

# Independence and maximal volume of $d$ -dimensional random convex hull

Won Son<sup>a</sup>, Seongoh Park<sup>a</sup>, Johan Lim<sup>1,a</sup>

<sup>a</sup>Department of Statistics, Seoul National University, Korea

---

## Abstract

In this paper, we study the maximal property of the volume of the convex hull of  $d$ -dimensional independent random vectors. We show that the volume of the random convex hull from a multivariate location-scale family indexed by  $\Sigma$  is stochastically maximized in simple stochastic order when  $\Sigma$  is diagonal. The claim can be applied to a broad class of multivariate distributions that include skewed/unskewed multivariate  $t$ -distributions. We numerically investigate the proven stochastic relationship between the dependent and independent random convex hulls with the Gaussian random convex hull. The numerical results confirm our theoretical findings and the maximal property of the volume of the independent random convex hull.

**Keywords:** convex hull, independence, multivariate location-scale family, simple stochastic order, stochastic geometry

---

## 1. Introduction

Random convex hull, the convex hull of independent and identically distributed random points, have been studied for decades after the seminal work by Rényi and Sulanke (1963, 1964). In particular, researchers in stochastic geometry focus on the functionals of random convex hull (where the two most important functionals are the volume and number of faces) and investigate their finite and asymptotic properties.

The random convex hull is also employed in many multivariate statistical procedures. For example, Barnett (1976) defines an ordering of the multivariate data based on the notion of convex hull peeling depth. In Cook (1979), the random convex hull generated by the data points is used to identify the influential observations in linear regression. It is also used to find an optimal classifier in machine learning literature (Fawcett and Niculescu-Mizil, 2007; Lim and Won, 2012; Son *et al.*, 2015). Ng *et al.* (2014) recently developed an efficient algorithm to simulate a random convex hull on a plane, and applied it to testing the independence of a  $d$ -dimensional random vector. They consider the use of the area (volume) of the convex hull for testing independence. However, no statistical justification is given for the use of the volume of the convex hull.

In this paper, we are interested in the maximal property of the volume of the convex hull of  $d$ -dimensional random vectors under independence. This paper argues that “the volume of the random convex hull of  $d$ -dimensional vectors  $\mathbf{X}_i = (X_i^1, X_i^2, \dots, X_i^d)^T$ ,  $i = 1, 2, \dots, n$ , is maximized in simple stochastic order (defined later) when  $X_i^1, X_i^2, \dots, X_i^d$  are independent (or uncorrelated) to each other.”

---

<sup>1</sup> Corresponding author: Department of Statistics, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Korea. E-mail: [johanlim@snu.ac.kr](mailto:johanlim@snu.ac.kr)

We show the claim is true for a broad class of multivariate distributions indexed by covariance matrix  $\Sigma$  which includes skewed/unskewed multivariate  $t$ -distributions.

The remainder of the paper is organized as follows. In Section 2, we introduce general notations to be used in the paper. In addition, we show the invariance of the volume of the random convex hull with respect to rotational and axis-scalable transformations. We also introduce multivariate location-scale (MLS) family indexed by  $\Sigma$ , denoted by  $\text{MLS}(\Sigma)$ , which is invariant to the two transformations above. In Section 3, we prove the maximal property of the random convex hull under independence when the random vectors are from a distribution in  $\text{MLS}(\Sigma)$ . We then discuss the Gaussian random convex hull as an illustrative example and provide numerical illustrations of the results in the section. Finally, in Section 4, we conclude the paper with a discussion on the extension of the results to serially correlated data.

## 2. Preliminaries

### 2.1. Notation

Suppose we consider  $n$  data points in  $\mathcal{R}^d$  with  $d \geq 1$ . The convex hull of the data points  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$  is defined as

$$\text{chull}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) := \left\{ \sum_{i=1}^n \alpha_i \mathbf{x}_i \mid \sum_{i=1}^n \alpha_i = 1, \alpha_i \geq 0, i = 1, 2, \dots, n \right\}. \quad (2.1)$$

A vertex of a convex set  $S \subset \mathcal{R}^d$  is a point  $x$  which cannot be written as a convex combination of the points in  $S \setminus \{x\}$ . The vertexes of the convex hull  $\text{chull}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$  are elements of the set  $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$  and forms the vertexes of the convex hull denoted by  $V = \{v_1, v_2, \dots, v_K\}$  (for some  $K \leq n$ ). Finally, we find that  $\text{chull}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$  equals to the polytope with vertexes  $v_1, v_2, \dots, v_K$  which is denoted as  $\text{pt}(v_1, v_2, \dots, v_K)$ .

