Available online at www.sciencedirect.com
ScienceDirect
Statistics and Probability Letters 78 (2008) 1823–1825
STATISTICS & PROBABILITY LETTERS
www.elsevier.com/locate/stapro

A matrix version of Chernoff inequality

Zhengyuan Wei[a,b,*], Xinsheng Zhang[a]

[a] *Department of Statistics, Fudan University, Shanghai, 200433, PR China*
[b] *Department of Statistics, Chongqing Institute of Technology, Chongqing, 400050, PR China*

Received 15 December 2007; received in revised form 9 January 2008; accepted 17 January 2008
Available online 30 January 2008

Abstract

An interesting result from the point of view of upper variance bounds is the inequality of Chernoff [Chernoff, H., 1981. A note on an inequality involving the normal distribution. Annals of Probability 9, 533–535]. Namely, that for every absolutely continuous function g with derivative g' such that $\mathrm{Var}\{|g(\xi)|\} < \infty$, and for standard normal r.v. ξ, $\mathrm{Var}(|g(\xi)|) \le \mathbb{E}\{(g'(\xi))^2\}$. Both the usefulness and simplicity of this inequality have generated a great deal of extensions, as well as alternative proofs. Particularly, Olkin and Shepp [Olkin, I., Shepp, L., 2005. A matrix variance inequality. Journal of Statistical Planning and Inference 130, 351–358] obtained an inequality for the covariance matrix of k functions. However, all the previous papers have focused on univariate function and univariate random variable. We provide here a covariance matrix inequality for multivariate function of multivariate normal distribution.

1. Introduction

The well-known upper variance bounds obtained by Chernoff (1981) for the standard normal r.v. $\xi \sim N(0, 1)$ are formulated as follows. For every absolutely continuous function g with derivative g' such that $\mathrm{Var}\{|g(\xi)|\} < \infty$, then

$$\mathrm{Var}(|g(\xi)|) \le \mathbb{E}\left\{\left(g'(\xi)\right)^2\right\}, \tag{1}$$

with equality if and only if g is linear. Chen (1982) extended Eq. (1) to the case that $\xi_1, \ldots, \xi_p$ are independent $N(0, 1)$ r.v.s and g defined on $\mathbb{R}^p$. Then

$$\mathrm{Var}(g(\boldsymbol{\xi})) \le \mathbb{E}\left\{\left(\frac{\partial}{\partial \xi_1} g(\boldsymbol{\xi})\right)^2\right\} + \cdots + \mathbb{E}\left\{\left(\frac{\partial}{\partial \xi_p} g(\boldsymbol{\xi})\right)^2\right\}. \tag{2}$$

Olkin and Shepp (2005) provided a matrix version of inequality (1) as follows. Let r.v. $\xi \sim N(0, 1)$ and g_j absolutely continuous functions with $\mathrm{Var}(g_j(\xi)) < \infty$, for $j = 1, \ldots, p$. Write $\boldsymbol{g}(\xi) = \left(g_1(\xi), \ldots, g_p(\xi)\right)^T$. Then

$$\mathbb{E}\left\{\left(\frac{\partial}{\partial \xi}\boldsymbol{g}(\xi)\right)\left(\frac{\partial}{\partial \xi}\boldsymbol{g}(\xi)\right)^T\right\} - \mathrm{Cov}\left(\boldsymbol{g}(\xi), \boldsymbol{g}(\xi)\right) \ge 0. \tag{3}$$

* Corresponding author at: Department of Statistics, Fudan University, Shanghai, 200433, PR China. Tel.: +86 2161443889.
E-mail address: 061025012@fudan.edu.cn (Z. Wei).

0167-7152/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.spl.2008.01.044

(i.e. nonnegative definite matrix.) Many papers deal with the inequality (1), and in all cases they relate to univariate r.v. and univariate function. Multivariate distribution play an important role in statistics and probability. This paper provides a version of covariance matrix bounds for multivariate function of multivariate normal distribution, which extended the results by Olkin and Shepp (2005) and Chen (1982).

