This is an Accepted Manuscript of an article published by Taylor & Francis Group in The American Statistician, 2017, available online: http://doi.org/10.1080/00031305.2016.1186559

Multivariate Chebyshev Inequality with Estimated Mean and Variance

Bartolomeo Stellato[1], Bart P. G. Van Parys[2], and Paul J. Goulart[1]

[1]Department of Engineering Science, University of Oxford
[2]Operations Research Center, Massachusetts Institute of Technology

Abstract

A variant of the well-known Chebyshev inequality for scalar random variables can be formulated in the case where the mean and variance are estimated from samples. In this paper we present a generalization of this result to multiple dimensions where the only requirement is that the samples are independent and identically distributed. Furthermore, we show that as the number of samples tends to infinity our inequality converges to the theoretical multi-dimensional Chebyshev bound.

Keywords: Chebyshev's Inequality, Probability Bounds, Sampling.

Bartolomeo Stellato (e-mail: *bartolomeo.stellato@eng.ox.ac.uk*) and Paul J. Goulart (e-mail: *paul.goulart@eng.ox.ac.uk*), Control Group, Department of Engineering Science, University of Oxford, Parks Road, Oxford OX1 3PJ, United Kingdom; Bart P. G. Van Parys (e-mail: *vanparys@mit.edu*) Operations Research Center, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139. The authors would like to thank Sergio Grammatico for the fruitful suggestions.

1 Introduction

The Chebyshev inequality (1867) is a fundamental result from probability theory and has been studied extensively for more than a century in a wide range of sciences. The most common version of this result asserts that the probability that a scalar random variable ξ with distribution $\mathbb{P}$ differs from its mean $\mu \in \mathbb{R}$ by more than $\lambda \in \mathbb{R}_{>0}$ standard deviations $\sigma \in \mathbb{R}_{>0}$ satisfies the relation

$$\mathbb{P}\left(|\xi - \mu| \geq \lambda\sigma\right) \leq \min\left\{1, \frac{1}{\lambda^2}\right\}. \tag{1}$$

Recent works by Chen (2011) and Navarro (2013) provide a closed form extension of (1) to the multivariate case where the confidence intervals are ellipsoids centered at the population mean. Moreover, Navarro (2014b) shows that the derived extension is tight. Another extension of (1) to more general ellipsoidal and polyhedral sets has been described in (Vandenberghe et al., 2007) where a multivariate version of the Chebyshev bound is computed as a solution to a semidefinite program (SDP) (Vandenberghe and Boyd, 1996).

Although these results provide means to explicitly compute distribution-free probability bounds based only on the first two moments of $\mathbb{P}$, they are of limited practical value since one often does not know μ and σ exactly. In practice, a common approach is to compute empirical estimates of μ and σ via sampling and to substitute these estimated values into (1), although it can be shown that this method can lead to unreliable results in the event of poor estimates of the moments. There is an extensive literature about empirical processes where the quality of the estimates is investigated, see e.g. Dudley (1978) and van de Geer (2010). However, these approaches suffer from two main problems. They either assume that the underlying distribution $\mathbb{P}$ has bounded support (e.g. Hoeffding's inequality), or they provide asymptotic results on the convergence rate that are valid only as the sample size tends to infinity. Unfortunately, neither of these two cases turns out to be helpful when we make no assumption on the support of the distribution and the number of samples is limited.

In the univariate case, Saw et al. (1984) approached the problem of formulating an empirical Chebyshev inequality from a different direction. Given N i.i.d samples $\xi^{(i)}, \ldots, \xi^{(N)} \in \mathbb{R}$ from an unknown distribution $\mathbb{P}$, and their empirical mean μ_N and empirical standard deviation (SD) σ_N, Saw derives a Chebyshev inequality with respect to the $(N+1)^{\text{th}}$ sample. The bound derived is remarkably simple and requires only a modification of the right-hand side of the theoretical bound in (1), i.e.

$$\begin{aligned}&\mathbb{P}^{N+1}\left(\left|\xi^{(N+1)} - \mu_N\right| \geq \lambda\sigma_N\right)\\ &\quad\leq \min\left\{1, \frac{1}{N+1}\left\lfloor\frac{(N+1)(N^2 - 1 + N\lambda^2)}{N^2\lambda^2}\right\rfloor\right\},\end{aligned} \tag{2}$$

where $\lfloor\cdot\rfloor$ denotes the floor function[2].

