probabilities in multidimensional spaces is a very hard problem in mathematics, being expressed by multiple integrals. One of the characteristic features of these problems that contributes to its difficulties is the dimensional effect or the "curse" of dimensionality (i.e., any increase in dimension implies substantial computational difficulties). In the light of this phenomenon, any easy–to–compute bound on multivariate probabilities conveys valuable information. In general, bounds on multivariate probability distributions are useful in various areas (Monhor [14]).

Berge [2] initiated the multivariate generalization of the Chebyshev inequality. Considering correlated bivariate random variables, he derived an inequality for the rectangular area with sides whose lengths are the variances multiplied by a positive common factor. After this pioneering work, Lal [12] generalized Berge's result to the case of different factors. Finally, Olkin and Pratt [16] succeeded in deriving Lal's result to arbitrary n-dimensional space. In a theoretical setting, this is a significant result in the studies of multivariate generalization of the Chebyshev inequality. However, the inequality of Olkin and Pratt is too complicated to be applied in practical problems. The above-described background and motivations clarify the mathematical aspects of the topic in question.

The engineering motivation is as follows. The need and importance of bounding the probability content that the multivariate normal random vector falls into an n-dimensional Euclidean ball stemmed from several problems in engineering and informational sciences: GPS ambiguity problems in Hassibi and Boyd [9] and image

https://doi.org/10.1017/S0269964807070179 Published online by Cambridge University Press

estimation problems in the context of Gauss–Markov random fields in Jeng [10], Jeng and Woods [11], Molina, Katseggelos, Mateos, Hermoso, and Segall [13].

Consequently, in the present article we address the above problem in terms of the multivariate Chebyshev inequality (i.e., the question of how the one-dimensional Chebyshev inequality can be generalized for estimating the probability content of an n-dimensional Euclidean ball in the case of the multivariate Normal distribution). As a solution to this problem, we derived a new multivariate Chebyshev inequality that gives a lower bound on the probability in question. The bound is given in terms of variances of the random vector in the same way as in the classical one-dimensional Chebyshev inequality. This fact makes the inequality attractive by its being easy to compute and its similarity to the original Chebyshev inequality, in contrast to well-known complicated multivariate Chebyshev inequalities.

The present article also gives some insights into the very origin of the Chebyshev inequality in both one-dimensional and multidimensional settings. This gives the article an expository value and, at the same time, makes it self-contained.

The article is structured as follows. In Sections 2 and 3 we consider brief outlines of the developments in one-dimensional and multidimensional Chebyshev inequalities in an historical setting. Section 4 presents an auxiliary lemma that plays a key role in proving the main result of the article. In Section 5 we prove a new multivariate Chebyshev inequality for multivariate Normal distribution. Section 6 summarizes the results of the article and draws conclusions.

2. ORIGIN OF ONE-DIMENSIONAL CHEBYSHEV INEQUALITY

In any advanced textbook of modern probability theory (e.g., Ross [17]) or a monograph that deals with convergence of sequences of random variables, one inevitably finds the following inequality under the name Chebyshev inequality:

$$\mathbf{P}(|\xi - \mathbf{E}(\xi)| > \varepsilon) \leq \frac{\mathrm{Var}(\xi)}{\varepsilon^2}, \tag{2.1}$$

where ε is an arbitrary positive number.

This form of the inequality is suitable for proving the convergence in probability of sequences of random variables in terms of variance. The ε should be chosen as small as possible, in order to prove a convergence of a sequence of random variables. In this situation, inequality (2.1) is not used for a single random variable ξ, but for a sequence of random variables (say, for ξ_n, $n = 1, 2, \ldots$). This means that the convergence to zero of the sequence of numerators in the right-hand side of inequality (2.1) should be managed. This goal can be achieved in numerous ways, not only for simple sequences of random variables but also for sequences of various functions of random variables. This is an important research area in probability theory and, here, Chebyshev inequality is an extremely flexible tool. However, this is a theoretical issue and going

https://doi.org/10.1017/S0269964807070179 Published online by Cambridge University Press

into details further would be a digression from the main line of the article but what we mentioned above is helpful in grasping the very essence of the Chebyshev inequality.

Below, we reformulate inequality (2.1) in the following slightly different form in a precise mathematical setting.