2. Main results

Proposition 1. *Let $\boldsymbol{\eta} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ be multivariate normally distributed with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, and every $g_j(\boldsymbol{x})$ is absolutely continuous function defined on $\mathbb{R}^p$ with $\mathrm{Var}\left(g_j(\boldsymbol{\eta})\right) < \infty$, $j = 1, \ldots, p$. Define $\mathbf{g}(\boldsymbol{x}) = \left(g_1(\boldsymbol{x}), \ldots, g_p(\boldsymbol{x})\right)^T$, $a_{ij} = \mathbb{E}\left\{\left(\frac{\partial g_i(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}\right)^T \boldsymbol{\Sigma} \left(\frac{\partial g_j(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}\right)\right\}$, $b_{ij} = \mathrm{Cov}(g_i(\boldsymbol{\eta}), g_j(\boldsymbol{\eta}))$, $\boldsymbol{A} = (a_{ij})_{p\times p}$, $\boldsymbol{B} = (b_{ij})_{p\times p}$. Then matrix $\boldsymbol{A} - \boldsymbol{B}$ is nonnegative definite, i.e.*

$$\mathbb{E}\left\{\left(\frac{\partial \boldsymbol{g}(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}^T}\right) \boldsymbol{\Sigma} \left(\frac{\partial \boldsymbol{g}(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}^T}\right)^T\right\} - \mathrm{Cov}\left(\boldsymbol{g}(\boldsymbol{\eta}), \boldsymbol{g}(\boldsymbol{\eta})\right) \geq 0. \tag{4}$$

Lemma 1. *Suppose that $\boldsymbol{\xi} \sim N_p(\mathbf{0}, \boldsymbol{I}_p)$, where $\boldsymbol{I}_p$ is the p-dimensional identity matrix, $g_j(\boldsymbol{x})$ is the absolutely continuous function defined on $\mathbb{R}^p$ with $\mathrm{Var}\left(g_j(\boldsymbol{\xi})\right) < \infty$, $j = 1, \ldots, p$. Write $\mathbf{g}(\boldsymbol{x}) = \left(g_1(\boldsymbol{x}), \ldots, g_p(\boldsymbol{x})\right)^T$. Then*

$$\mathbb{E}\left\{\left(\frac{\partial \boldsymbol{g}(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}^T}\right) \left(\frac{\partial \boldsymbol{g}(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}^T}\right)^T\right\} - \mathrm{Cov}\left(\boldsymbol{g}(\boldsymbol{\xi}), \boldsymbol{g}(\boldsymbol{\xi})\right) \geq 0. \tag{5}$$

Proof. For the definition and properties of multivariate Hermite polynomials $\{He_{\boldsymbol{r}}(\boldsymbol{\xi}; \boldsymbol{\Sigma})\}_{\boldsymbol{r}=\mathbf{0}}^{\infty}$, please refer to Willink (2005), here $\boldsymbol{r} = (r_1, \ldots, r_p)^T \in \mathbb{N}^p$. Expand $g_j(\boldsymbol{\xi})$ in multivariate orthogonal Hermite polynomials $\left\{He_{\boldsymbol{r}}(\boldsymbol{\xi}; \boldsymbol{I}_p)\right\}_{\boldsymbol{r}=\mathbf{0}}^{\infty}$,

$$g_j(\boldsymbol{\xi}) = \sum_{\boldsymbol{r}=\mathbf{0}}^{\infty} a_{j,\boldsymbol{r}} He_{\boldsymbol{r}}(\boldsymbol{\xi}; \boldsymbol{I}_p), \quad j = 1, \ldots, p,$$

with probability 1. We know that

$$\begin{aligned}
&He_{\boldsymbol{r}}(\boldsymbol{\xi}; \boldsymbol{I}_p) = He_{r_1}(\xi_1) \cdots He_{r_p}(\xi_p), \\
&\mathbb{E}\{\left(He_{r_1}(\xi_1)\right)^2\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} He_{r_1}^2(x) \exp(-x^2/2) \mathrm{d}x = r_1!, \\
&\mathbb{E}\left\{He_{\boldsymbol{r}}(\boldsymbol{\xi}; \boldsymbol{I}_p)\, He_{\boldsymbol{s}}(\boldsymbol{\xi}; \boldsymbol{I}_p)\right\} = \delta_{r_1 s_1} \cdots \delta_{r_p s_p} \boldsymbol{r}!,
\end{aligned}$$

where δ_{mn} is the Kronecker delta function, $\boldsymbol{r}!$ denotes $r_1! \cdots r_p!$. Additionally,