Currently, there exists no counterpart of (2) for the multivariate case. There have been only limited efforts to extend these results to multiple dimensions by making strong assumptions on the

[2] λ corresponds to $k\sqrt{\frac{N+1}{N}}$ in (Saw et al., 1984).

population. In (Navarro, 2014a) the author derives a multivariate equivalent by assuming that the true distribution of the population is the empirical distribution over a given data set.

In this paper we derive a multivariate version of the inequality in (2) using the Euclidean norm, without requiring any further assumptions on the distrubution. In addition, we show that the result converges to the multivariate Chebyshev inequality as computed in (Vandenberghe et al., 2007) for an ellipsoidal set centered at the mean.

2 Main Results

Before stating the main result, we require the following definition:

Definition 2.1. *Let $\boldsymbol{\xi} \in \mathbb{R}^{n_{\boldsymbol{\xi}}}$ be a random variable and let $N \in \mathbb{Z}_{\geq n_{\boldsymbol{\xi}}}$. Given $(N+1)$ i.i.d. samples $\boldsymbol{\xi}^{(1)}, \ldots, \boldsymbol{\xi}^{(N)}, \boldsymbol{\xi}^{(N+1)} \in \mathbb{R}^{n_{\boldsymbol{\xi}}}$ of $\boldsymbol{\xi}$ with mean $\mu \in \mathbb{R}^{n_{\boldsymbol{\xi}}}$ and covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{n_{\boldsymbol{\xi}} \times n_{\boldsymbol{\xi}}}$, we define the* empirical mean *as*

$$\boldsymbol{\mu}_N \coloneqq \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{\xi}^{(i)}, \tag{3}$$

and the unbiased *and* biased empirical covariances *as*

$$\boldsymbol{\Sigma}_N \coloneqq \frac{1}{N-1} \sum_{i=1}^{N} (\boldsymbol{\xi}^{(i)} - \boldsymbol{\mu}_N)(\boldsymbol{\xi}^{(i)} - \boldsymbol{\mu}_N)^\top, \qquad \hat{\boldsymbol{\Sigma}}_N \coloneqq \frac{1}{N} \sum_{i=1}^{N} (\boldsymbol{\xi}^{(i)} - \boldsymbol{\mu}_N)(\boldsymbol{\xi}^{(i)} - \boldsymbol{\mu}_N)^\top,$$

respectively.

We can now state our main result, which is a multivariate version of the univariate result of Saw et al. (1984):

Theorem 2.1. *Let $\boldsymbol{\xi} \in \mathbb{R}^{n_{\boldsymbol{\xi}}}$ be a random variable and let $N \in \mathbb{Z}_{\geq n_{\boldsymbol{\xi}}}$. Given $N+1$ i.i.d samples of $\boldsymbol{\xi}$ denoted as $\boldsymbol{\xi}^{(1)}, \ldots, \boldsymbol{\xi}^{(N)}, \boldsymbol{\xi}^{(N+1)} \in \mathbb{R}^{N_{\boldsymbol{\xi}}}$, if we assume that $\boldsymbol{\Sigma}_N$ is nonsingular, then for all $\lambda \in \mathbb{R}_{>0}$ it holds that:*

$$\begin{aligned} &\mathbb{P}^{N+1}\left((\boldsymbol{\xi}^{(N+1)} - \boldsymbol{\mu}_N)^\top \boldsymbol{\Sigma}_N^{-1} (\boldsymbol{\xi}^{(N+1)} - \boldsymbol{\mu}_N) \geq \lambda^2\right) \\ &\quad \leq \min\left\{1, \frac{1}{N+1}\left\lfloor \frac{n_{\boldsymbol{\xi}}(N+1)(N^2 - 1 + N\lambda^2)}{N^2 \lambda^2} \right\rfloor\right\}. \end{aligned} \tag{4}$$