### 2.2. Multivariate location-scale family indexed by $\Sigma$

In this section, we introduce the MLS family that we assume for the distribution of random vectors in the paper.

The MLS family is one of the important parametric families and many important distributions are included in location-scale family. The family  $\{P_{(\mu, \Sigma)} : \mu \in \mathcal{R}^d, \Sigma \in \mathcal{M}_d\}$  is defined to be a location-scale family on  $\mathcal{R}^d$  if

$$P_{(\mu, \Sigma)}(\mathbf{B}) = G\left(\Sigma^{-\frac{1}{2}}(\mathbf{B} - \mu)\right), \quad \mathbf{B} \in \mathcal{B}^d, \quad (2.2)$$

where  $G(\cdot)$  is an arbitrary given probability measure on the  $d$ -dimensional Borel field,  $\Sigma^{-1/2}(\mathbf{B} - \mu) = \{\Sigma^{-1/2}(\mathbf{x} - \mu) : \mathbf{x} \in \mathbf{B} \subset \mathcal{R}^d\}$ , and  $\mathcal{M}_d$  is a collection of  $d \times d$  symmetric positive definite matrices (Shao, 2003).

Some examples of the MLS family are as follows. Elliptically symmetric distributions (or simply elliptical distributions) belong to a location-scale family (Ollila *et al.*, 2003). A  $d$ -dimensional random vector  $\mathbf{X}$  is said to have an elliptical distribution, denoted by  $\mathbf{X} \sim EC_d(\mu, \Sigma, \xi)$ , if it has a stochastic representation

$$\mathbf{X} = \mu + \xi \Sigma^{\frac{1}{2}} \mathbf{U}, \quad (2.3)$$

where  $\mathbf{U}$  is a random vector uniformly distributed on the unit sphere in  $\mathcal{R}^d$ ,  $\xi \geq 0$  is a scalar random variable independent of  $\mathbf{U}$ , and  $\Sigma^{1/2}$  is a deterministic symmetric matrix such that  $\Sigma^{1/2}(\Sigma^{1/2})^T = \Sigma$ . There are many subclasses of elliptical distributions. Multivariate normal distributions, multivariate  $t$ -distributions, logistic distributions and multivariate Cauchy distributions are well-known examples of them (Fang *et al.*, 1990). While the underlying population distributions are free of the assumption of symmetry, some distributions are also included in the MLS family. Multivariate Pareto of the second type (Asimit *et al.*, 2010), a generalized multivariate gamma distributions (Carpenter and Diawara, 2007), and multivariate skew normal distributions and multivariate skew  $t$ -distributions (Azzalini and Capitanio, 2003) are examples of asymmetric distributions which can be classified as the MLS family. Some further extensions can also be found from Zhao and Kim (2016).

In this paper, we assume  $\boldsymbol{\mu} = \mathbf{0}$  without loss of generality, and index the MLS family with the covariance matrix  $\Sigma$ . In sequel, we simply write the MLS family indexed by  $\Sigma$  as  $\text{MLS}(\Sigma)$  omitting the distribution notation of  $\xi$  of (2.3). Further, we let  $\lambda_j$ s for  $j = 1, 2, \dots, d$  be the eigenvalues of the covariance matrix  $\Sigma$ .

### 2.3. Vertex invariance to rotation and scale transformation

Let us consider  $n$  independent random samples  $\mathbf{X}_i = (X_i^1, X_i^2, \dots, X_i^d)^T$ ,  $i = 1, 2, \dots, n$ , which are identically from the  $d$ -dimensional multivariate distribution with mean  $\mathbf{0}$  and covariance matrix  $\Sigma$ . Let  $v.\text{chull}(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n)$  be the volume of  $\text{chull}(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n)$ .