Chebyshev Inequality: *If a random variable ξ has finite variance, then, for any $r > 0$, we have*

$$\mathbf{P}(|\xi - \mathbf{E}(\xi)| \le r) \ge 1 - \frac{\mathrm{Var}(\xi)}{r^2}. \tag{2.2}$$

Chebyshev (1821–1894) [5,6] proved his inequality for the discrete random variables taking a finite number of possible values and the inequality was published in 1867. The proof was based on very long and tedious elementary algebraic calculations. In 1821 (the year of the birth of Chebyshev), the great German mathematician Karl Friedrich Gauss [7,8] proved the following inequality and presented it to the Royal Scientific Society of Göttingen; it appeared in 1823.

Gauss Inequality (Gauss (1821)): *If the probability distribution function of a random variable ξ is unimodal, then, for $k > 0$, we have*

$$\mathbf{P}(|\xi - x_0| > k\tau) \le \frac{4}{9k^2}, \tag{2.3}$$

where x_0 is the mode and $\tau^2 = \mathbf{E}[(\xi - x_0)^2]$.

From this inequality we see that Gauss considered the mode instead of the mathematical expectation for a continuous unimodal probability distribution. In the case of symmetric unimodal distributions, the mode and the mathematical expectation coincide. It is most likely that Gauss had symmetric distributions in mind because, at that time, error distributions were assumed to be symmetric.

Here, we see that precisely comparing the Gauss inequality and the original Chebyshev 1867 inequality is impossible, because Chebyshev proved his inequality for discrete random variables having only a finite number of values, whereas Gauss' statement is valid for continuous distributions, but with the restriction of unimodality. However, from the standpoint of bounding probability content, the Gauss inequality is much better than the Chebyshev inequality; at the same time, the latter is more suitable for the consideration of convergence process related to the law of large numbers. (The law of large numbers with relevance to the assessment of the measurement data quality is discussed in Monhor and Takemoto [15].)

In 1853 (i.e., 14 years after Chebyshev), Irenée-Jules Bienaymé (1796–1878) [4], a French statistician, proved an inequality that is identical with the Chebyshev

https://doi.org/10.1017/S0269964807070179 Published online by Cambridge University Press

inequality, but valid for continuous probability distributions without the restriction of unimodality.

We note another form of the Chebyshev inequality.

CHEBYSHEV INEQUALITY: *If a random variable ξ has finite variance, then, for any $\lambda > 1$, we have*

$$\mathbf{P}(|\xi - \mathbf{E}(\xi)| > \lambda\sigma(\xi)) \leq \frac{1}{\lambda^2}. \tag{2.4}$$

Remark 2.1: In (2.4), $\mathrm{Var}(\xi) = \sigma(\xi)$. If $0 < \lambda \leq 1$, then we have $1/\lambda^2 \geq 1$; that is, the right-hand side of (2.4) becomes equal to or greater than 1. This means that the case $0 < \lambda \leq 1$ leads to a trivial inequality. Consequently, the condition that $\lambda > 1$ is not a restriction, but avoids the trivial case.

Remark 2.2: Chebyshev originally proved his inequality in the form (2.4), which is more similar to the Gauss inequality.

Bickel and Krieger [3] considered the accuracy of the Chebyshev inequality and, recently, Monhor and Takemoto [15] observed that the increase in λ improves the accuracy of the Chebyshev inequality in comparison with the Normal distribution.

3. CHEBYSHEV INEQUALITY IN MULTIDIMENSIONAL SPACES

Berge [2] generalized inequality (2.4) to the two-dimensional space as follows. If ξ_1 and ξ_2 are bivariate random variables with zero means, variances $\sigma(\xi_1)$ and $\sigma(\xi_2)$, and correlation coefficient ρ, then we have

$$\mathbf{P}(|\xi_1| > \lambda\sigma(\xi_1) \quad \text{or} \quad |\xi_2| > \lambda\sigma(\xi_2)) \leq \frac{1 + \sqrt{1 - \rho}}{\lambda^2}. \tag{3.1}$$