Remark. *The inequality* (4) *can be simplified by upper bounding the floor function by its argument*

$$\begin{aligned} &\mathbb{P}^{N+1}\left((\boldsymbol{\xi}^{(N+1)} - \boldsymbol{\mu}_N)^\top \boldsymbol{\Sigma}_N^{-1} (\boldsymbol{\xi}^{(N+1)} - \boldsymbol{\mu}_N) \geq \lambda^2\right) \\ &\qquad \leq \min\left\{1, \frac{n_{\boldsymbol{\xi}}(N^2 - 1 + N\lambda^2)}{N^2 \lambda^2}\right\}. \end{aligned}$$

We can also show that our empirical bound is well behaved in the limit as $N \to \infty$, coinciding with the (tight) analytical bound computable using the method of Vandenberghe et al. (2007):

Theorem 2.2. *As $N \to \infty$, the right-hand side of (4) tends to*

$$\min\left\{1, \frac{n_{\boldsymbol{\xi}}}{\lambda^2}\right\},$$

which corresponds to the Multivariate Chebyshev inequality over ellipsoids shaped according to $\boldsymbol{\Sigma}$ and centered in $\boldsymbol{\mu}$.

3 Proof of the Main Results

In order to prove our main results, we require two supporting lemmas.

Lemma 3.1. *Let $k \in \mathbb{R}_{>0}$ and $N \in \mathbb{Z}_{\geq n_{\boldsymbol{\xi}}}$. Consider a set of vectors $\mathcal{U}_N \coloneqq \{\boldsymbol{u}_i\}_{i=1}^N$ with $\boldsymbol{u}_i \in \mathbb{R}^{n_{\boldsymbol{\xi}}}$ for all $i \in \{1, \ldots, N\}$ satisfying the conditions*

$$\sum_{i=1}^N \boldsymbol{u}_i = 0_{n_{\boldsymbol{\xi}}}, \qquad \sum_{i=1}^N \boldsymbol{u}_i \boldsymbol{u}_i^\top = N \boldsymbol{I}_{n_{\boldsymbol{\xi}} \times n_{\boldsymbol{\xi}}}.$$

Define the subset of vectors in $\mathcal{U}_N$ with norm greater or equal to k as

$$J(\mathcal{U}_N, k) \coloneqq \{\boldsymbol{u}_i \in \mathcal{U}_N \ : \ \|\boldsymbol{u}_i\|_2 \geq k\},$$

where $\|\cdot\|_2$ is the Euclidean norm. Then the cardinality of $J(\mathcal{U}_N, k)$ is bounded by $|J(\mathcal{U}_N, k)| \leq \left\lfloor \frac{n_{\boldsymbol{\xi}} N}{k^2} \right\rfloor$.

Proof. Observe that

$$\|\boldsymbol{u}_i\|_2 \geq k \quad \iff \quad \boldsymbol{u}_i^\top \boldsymbol{u}_i \geq k^2.$$

Summing both sides of the preceding inequality over $J(\mathcal{U}_N, k)$ produces

$$k^2 \, |J(\mathcal{U}_N, k)| \leq \sum_{\boldsymbol{u}_i \in J(\mathcal{U}_N, k)} \boldsymbol{u}_i^\top \boldsymbol{u}_i \leq \sum_{i=1}^N \boldsymbol{u}_i^\top \boldsymbol{u}_i = \operatorname{tr}\left(\sum_{i=1}^N \boldsymbol{u}_i \boldsymbol{u}_i^\top\right) = n_{\boldsymbol{\xi}} N$$

and the result follows immediately. ■

Lemma 3.2. *The following relations hold:*

$$\boldsymbol{\xi}^{(N+1)} - \boldsymbol{\mu}_{N+1} = \frac{N}{N+1}(\boldsymbol{\xi}^{(N+1)} - \boldsymbol{\mu}_N) \tag{5}$$

$$\hat{\boldsymbol{\Sigma}}_{N+1} = \frac{N-1}{N+1}\boldsymbol{\Sigma}_N + \frac{N}{(N+1)^2}(\boldsymbol{\xi}^{(N+1)} - \boldsymbol{\mu}_N)(\boldsymbol{\xi}^{(N+1)} - \boldsymbol{\mu}_N)^\top. \tag{6}$$