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{x}} He_{\boldsymbol{r}}(\boldsymbol{x}; \boldsymbol{I}_p) &= \frac{\partial}{\partial \boldsymbol{x}} \left(\prod_{j=1}^{p} He_{r_j}(x_j)\right) \\
&= \begin{pmatrix} r_1 & & \\ & \ddots & \\ & & r_p \end{pmatrix} \begin{pmatrix} He_{\boldsymbol{r}-\boldsymbol{e}_1}(\boldsymbol{x}; \boldsymbol{I}_p) \\ \vdots \\ He_{\boldsymbol{r}-\boldsymbol{e}_p}(\boldsymbol{x}; \boldsymbol{I}_p) \end{pmatrix}.
\end{aligned}$$

Here $\boldsymbol{e}_k = (0, \ldots, 1, \ldots, 0)^T$ denotes a p-dimensional unit column vector with kth is 1, for $k = 1, \ldots, p$. Consequently,

$$\frac{\partial g_j(\boldsymbol{x})}{\partial \boldsymbol{x}} = \sum_{\boldsymbol{r}\neq\mathbf{0}}^{\infty} a_{j,\boldsymbol{r}} \begin{pmatrix} r_1 & & \\ & \ddots & \\ & & r_p \end{pmatrix} \begin{pmatrix} He_{\boldsymbol{r}-\boldsymbol{e}_1}(\boldsymbol{x}; \boldsymbol{I}_p) \\ \vdots \\ He_{\boldsymbol{r}-\boldsymbol{e}_p}(\boldsymbol{x}; \boldsymbol{I}_p) \end{pmatrix}.$$

On the one hand,

$$\operatorname{Cov}\left(g_i(\boldsymbol{\xi}), g_j(\boldsymbol{\xi})\right) = \sum_{\boldsymbol{r}\neq\boldsymbol{0}}^{\infty} a_{i,\boldsymbol{r}} a_{j,\boldsymbol{r}} \boldsymbol{r}!. \tag{6}$$

On the other hand,

$$\begin{aligned}\mathbb{E}\left\{\left(\frac{\partial}{\partial\boldsymbol{\xi}^T} g_i(\boldsymbol{\xi})\right)\left(\frac{\partial}{\partial\boldsymbol{\xi}^T} g_j(\boldsymbol{\xi})\right)^T\right\} &= \sum_{\boldsymbol{r}\neq\boldsymbol{0}}^{\infty} a_{i,\boldsymbol{r}} a_{j,\boldsymbol{r}} \left(\sum_{k=1}^{p} r_k^2 (\boldsymbol{r}-\boldsymbol{e}_k)!\right) \\ &= \sum_{\boldsymbol{r}\neq\boldsymbol{0}}^{\infty} a_{i,\boldsymbol{r}} a_{j,\boldsymbol{r}} \boldsymbol{r}! \left(\sum_{k=1}^{p} r_k\right). \end{aligned} \tag{7}$$

From Eqs. (6) and (7), it follows that

$$\mathbb{E}\left\{\left(\frac{\partial}{\partial\boldsymbol{\xi}^T} g_i(\boldsymbol{\xi})\right)\left(\frac{\partial}{\partial\boldsymbol{\xi}^T} g_j(\boldsymbol{\xi})\right)^T\right\} - \operatorname{Cov}\left(g_i(\boldsymbol{\xi}), g_j(\boldsymbol{\xi})\right) = \sum_{\boldsymbol{r}\neq\boldsymbol{0}}^{\infty} a_{i,\boldsymbol{r}} a_{j,\boldsymbol{r}} \boldsymbol{r}! \left(\sum_{k=1}^{p} r_k - 1\right).$$