It is Berge [2] who first considered the Chebyshev inequality in the multivariate setting. Thus, his result is the first multivariate Chebyshev inequality. After Berge, Lal [12] derived another multivariate Chebyshev inequality considering two different λ's in inequality (3.1). The Lal inequality is as follows:

$$\mathbf{P}(|\xi_1| > \lambda_1\sigma(\xi_1) \quad \text{or} \quad |\xi_2| > \lambda_2\sigma(\xi_2)) \leq \frac{\lambda_1^2 + \lambda_2^2 + \sqrt{(\lambda_1^2 + \lambda_2^2)^2 - 4\rho^2\lambda_1^2\lambda_2^2}}{2\rho^2\lambda_1^2\lambda_2^2}. \tag{3.2}$$

We see that the Lal inequality is much more complicated, although it establishes mathematically a more general environment.

https://doi.org/10.1017/S0269964807070179 Published online by Cambridge University Press

Olkin and Pratt [16] generalized the above results to the arbitrary n-dimensional space. Let $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ be a random vector with zero mean and with variances $\sigma_1, \sigma_2, \ldots, \sigma_n$. Let ρ_{ij} be the elements of the correlation matrix of the random vector. Furthermore, we introduce the notations $t = \sum(1/k_i)$, $k_i > 0$, and $u = \sum \rho_{ij}/(k_i k_j)$. The theorem of Olkin and Pratt then reads as follows:

THEOREM (Olkin and Pratt):

$$\mathbf{P}(|\xi_i| > k_i \sigma_i \text{ for some } i) \leq \frac{n-1}{n}\, t - \frac{n-2}{n^2}\, u + \frac{2}{n^2}\sqrt{u(nt-u)(n-1)}. \qquad \textbf{(3.3)}$$

We see that the right-hand side of inequality (3.3) is quite complicated and it makes the inequality impractical. However, from the standpoint of pure mathematics, the result of Olkin and Pratt is the culmination point in the multivariate generalization of the classic Chebyshev inequality.

4. CONVENTIONS ON NOTATIONS AND AN AUXILIARY LEMMA

Throughout the article we will denote the complement of a set S in $\mathbb{R}^n$ by S^c and the indicator function of S by $I_S(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^n$; that is, we have

$$I_S(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} \in S \\ 0 & \text{if } \mathbf{x} \in S^c. \end{cases}$$

The Euclidean ball with center at $\mathbf{x}$, $\mathbf{x} \in \mathbb{R}^n$ and radius r ($r > 0$) will be denoted by $B(\mathbf{x}, r)$:

$$B(\mathbf{x}, r) = \left\{ \mathbf{y} \in \mathbb{R}^n : \sqrt{\sum_{i=1}^{n} (y_i - x_i)^2} \leq r \right\}.$$

The determinant of an $n \times n$ matrix $\mathbf{A} = (a_{ij})$ will be denoted by $|\mathbf{A}|$. In order to simplify and to shorten mathematical technicalities in the proof of the theorem in Section 5, we introduce a function $\varphi(t, \mathbf{A})$ of real variable t and of matrix parameter $\mathbf{A}$ and prove a property of the function $\varphi(t, \mathbf{A})$ in the form of an independent lemma. The definition of the function $\varphi(t, \mathbf{A})$ is as follows: $\varphi(t, \mathbf{A}) = |\mathbf{I} + \mathbf{tA}|$. Furthermore, in what follows we will also use the standard notation $tr(\mathbf{A}) = \sum_{i=1}^{n} a_{ii}$.

LEMMA: *Let* $\mathbf{A}$ *be an* $n \times n$ *matrix and* $\mathbf{I}$ *be a unit matrix of the same order. Then we have*

$$\left.\frac{d\varphi(t, \mathbf{A})}{dt}\right|_{t=0} = \frac{d}{dt}\left(|\mathbf{I} + \mathbf{tA}|\right)\Big|_{t=0} = tr(\mathbf{A})$$

https://doi.org/10.1017/S0269964807070179 Published online by Cambridge University Press

PROOF: $\mathbf{I} + \mathbf{tA} = (\delta_{ij} + ta_{ij})$, where δ_{ij} are Kroneker symbols; that is,

$$\delta_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j. \end{cases}$$

By the definition of determinants we have

$$|\mathbf{I} + \mathbf{tA}| = \sum (-1)^{\text{sign}(\tau)} (\delta_{1\tau(1)} + ta_{1\tau(1)})(\delta_{2\tau(2)} + ta_{2\tau(2)}) \ldots (\delta_{n\tau(n)} + ta_{n\tau(n)}), \tag{4.1}$$

where the summation is taken over all permutations τ of the numbers $1, 2, \ldots, n$, and $\text{sign}(\tau)$ equals 0 or $+1$ depending on whether τ is even or odd.