Proof. The first relation can be obtained directly by writing

$$\boldsymbol{\xi}^{(N+1)} - \boldsymbol{\mu}_{N+1} = \boldsymbol{\xi}^{(N+1)} - \frac{1}{N+1}\left(\boldsymbol{\xi}^{(N+1)} + \sum_{i=1}^{N} \boldsymbol{\xi}^{(i)}\right)$$

and collecting terms; the same result appears in (Welford, 1962). The second relation can be found by first defining the partial sums $\boldsymbol{S}_N$ as

$$\boldsymbol{S}_N \coloneqq \sum_{i=1}^{N} \left(\boldsymbol{\xi}^{(i)} - \boldsymbol{\mu}_N\right)\left(\boldsymbol{\xi}^{(i)} - \boldsymbol{\mu}_N\right)^{\top},$$

for which Welford (1962) provides (with obvious modifications) the recurrence relation

$$\boldsymbol{S}_{N+1} = \boldsymbol{S}_N + \frac{N}{N+1}\left(\boldsymbol{\xi}^{(N+1)} - \boldsymbol{\mu}_N\right)\left(\boldsymbol{\xi}^{(N+1)} - \boldsymbol{\mu}_N\right)^{\top}.$$

The result then follows by applying the identities $(N+1)\hat{\boldsymbol{\Sigma}}_{N+1} = \boldsymbol{S}_{N+1}$ and $(N-1)\boldsymbol{\Sigma}_N = \boldsymbol{S}_N$. ■

We are now in a position to prove both of our main results:

Proof of Theorem 2.1

Since $\boldsymbol{\Sigma}_N$ is assumed nonsingular, it follows that $\hat{\boldsymbol{\Sigma}}_{N+1} \succeq \boldsymbol{\Sigma}_N \succ 0$, i.e. $\hat{\boldsymbol{\Sigma}}_{N+1}$ is positive definite. Normalize each of the $N+1$ samples $\boldsymbol{\xi}^{(i)}$ using

$$\boldsymbol{u}_i \coloneqq \left(\hat{\boldsymbol{\Sigma}}_{N+1}\right)^{-1/2}\left(\boldsymbol{\xi}^{(i)} - \boldsymbol{\mu}_{N+1}\right), \quad \forall i \in \{1, \ldots, N+1\} \tag{7}$$

so that

$$\sum_{i=1}^{N+1} \boldsymbol{u}_i = 0_d, \quad \sum_{i=1}^{N+1} \boldsymbol{u}_i \boldsymbol{u}_i^{\top} = (N+1)\boldsymbol{I}_{n_{\boldsymbol{\xi}} \times n_{\boldsymbol{\xi}}}, \tag{8}$$

and (8) satisfies Lemma 3.1. Since all of the vectors $\boldsymbol{u}_i$ are i.i.d. and not more than $J(\mathcal{U}_{N+1}, k)$ of these $N+1$ vectors have norm greater or equal to k, we have from Lemma 3.1 that

$$\mathbb{P}^{N+1}\left(\|\boldsymbol{u}_{N+1}\|_2 \geq k\right) \leq \frac{J(\mathcal{U}_{N+1}, k)}{N+1} = \frac{1}{N+1}\left\lfloor \frac{n_{\boldsymbol{\xi}}(N+1)}{k^2} \right\rfloor. \tag{9}$$

Considering next the inequality $\|\boldsymbol{u}_{N+1}\|_2 \geq k$, apply (5) and (7) to obtain the equivalent condition

$$\left(\boldsymbol{\xi}^{(i)} - \boldsymbol{\mu}_{N+1}\right)^{\top} \hat{\boldsymbol{\Sigma}}_{N+1}^{-1}\left(\boldsymbol{\xi}^{(i)} - \boldsymbol{\mu}_{N+1}\right) \geq k^2.$$

By using Lemma 3.2 it is possible to define $\hat{\boldsymbol{\Sigma}}_{N+1}$ as

$$\hat{\boldsymbol{\Sigma}}_{N+1} = \frac{N-1}{N+1}\boldsymbol{\Sigma}_N + \frac{1}{N}\left(\boldsymbol{\xi}^{(i)} - \boldsymbol{\mu}_{N+1}\right)\left(\boldsymbol{\xi}^{(i)} - \boldsymbol{\mu}_{N+1}\right)^{\top}. \tag{10}$$