Evidently, $\sum_{k=1}^{p} r_k - 1 \geq 0$, for every $\boldsymbol{r} \neq \boldsymbol{0}$, $\boldsymbol{r} \in \mathbb{N}^p$. Let $\lambda_{j,\boldsymbol{r}} = \sqrt{\boldsymbol{r}!\left(\sum_{k=1}^{p} r_k - 1\right)}\, a_{j,\boldsymbol{r}}$, $\boldsymbol{\lambda}_j = (\lambda_{j,\boldsymbol{r}_1}, \lambda_{j,\boldsymbol{r}_2}, \ldots)$, for $j = 1, \ldots, p$, $\{\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots\} \subset \mathbb{N}^p \setminus \{\boldsymbol{0}\}$, and write

$$\boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{\lambda}_1 \\ \vdots \\ \boldsymbol{\lambda}_p \end{pmatrix},$$

obviously, the previous identity equals to $\boldsymbol{\lambda}_i (\boldsymbol{\lambda}_j)^T$. Then

$$\mathbb{E}\left\{\left(\frac{\partial \boldsymbol{g}(\boldsymbol{\xi})}{\partial\boldsymbol{\xi}^T}\right)\left(\frac{\partial \boldsymbol{g}(\boldsymbol{\xi})}{\partial\boldsymbol{\xi}^T}\right)^T\right\} - \operatorname{Cov}\left(\boldsymbol{g}(\boldsymbol{\xi}), \boldsymbol{g}(\boldsymbol{\xi})\right) = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^T \geq 0.$$

which completes Lemma 1. □

Proof of Proposition 1. For p-dimensional random vector $\boldsymbol{\eta} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, we express $\boldsymbol{\eta} = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\boldsymbol{\xi}$ and let $\mathbf{G}(\boldsymbol{\xi}) = \boldsymbol{g}(\boldsymbol{\eta}) = \boldsymbol{g}(\boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\boldsymbol{\xi})$, where $\boldsymbol{\xi} \sim N_p(\boldsymbol{0}, \boldsymbol{I}_p)$. Hence

$$\frac{\partial \boldsymbol{g}(\boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\boldsymbol{\xi})}{\partial\boldsymbol{\xi}^T} = \frac{\partial \boldsymbol{g}(\boldsymbol{\eta})}{\partial\boldsymbol{\eta}^T}\boldsymbol{\Sigma}^{1/2} = \frac{\partial \boldsymbol{G}(\boldsymbol{\xi})}{\partial\boldsymbol{\xi}^T}.$$

Applying Lemma 1, it follows immediately that

$$\mathbb{E}\left\{\left(\frac{\partial \boldsymbol{g}(\boldsymbol{\eta})}{\partial\boldsymbol{\eta}^T}\right)\boldsymbol{\Sigma}\left(\frac{\partial \boldsymbol{g}(\boldsymbol{\eta})}{\partial\boldsymbol{\eta}^T}\right)^T\right\} - \operatorname{Cov}\left(\boldsymbol{g}(\boldsymbol{\eta}), \boldsymbol{g}(\boldsymbol{\eta})\right) = \mathbb{E}\left\{\left(\frac{\partial \boldsymbol{G}(\boldsymbol{\xi})}{\partial\boldsymbol{\xi}^T}\right)\left(\frac{\partial \boldsymbol{G}(\boldsymbol{\xi})}{\partial\boldsymbol{\xi}^T}\right)^T\right\} - \operatorname{Cov}\left(\boldsymbol{G}(\boldsymbol{\xi}), \boldsymbol{G}(\boldsymbol{\xi})\right) \geq 0. \quad \square$$

Remark 1. We provided an alternative proof for Eq. (2) in the course of proving Lemma 1. Additionally, we can obtain an extension of Eq. (2) by using differential method directly. Namely, suppose that $\boldsymbol{\eta} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then

$$\operatorname{Var}(g(\boldsymbol{\eta})) \leq \mathbb{E}\left\{\left(\frac{\partial}{\partial\boldsymbol{\eta}^T} g(\boldsymbol{\eta})\right)\boldsymbol{\Sigma}\left(\frac{\partial}{\partial\boldsymbol{\eta}} g(\boldsymbol{\eta})\right)\right\}. \tag{8}$$

References

Chen, L.H.Y., 1982. An inequality for multivariate normal distribution. Journal of Multivariate Analysis 12, 306–315.
Chernoff, H., 1981. A note on an inequality involving the normal distribution. Annals of Probability 9, 533–535.
Olkin, I., Shepp, L., 2005. A matrix variance inequality. Journal of Statistical Planning and Inference 130, 351–358.
Willink, R., 2005. Normal moments and Hermite polynomials. Statistics & Probability Letters 73, 271–275.