Now we write (4.1) in the following form:

$$|\mathbf{I} + \mathbf{tA}| = \prod_{k=1}^{n} (1 + ta_{kk}) + \widehat{\sum} (-1)^{\text{sign}(\tau)} (\delta_{1\tau(1)} + ta_{1\tau(1)}) \times (\delta_{2\tau(2)} + ta_{2\tau(2)}) \ldots (\delta_{n\tau(n)} + ta_{n\tau(n)}), \tag{4.2}$$

where the summation $\widehat{\sum}$ is taken over all of the permutations for τ except $(1, 2, \ldots, n)$.

Differentiating both sides of (4.2), we get the following expression for $d\varphi(t, \mathbf{A})/dt$:

$$\frac{d\varphi(t, \mathbf{A})}{dt} = \frac{d}{dt}\left(\prod_{k=1}^{n} (1 + ta_{kk}) \right) + \frac{d}{dt}\left(\widehat{\sum} (-1)^{\text{sign}(\tau)} (\delta_{1\tau(1)} + ta_{1\tau(1)}) \times (\delta_{2\tau(2)} + ta_{2\tau(2)}) \ldots (\delta_{n\tau(n)} + ta_{n\tau(n)}) \right). \tag{4.3}$$

In (4.3), each product that is of the type

$$(\delta_{1\tau(1)} + ta_{1\tau(1)})(\delta_{2\tau(2)} + ta_{2\tau(2)}) \ldots (\delta_{n\tau(n)} + ta_{n\tau(n)})$$

contains at least two factors of the form

$$\delta_{r\tau(r)} + ta_{r\tau(r)} = ta_{r\tau(r)},$$

https://doi.org/10.1017/S0269964807070179 Published online by Cambridge University Press

where $r \neq \tau(r)$. This fact implies that

$$\frac{d}{dt}\left(\widehat{\sum}(-1)^{\mathrm{sign}(\tau)}(\delta_{1\tau(1)}+ta_{1\tau(1)})(\delta_{2\tau(2)}+ta_{2\tau(2)})\ldots(\delta_{n\tau(n)}+ta_{n\tau(n)})\right)\Bigg|_{t=0}=0.$$

Consequently, we have

$$\begin{aligned}\frac{d\varphi(t,\mathbf{A})}{dt}\Bigg|_{t=0} &= \frac{d}{dt}\left(\prod_{k=1}^{n}(1+ta_{kk})\right)\Bigg|_{t=0}\\ &= a_{11}\prod_{k=2}^{n}(1+ta_{kk})|_{t=0}+a_{22}\prod_{k=1,k\neq 2}^{n}(1+ta_{kk})|_{t=0}\\ &\quad+\cdots+a_{nn}\prod_{k=1}^{n-1}(1+ta_{kk})|_{t=0}.\end{aligned}$$

This means that

$$\frac{d\varphi(t,\mathbf{A})}{dt}\Bigg|_{t=0}=tr(\mathbf{A});$$

hence, the lemma is proven. ■

5. A MULTIVARIATE CHEBYSHEV INEQUALITY

THEOREM: *Let the random vector $\boldsymbol{\xi}=(\xi_1,\xi_2,\ldots,\xi_n)$ have a multivariate Normal distribution with mean vector $\mathbf{m}=(m_1,m_2,\ldots,m_n)$. For any $r>0$, we then have*

$$\mathbf{P}(\boldsymbol{\xi}\in B(\mathbf{m},r))\geq 1-\frac{\sum_{i=1}^{n}\mathrm{Var}(\xi_i)}{r^2}. \tag{5.1}$$

PROOF: Let $f(\mathbf{x}),\mathbf{x}\in\mathbb{R}^n$ denote the probability density of $\boldsymbol{\xi}=(\xi_1,\xi_2,\ldots,\xi_n)$. Also, we designate the inverse of the covariance matrix of $\boldsymbol{\xi}$ by $\boldsymbol{\Sigma}$. Here, instead of the covariance matrix itself, we consider its inverse matrix in order to make the mathematical manipulations easier.