Let us define $q_N \coloneqq \left(\boldsymbol{\xi}^{(i)} - \boldsymbol{\mu}_N\right)^\top \boldsymbol{\Sigma}_N^{-1} \left(\boldsymbol{\xi}^{(i)} - \boldsymbol{\mu}_N\right)$ and $q_{N+1} \coloneqq \left(\boldsymbol{\xi}^{(i)} - \boldsymbol{\mu}_{N+1}\right)^\top \boldsymbol{\Sigma}_N^{-1} \left(\boldsymbol{\xi}^{(i)} - \boldsymbol{\mu}_{N+1}\right)$. Note that $q_{N+1} = \frac{N^2}{(N+1)^2} q_N$ from (5). Applying the Sherman-Woodbury-Morrison identity (Sherman and Morrison, 1949) to (10) we can invert matrix $\hat{\boldsymbol{\Sigma}}_{N+1}$ obtaining

$$\begin{aligned}\left(\boldsymbol{\xi}^{(i)} - \boldsymbol{\mu}_{N+1}\right)^\top \hat{\boldsymbol{\Sigma}}_{N+1}^{-1} \left(\boldsymbol{\xi}^{(i)} - \boldsymbol{\mu}_{N+1}\right) &= \frac{N+1}{N-1} q_{N+1} - \left(1 + \frac{1}{N}\frac{N+1}{N-1} q_{N+1}\right)^{-1} \frac{1}{N}\left(\frac{N+1}{N-1}\right)^2 q_{N+1}^2 \\ &= \frac{(N+1) q_{N+1}}{(N-1)N + (N+1) q_{N+1}} \geq k^2.\end{aligned}$$

where the first equality has been pre- and post-multiplied by $\left(\boldsymbol{\xi}^{(i)} - \boldsymbol{\mu}_{N+1}\right)^\top$ and $\left(\boldsymbol{\xi}^{(i)} - \boldsymbol{\mu}_{N+1}\right)$ respectively. The latter inequality can be rewritten in terms of q_N and rearranged to

$$q_N \geq \frac{(N^2-1)k^2}{N(N-k^2)},$$

so that (9) is equivalent to

$$\mathbb{P}^{N+1}\left(\left(\boldsymbol{\xi}^{(i)} - \boldsymbol{\mu}_N\right)^\top \boldsymbol{\Sigma}_N^{-1} \left(\boldsymbol{\xi}^{(i)} - \boldsymbol{\mu}_N\right) \geq \frac{(N^2-1)k^2}{N(N-k^2)}\right) \leq \frac{1}{N+1}\left\lfloor \frac{n_{\boldsymbol{\xi}}(N+1)}{k^2} \right\rfloor. \tag{11}$$

Finally, define λ such that

$$\lambda^2 = \frac{(N^2-1)k^2}{N(N-k^2)}, \quad \text{so that} \quad k^2 = \frac{N^2\lambda^2}{N^2 - 1 + N\lambda^2}.$$

Direct substitution into (11) then produces the desired inequality

$$\mathbb{P}^{N+1}\left((\boldsymbol{\xi}^{(N+1)} - \boldsymbol{\mu}_N)^\top \boldsymbol{\Sigma}_N^{-1} (\boldsymbol{\xi}^{(N+1)} - \boldsymbol{\mu}_N) \geq \lambda^2\right) \leq \frac{1}{N+1}\left\lfloor \frac{n_{\boldsymbol{\xi}}(N+1)(N^2-1+N\lambda^2)}{N^2\lambda^2} \right\rfloor. \qquad \blacksquare$$

Proof of Theorem 2.2

Given $\boldsymbol{\mu} \in \mathbb{R}^{n_{\boldsymbol{\xi}}}$ and $\boldsymbol{\Sigma} \in \mathbb{R}^{n_{\boldsymbol{\xi}} \times n_{\boldsymbol{\xi}}}$, $\boldsymbol{\Sigma} \succeq 0$, as the mean and covariance of the random variable $\boldsymbol{\xi} \in \mathbb{R}^{n_{\boldsymbol{\xi}}}$ respectively, we now derive the multivariate Chebyshev inequality bounding the probability

$$\mathbb{P}\left((\boldsymbol{\xi} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{\xi} - \boldsymbol{\mu}) \geq \lambda^2\right), \tag{12}$$

which is the probability of the complement of the ellipsoid shaped by $\boldsymbol{\Sigma}$ centered at the mean $\boldsymbol{\mu}$.