The following lemma shows that the vertex set of  $\text{chull}(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n)$  is invariant to both rotation and scale (according to the axis) transformations of  $\{\mathbf{X}_i, i = 1, 2, \dots, n\}$ .

**Lemma 1.** *Let  $P$  be a  $d$ -dimensional orthonormal matrix and  $Q$  be the diagonal matrix with diagonal elements  $q_1, q_2, \dots, q_d$ . If  $V = \{v_1, v_2, \dots, v_K\}$  is the set of the vertexes of  $\text{chull}(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n)$ , then (i) the vertexes of  $\text{chull}(P\mathbf{X}_1, P\mathbf{X}_2, \dots, P\mathbf{X}_n)$  is  $PV = \{Pv_1, Pv_2, \dots, Pv_K\}$  and (ii) the vertex set of  $\text{chull}(Q\mathbf{X}_1, Q\mathbf{X}_2, \dots, Q\mathbf{X}_n)$  is  $QV = \{Qv_1, Qv_2, \dots, Qv_K\}$ .*

**Proof:** The proof of (ii) is very similar to that of (i). Thus, we only prove (i) at here.

We first show that  $Pv_k$  is a vertex of  $\text{chull}(P\mathbf{X}_1, P\mathbf{X}_2, \dots, P\mathbf{X}_n)$ . Suppose, without loss of generality,  $Pv_k = P\mathbf{X}_1$  and assume it is not a vertex. Then, it can be written as a convex combination of  $P\mathbf{X}_2, P\mathbf{X}_3, \dots, P\mathbf{X}_n$  as

$$P\mathbf{X}_1 = \sum_{i=2}^n \alpha_i P\mathbf{X}_i, \quad \sum_{i=2}^n \alpha_i = 1, \quad \alpha_i \geq 0.$$

By multiplying  $P^T$  in both sides of the above, we have

$$v_k = \mathbf{X}_1 = \sum_{i=2}^n \alpha_i \mathbf{X}_i, \quad \sum_{i=2}^n \alpha_i = 1, \quad \alpha_i \geq 0,$$

which implies  $v_k$  is not a vertex of  $\text{chull}(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n)$ . This introduces a contradiction.

We now show that any point in  $\text{chull}(P\mathbf{X}_1, P\mathbf{X}_2, \dots, P\mathbf{X}_n)$  has a convex representation of  $PV = \{Pv_1, Pv_2, \dots, Pv_K\}$ . Suppose  $\mathbf{y}$  is a vertex of the convex hull but not in  $PV$ . Since it is within a convex hull, it can be written as

$$\mathbf{y} = \sum_{i=1}^n \alpha_i P\mathbf{X}_i, \quad \alpha_i \geq 0, \quad \sum_{i=1}^n \alpha_i = 1.$$

Again, by multiplying  $P^T$  both sides,

$$P^T \mathbf{y} = \sum_{i=1}^n \alpha_i \mathbf{X}_i = \sum_{k=1}^K \beta_k \mathbf{v}_k, \quad \alpha_i \geq 0, \quad \sum_{i=1}^n \alpha_i = 1 \quad \text{and} \quad \beta_k \geq 0, \quad \sum_{k=1}^K \beta_k = 1,$$

where the second equation is from that  $V = \{v_1, v_2, \dots, v_K\}$  is the vertex set of  $\text{chull}(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n)$ . In the above, the last equation

$$P^T \mathbf{y} = \sum_{k=1}^K \beta_k \mathbf{v}_k, \quad \beta_k \geq 0, \quad \sum_{k=1}^K \beta_k = 1,$$

is equivalent to

$$\mathbf{y} = \sum_{k=1}^K \beta_k P \mathbf{v}_k, \quad \beta_k \geq 0, \quad \sum_{k=1}^K \beta_k = 1,$$