Both the summation and the inverse of the covariance matrix were designated by the same capital Greek Σ, which does not cause any ambiguity or confusion; also, we followed the well-established tradition of the use of Σ for both the covariance matrix and its inverse in multivariate statistical analysis (e.g., see Anderson [1]).

In the light of the above notation, we have

$$f(\mathbf{x})=\frac{\sqrt{|\boldsymbol{\Sigma}|}}{(2\pi)^{n/2}}\exp\left\{-\frac{1}{2}(\mathbf{x}-\mathbf{m})\boldsymbol{\Sigma}(\mathbf{x}-\mathbf{m})'\right\},$$

https://doi.org/10.1017/S0269964807070179 Published online by Cambridge University Press

where $\mathbf{\Sigma}$ is a symmetric, positive-definite $n \times n$ matrix. Hence, it is obvious that if $t > 0$, then $\mathbf{\Sigma} + t\mathbf{I}$, where $\mathbf{I}$ stands for $n \times n$ unit matrix, is also symmetric and positive-definite. This means that $\mathbf{\Sigma} + t\mathbf{I}$ can be considered as the inverse matrix of covariance of a multivariate Normal distribution. Consequently, for any $t > 0$, we have

$$\frac{\sqrt{|\mathbf{\Sigma} + t\mathbf{I}|}}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mathbf{m})(\mathbf{\Sigma}+t\mathbf{I})(\mathbf{x}-\mathbf{m})'\right\} dx_1\, dx_2 \ldots dx_n = 1. \quad \textbf{(5.2)}$$

In (5.2), we carry out the following algebraic transformations:

$$\frac{\sqrt{|\mathbf{\Sigma}|}\cdot\sqrt{|\mathbf{I}+t\mathbf{\Sigma}^{-1}|}}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mathbf{m})(\mathbf{\Sigma}+t\mathbf{I})(\mathbf{x}-\mathbf{m})'\right\} dx_1\, dx_2 \ldots dx_n = 1,$$

or

$$\frac{\sqrt{|\mathbf{\Sigma}|}}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mathbf{m})(\mathbf{\Sigma}+t\mathbf{I})(\mathbf{x}-\mathbf{m})'\right\} dx_1\, dx_2 \ldots dx_n$$

$$= \left(\sqrt{|\mathbf{I}+t\mathbf{\Sigma}^{-1}|}\right)^{-1/2}. \quad \textbf{(5.3)}$$

Differentiating both sides of (5.3) with respect to t, we obtain

$$-\frac{1}{2}\frac{\sqrt{|\mathbf{\Sigma}|}}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \left(\sum_{i=1}^n (x_i - m_i)^2\right) \exp\left\{-\frac{1}{2}(\mathbf{x}-\mathbf{m})(\mathbf{\Sigma}+t\mathbf{I})(\mathbf{x}-\mathbf{m})'\right\} dx_1\, dx_2 \ldots dx_n$$

$$= -\frac{1}{2}\left(\sqrt{|\mathbf{I}+t\mathbf{\Sigma}^{-1}|}\right)^{-3/2} \frac{d}{dt}\left(\varphi(t, \mathbf{\Sigma}^{-1})\right). \quad \textbf{(5.4)}$$

It is easy to see that $tr(\mathbf{\Sigma}^{-1}) = \sum_{i=1}^n \mathrm{Var}(\xi_i)$. Taking $t \searrow 0$ to both sides of (5.4) and using the lemma, we get

$$\frac{\sqrt{|\mathbf{\Sigma}|}}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \frac{\sum_{i=1}^n (x_i - m_i)^2}{r^2} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mathbf{m})\mathbf{\Sigma}(\mathbf{x}-\mathbf{m})'\right\} dx_1\, dx_2 \ldots dx_n$$

$$= \frac{1}{r^2}\sum_{i=1}^n \mathrm{Var}(\xi_i).$$

https://doi.org/10.1017/S0269964807070179 Published online by Cambridge University Press