Without loss of generality, we shift the coordinate system to the mean $\boldsymbol{\mu}$ by defining the variable $\boldsymbol{\eta} \coloneqq \boldsymbol{\xi} - \boldsymbol{\mu}$ with zero mean $\boldsymbol{\mu}_{\boldsymbol{\eta}} = 0$ and variance $\boldsymbol{\Sigma}_\eta = \boldsymbol{\Sigma}$. Let us define $\mathcal{E}$ as the ellipsoid

$$\mathcal{E} \coloneqq \left\{ \boldsymbol{\eta}^\top \frac{\boldsymbol{\Sigma}^{-1}}{\lambda^2} \boldsymbol{\eta} - 1 < 0 \right\}.$$

The problem of computing an upper bound on the probability of $\boldsymbol{\eta}$ falling in the complement $\mathcal{E}^c$ of the ellipsoid $\mathcal{E}$ is equivalent to bounding the probability (12). Let $1_{\mathcal{E}^c}(\cdot)$ denote the indicator function of set $\mathcal{E}^c$, i.e. if $1_{\mathcal{E}^c}(\boldsymbol{\eta}) = 0$ if $\boldsymbol{\eta} \notin \mathcal{E}^c$ and $1_{\mathcal{E}^c}(\boldsymbol{\eta}) = 1$ if $\boldsymbol{\eta} \in \mathcal{E}^c$; with the obvious relation $\mathbb{P}(\boldsymbol{\eta} \in \mathcal{E}^c) = \mathbb{E}(1_{\mathcal{E}^c})$. In order to bound the latter, we can define a quadratic function (Boyd and Vandenberghe, 2004, Section 7.4.1) $f(\boldsymbol{\eta}) = \boldsymbol{\eta}^\top \boldsymbol{P}\boldsymbol{\eta} + 2\boldsymbol{q}^\top \boldsymbol{\eta} + r$ such that $f(\boldsymbol{\eta}) \geq 1_{\mathcal{E}^c}(\boldsymbol{\eta})$ for all $\boldsymbol{\eta} \in \mathbb{R}^{n_{\boldsymbol{\eta}}}$. Equivalently, this inequality can be written as $f(\boldsymbol{\eta}) \geq 1, \forall \boldsymbol{\eta} \in \mathcal{E}^c$ and $f(\boldsymbol{\eta}) \geq 0, \forall \boldsymbol{\eta} \in \mathbb{R}^{n_{\boldsymbol{\eta}}}$. By taking the expected value we obtain

$$\mathbb{E}(f(\boldsymbol{\eta})) \geq \mathbb{E}(1_{\mathcal{E}^c}(\boldsymbol{\eta})) = \mathbb{P}(\boldsymbol{\eta} \in \mathcal{E}^c). \tag{13}$$

Hence, the problem of upper-bounding (12) is equivalent to solving the convex problem

$$\begin{aligned} &\text{minimize} && \mathbb{E}(f(\boldsymbol{\eta})) && \text{(14a)}\\ &\text{subject to:} && f(\boldsymbol{\eta}) \geq 1,\ \boldsymbol{\eta} \in \mathcal{E}^c && \text{(14b)}\\ & && f(\boldsymbol{\eta}) \geq 0,\ \boldsymbol{\eta} \in \mathbb{R}^{n_{\boldsymbol{\eta}}}. && \text{(14c)} \end{aligned}$$

Since $f(\boldsymbol{\eta})$ is a quadratic function and we know the first and second moments of $\boldsymbol{\eta}$, we can compute its expected value as

$$\mathbb{E}(f(\boldsymbol{\eta})) = \operatorname{tr}(\Sigma_{\boldsymbol{\eta}} \boldsymbol{P}) + 2\boldsymbol{q}^\top \boldsymbol{\mu}_{\boldsymbol{\eta}} + r = \operatorname{tr}(\Sigma \boldsymbol{P}) + r, \tag{15}$$

where $\boldsymbol{\mu}_{\boldsymbol{\eta}} = 0$ and $\boldsymbol{\Sigma}_{\boldsymbol{\eta}} = \boldsymbol{\Sigma}$. The constraint (14c) can be rewritten as the following linear matrix inequality (LMI)