On the other hand,

$$\begin{aligned}
&\mathbf{P}(\boldsymbol{\xi} \in B(\mathbf{m}, r)) \\
&\quad = 1 - \frac{\sqrt{|\boldsymbol{\Sigma}|}}{(2\pi)^{n/2}} \int_{B^c(\mathbf{m}, r)} \exp\left\{ -\frac{1}{2} (\mathbf{x} - \mathbf{m}) \boldsymbol{\Sigma} (\mathbf{x} - \mathbf{m})' \right\} dx_1\, dx_2 \ldots dx_n \\
&\quad = 1 - \frac{\sqrt{|\boldsymbol{\Sigma}|}}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} I_{B^c}(\mathbf{m}, r)(\mathbf{x}) \exp\left\{ -\frac{1}{2} (\mathbf{x} - \mathbf{m}) \boldsymbol{\Sigma} (\mathbf{x} - \mathbf{m})' \right\} dx_1\, dx_2 \ldots dx_n \\
&\quad \geq 1 - \frac{\sqrt{|\boldsymbol{\Sigma}|}}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \frac{\sum_{i=1}^{n} (x_i - m_i)^2}{r^2} \exp\left\{ -\frac{1}{2} (\mathbf{x} - \mathbf{m}) \boldsymbol{\Sigma} (\mathbf{x} - \mathbf{m})' \right\} dx_1\, dx_2 \ldots dx_n \\
&\quad = 1 - \frac{\sum_{i=1}^{n} \mathrm{Var}(\xi_i)}{r^2},
\end{aligned}$$

or

$$\mathbf{P}(\boldsymbol{\xi} \in B(\mathbf{m}, r)) \geq 1 - \frac{\sum_{i=1}^{n} \mathrm{Var}(\xi_i)}{r^2};$$

thus, the theorem is proven. ■

Jeng [10,11] proved the following inequality, which gives an upper bound on the probability content of a Euclidean ball for a multivariate Normal distribution.

JENG'S INEQUALITY: *Assume that* $\mathbf{B}$ *is a n-dimensional positive-definite matrix with eigenvalues* $\rho_1 \geq \rho_2 \geq \cdots \geq \rho_n > 0$ *and* $\mathbf{B} = \mathbf{JDJ}$, *where* $\mathbf{D}$ *is a diagonal matrix that consists of the eigenvalues.*

Then, for any $r > 0$, *we have*

$$\begin{aligned}
&\frac{1}{\sqrt{(2\pi)^n |\mathbf{B}|}} \int_{B^c(\mathbf{m}, r)} \exp\left\{ -\frac{1}{2} (\mathbf{x} - \mathbf{m}) \mathbf{B}^{-1} (\mathbf{x} - \mathbf{m})' \right\} dx_1\, dx_2 \ldots dx_n \\
&\quad \geq q \left(\frac{r}{\sqrt{\rho_n}} \right)^{n-2} \exp\left\{ -\frac{r^2}{2\rho_n} \right\}, \qquad \textbf{(5.5)}
\end{aligned}$$

where q *is a constant that does not depend on* r.

https://doi.org/10.1017/S0269964807070179 Published online by Cambridge University Press

Remark 5.1: The proof of Jeng's inequality can be found in Jeng and Woods [11]. For computing and approximating the constant q, the reader is referred to Wozencraft and Jacobs [18, p. 82]. Note that in Jeng's inequality, $\mathbf{B}$ denotes the covariance matrix of the multivariate Normal random vector in question.

Remark 5.2: The bounding method proposed by Hassibi and Boyd [9] is not an explicit inequality, but an algorithmic procedure that is computationally quite complicated.

6. CONCLUDING REMARKS

From inequalites (3.1)–(3.3), we find that the right-hand sides of the inequalities of Lal [12], Berge [2], and Olkin and Pratt [16] are quite different from that of the original Chebyshev inequality. Unlikely as it might seem, the right-hand side of the new multivariate Chebyshev inequality is exactly the same as that of the classic one-dimensional Chebyshev inequality. In this sense, the multivariate generalization of the Chebyshev inequality given in this article can be considered as the most natural variant. This is an interesting fact.

In one-dimensional space, the usual interval that is the complementary event for the classic Chebyshev inequality can be conceived as both a "one-dimensional rectangle" and a "one-dimensional ball" simultaneously. As is well known, this "occasional coincidence" of these two sets (in the language of probability theory, "events") disappears and they become the sets of different structure in cases when the dimension of the space is higher than one. In multidimensional generalizations of the classic Chebyshev inequality, Berge [2], Lal [12], and Olkin and Pratt [16] followed the line of a rectangle set. However, here we followed the line of a ball, the consideration of which stemmed from the need coming from the engineering problems treated by Hassibi and Boyd [9], Jeng and Woods [11], and Molina et al. [13].

In a practical setting, the most important feature of the new multivariate Chebyshev inequality is its easy-to-compute form. The simultaneous application of the established multivariate Chebyshev inequality and Jeng's inequality is useful in practical problems by providing lower and upper bounds on the probability content. For example, this approach supplies us with an easy-to-compute estimation of the GPS verification probability considered in [9].

Apart from the establishment of the new multivariate Chebyshev inequality, in the present article we traced the very origin of the Chebyshev inequality back to the original source works by Gauss, Bienaymé, and Chebyshev, which is an interesting fact in its own right and adds an expository element to the topic. In addition, these facts make the article self-contained and convey a tutorial value, too.

References

1. Anderson, T.W. (1958). *An Introduction to multivariate statistical analysis*. New York: Wiley.
2. Berge, P.O. (1937). A note on a form of Tchebycheff's theorem for two variables. *Biometrika* 29: 405–406.

https://doi.org/10.1017/S0269964807070179 Published online by Cambridge University Press

3. Bickel, P.J. & Krieger, A.M. (1992). Extensions of Chebyshev's inequality with applications. *Probability and Mathematical Statistics* 13: 293–310.
4. Bienaymé, I.J. (1853). Considérations à l'appui de la découverte de Laplace sur loi de probabilité dans la méthode des moindres carrés. *Comptes Rendus de l'Academie des Sciences, Paris* 37: 309–324.
5. Chebyshev, P.L. (1867). Des valeurs Moyennes, *Journal de Mathématique Pures et Appliquées* 12: 177–184.
6. Chebyshev, P.L. (1929). In D.E. Smith (ed.), *A source book in mathematics*. New York: McGraw-Hill, pp. 580–587.
7. Gauss, C.F. (1823). *Theoria Combinationis Observationum Erroribus minimis Obnoxiae*. Pars Prior, Pars Posterior, Supplementum. Göttingen: Dieterich.
8. Gauss, K.F. (1995). *Theory of the combination of observations least subject to errors*, Part One, Part Two, Supplement. G.W. Stewart (trans.). Philadelphia: SIAM.
9. Hassibi, A. & Boyd, S. (1998). Integer parameter estimation in linear models with applications to GPS. *IEEE Transactions on Signal Processing* 46: 2938–2952.
10. Jeng, F.C. (1988). Compound Gauss–Markov random fields for image estimation and restoration. PhD thesis, Renselaer Polytechnic Institute, Troy, NY.
11. Jeng F.C. & Woods, J.W. (1990). Simulated annealing compound Gauss–Markov random fields. *IEEE Transactions on Information Theory* 36: 94–107.
12. Lal, D.N. (1955). A note on a form of Tchebycheff's inequality for two variables. *Sankhya* 15: 317–320.
13. Molina, R., Katseggelos, A.K., Mateos, J., Hermoso, A., & Segall, C.A. (2000). Restoration of severely blurred high range images using stochastic and deterministic relaxation algorithms in compound Gauss-Markov random fields. *Pattern Recognition* 33: 555–571.
14. Monhor, D. (1983). An inequality for multivariate Dirichlet distibution. *Acta Mathematica Hungarica* 47: 161–163.
15. Monhor, D. & Takemoto, S. (2005). Understanding the concept of outlier and its relevance to the assessment of data quality: Probabilistic background theory. *Earth, Planets and Space* 57: 1009–1018.
16. Olkin, I. & Pratt, J.W. (1958). A multivariate Tchebycheff's inequality. *Annals of Mathematical Statistics* 29: 226–234.
17. Ross, S.M. (1993). *Introduction to probability models*, 5th ed. San Diego: Academic Press.
18. Wozencraft, J.M. & Jacobs, I.M. (1965). *Principles of communications engineering*. New York: Wiley.

https://doi.org/10.1017/S0269964807070179 Published online by Cambridge University Press