$$\begin{bmatrix} \boldsymbol{P} & \boldsymbol{q} \\ \boldsymbol{q}^\top & r \end{bmatrix} \succeq 0, \tag{16}$$

see (Vandenberghe and Boyd, 1996). By making use of the S-procedure (Boyd and Vandenberghe, 2004, Section B.2), we can define a scalar $\tau \geq 0$ and rewrite (14b) as another LMI

$$\begin{bmatrix} \boldsymbol{P} & \boldsymbol{q} \\ \boldsymbol{q}^\top & r-1 \end{bmatrix} \succeq \tau \begin{bmatrix} \frac{\boldsymbol{\Sigma}^{-1}}{\lambda^2} & 0 \\ 0 & -1 \end{bmatrix},\ \tau \geq 0. \tag{17}$$

Finally, from (15), (16) and (17) we can rewrite (14) as a Semidefinite Program (SDP) (Vandenberghe and Boyd, 1996):

$$\begin{aligned} &\text{minimize} && \operatorname{tr}(\boldsymbol{\Sigma}\boldsymbol{P}) + r \\ &\text{subject to:} && \begin{bmatrix} \boldsymbol{P} & \boldsymbol{q} \\ \boldsymbol{q}^\top & r-1 \end{bmatrix} \succeq \tau \begin{bmatrix} \frac{\boldsymbol{\Sigma}^{-1}}{\lambda^2} & 0 \\ 0 & -1 \end{bmatrix} \\ & && \begin{bmatrix} \boldsymbol{P} & \boldsymbol{q} \\ \boldsymbol{q}^\top & r \end{bmatrix} \succeq 0, \quad \tau \geq 0. \end{aligned}$$

Since the ellipsoid is centered at the origin, it is possible to choose $\boldsymbol{q} = 0$ and rewrite the problem as:

$$\begin{aligned} &\text{minimize} && \operatorname{tr}(\boldsymbol{\Sigma}\boldsymbol{P}) + r && \text{(18)}\\ &\text{subject to:} && \boldsymbol{P} \succeq \tau \frac{\boldsymbol{\Sigma}^{-1}}{\lambda^2}, \quad r \geq 1 - \tau && \text{(19)}\\ & && \boldsymbol{P} \succeq 0,\ r \geq 0,\ \tau \geq 0. && \text{(20)} \end{aligned}$$

The objective function and the constraints in (18) are linear in the optimization variables. The optimal solution therefore exists at the boundary of the feasible region: i.e. $r = 1 - \tau$. We need to distinguish two different cases. The first corresponds to

$$r = 0 \iff \tau = 1 \implies \boldsymbol{P} = \frac{\boldsymbol{\Sigma}^{-1}}{\lambda^2},$$

with the optimum being $n_{\boldsymbol{\xi}}/\lambda^2$. The second case is

$$\tau = 0 \iff r = 1 \implies \boldsymbol{P} = 0,$$

and the corresponding optimum is 1. Finally, by computing the inverse coordinate transformation to get back $\boldsymbol{\xi}$, it is possible to write the multivariate inequality explicitly as

$$\mathbb{P}\left((\boldsymbol{\xi} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{\xi} - \boldsymbol{\mu}) \geq \lambda^2\right) \leq \min\left\{1, \frac{n_{\boldsymbol{\xi}}}{\lambda^2}\right\}. \qquad \blacksquare$$

4 Applications

Applications of this result include any field wherein the Chebyshev inequality must be applied to distributions for which the mean and covariance are unknown.

A direct application of this empirical Chebyshev inequality is outlier detection. Given a probability bound, it is possible to compute a threshold λ and construct a confidence ellipsoidal set from the sample mean and covariance of the first N samples. Then, if the Mahalanobis distance of the $N + 1^{\text{th}}$ sample exceeds λ, it can be considered an outlier. In (Hardin and Rocke, 2005) a similar approach is described making use of the quantiles of the chi-square- or F-distributions in case of normal data. We expect our bound to give more conservative results than the method proposed in (Hardin and Rocke, 2005), but with more general validity since we make no assumptions on the samples' distribution.

Another application involves solving stochastic optimization problems using data-driven information about the uncertainty without knowing its distribution. Following the approach in (Chen et al., 2007) and (Bertsimas et al., 2013), we can make use of our empirical Chebyshev inequality to construct ellipsoidal uncertainty sets with predefined probability guarantees. We can then approximate stochastic programs, that are intractable in their general form (Shapiro and Nemirovski, 2005), with robust optimization problems (Ben-Tal et al., 2009) and enforce the optimal solution to be feasible for all the uncertainty realizations inside our ellipsoidal uncertainty set. The latter condition implies the same probabilistic guarantees on the original stochastic program. In certain cases, e.g. when the constraints are linear and the uncertainty enters linearly in the coefficients, the robust reformulations are convex and can be solved efficiently as second-order cone programs (SOCPs) (Ben-Tal et al., 2009).

5 Conclusions

We have derived a generalization of the empirical Chebyshev inequality in multiple dimensions with the only requirement that the given samples are independent and identically distributed. The

derived bound scales linearly with the dimension of the random vector and has the same structure as the one-dimensional inequality.

Since many of the common distributions studied in both theory and practice are unimodal, an interesting improvement of this result could be to introduce the assumption of unimodality in order to derive less pessimistic bounds. Another possible extension is to investigate other norms (e.g. ∞ or 1-norm) and compare the right-hand side of the respective reformulations to understand which one is more appropriate for different kinds of distributions.

There are many possible application of this theoretical result appearing whenever the Chebyshev inequality is employed without knowing the population distribution. In particular, the inequality can be exploited to construct confidence sets which can be used in several situations such as outliers detection or stochastic programs reformulations.

References

Ben-Tal, A., L. El Ghaoui, and A. Nemirovski (2009). Robust optimization. Princeton University Press.

Bertsimas, D., V. Gupta, and N. Kallus (2013). Data-driven robust optimization. *arXiv.org*.

Boyd, S. and L. Vandenberghe (2004). *Convex Optimization*. Cambridge University Press.

Chebyshev, P. (1867). Des valeurs moyennes. *Journal de Mathématiques pures et Appliquées 12*(2), 177–184.

Chen, X. (2011). A New Generalization of Chebyshev Inequality for Random Vectors. *arXiv.org*.

Chen, X., M. Sim, and P. Sun (2007, December). A Robust Optimization Perspective on Stochastic Programming. *Operations Research 55*(6), 1058–1071.

Dudley, R. M. (1978). Central Limit-Theorems for Empirical Measures. *Annals of Probability 6*(6), 899–929.

Hardin, J. and D. M. Rocke (2005, December). The Distribution of Robust Distances. *Journal of Computational and Graphical Statistics 14*(4), 928–946.

Navarro, J. (2013). A very simple proof of the multivariate Chebyshev's inequality. *To appear Communications in Statistics - Theory and Methods*.

Navarro, J. (2014a). A note on confidence regions based on the bivariate Chebyshev inequality. Applications to order statistics and data sets. *Journal of the Turkish Statistical Association 7*.

Navarro, J. (2014b, August). Can the bounds in the multivariate Chebyshev inequality be attained? *Statistics and Probability Letters 91*, 1–5.

Saw, J. G., M. C. K. Yang, and T. C. Mo (1984). Chebyshev Inequality With Estimated Mean and Variance. *The American Statistician 38*(2), 130–132.

Shapiro, A. and A. Nemirovski (2005). On Complexity of Stochastic Programming Problems. In *Continuous Optimization*, pp. 111–146. New York: Springer US.

Sherman, J. and W. J. Morrison (1949). Adjustment of an inverse matrix corresponding to changes in the elements of a given column or a given row of the original matrix. In *Annals of Mathematical Statistics*, pp. 621–621.

van de Geer, S. A. (2010, January). *Empirical Processes in M-Estimation*. Cambridge University Press.

Vandenberghe, L. and S. Boyd (1996). Semidefinite Programming. *SIAM Review 38*(1), 49–95.

Vandenberghe, L., S. Boyd, and K. Comanor (2007). Generalized Chebyshev Bounds via Semidefinite Programming. *SIAM Review 49*(1), 52–64.

Welford, B. P. (1962). Note on a method for calculating corrected sums of squares and products. *Technometrics*.