which contradicts to that  $\mathbf{y}$  is a vertex not in PV. □

### 3. Maximal volume of random convex hull of samples from $\text{MLS}(\Sigma)$

#### 3.1. Main result

We now present our main results of the paper which show the volume of the convex hull of a  $\text{MLS}(\Sigma)$  is stochastically maximized when true covariance matrix  $\Sigma$  is diagonal, equivalently,  $X^1, X^2, \dots, X^d$  are uncorrelated to each other. If the data are from a multivariate normal distribution, the volume is maximized when the  $d$ -variables are independent. The stochastic order between two variables  $X$  and  $Z$  is defined as:  $Z$  is (simply) stochastically larger than  $X$ , denoted by  $X \leq_{\text{st}} Z$  if and only if  $P(Z \leq a) \leq P(X \leq a)$  for every  $a \in \mathcal{R}$ .

**Theorem 1.** *Suppose  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$  are independently from a distribution from a  $\text{MLS}$  with covariance matrix  $\Sigma$ ,  $\text{MLS}(\Sigma)$ , and  $\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_n$  are independently from  $\text{MLS}(\text{diag}(\Sigma))$ , where  $\text{diag}(\Sigma)$  is the diagonal matrix whose diagonal elements are same with those of  $\Sigma$ . Then, for every  $n$  and a positive definite covariance matrix  $\Sigma$ , we have*

$$\text{v.chull}(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n) \leq_{\text{st}} \text{v.chull}(\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_n). \quad (3.1)$$

**Proof:** Suppose  $V = \{v_1, v_2, \dots, v_K\}$  is the set of vertexes of  $\text{chull}(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n)$  and let the singular value decomposition of  $\Sigma$  be  $P^T \Lambda P$  with an orthonormal matrix  $P$  and  $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_d)$ .

$$\begin{aligned} \text{v.chull}(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n) &= \iint \mathbf{I}(\mathbf{x} \in \text{chull}(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n)) d\mathbf{x} \\ &= \iint \mathbf{I}(\mathbf{y} \in \text{chull}(P\mathbf{X}_1, P\mathbf{X}_2, \dots, P\mathbf{X}_n)) \cdot \det(P^T) d\mathbf{y} \\ &= \iint \mathbf{I}(\mathbf{y} \in \text{pt}(Pv_1, Pv_2, \dots, Pv_K)) \cdot 1 d\mathbf{y}, \end{aligned} \quad (3.2)$$

which is from the rotation transformation  $\mathbf{y} = P^T \mathbf{x}$  (equivalently,  $P\mathbf{y} = \mathbf{x}$ ) and the invariance from Lemma 1. In (3.2),  $\text{pt}(Pv_1, Pv_2, \dots, Pv_K)$  equals to  $\text{chull}(P\mathbf{X}_1, P\mathbf{X}_2, \dots, P\mathbf{X}_n)$ , where  $\mathbf{Y}_i = P\mathbf{X}_i$  are

independently and identically distributed (iid) as  $\text{MLS}(\Lambda)$ . Thus,

$$(3.2) = \iint \mathbf{I}(\mathbf{y} \in \text{chull}(\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_n)) d\mathbf{y} \\ = \text{v.chull}(\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_n).$$

We now show that  $\text{v.chull}(\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_n)$  is stochastically smaller than  $\text{v.chull}(\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_n)$ . This is simply by applying the axis-wise scale transformation to both  $\text{chull}(\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_n)$  and  $\text{chull}(\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_n)$ . First, for  $\text{chull}(\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_n)$ , we consider the transformation matrix  $\mathbf{Q}_Y = \Lambda^{-1/2}$  and have

$$\begin{aligned} \text{v.chull}(\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_n) &= \iint \mathbf{I}(\mathbf{y} \in \text{chull}(\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_n)) d\mathbf{y} \\ &= \iint \mathbf{I}(\mathbf{u} \in \text{chull}(\mathbf{Q}_Y \mathbf{Y}_1, \mathbf{Q}_Y \mathbf{Y}_2, \dots, \mathbf{Q}_Y \mathbf{Y}_n)) \cdot \det(\mathbf{Q}_Y^{-1}) d\mathbf{u} \\ &= \left( \prod_{j=1}^d \lambda_j \right)^{\frac{1}{2}} \text{v.chull}(\mathbf{U}_{1,y}, \mathbf{U}_{2,y}, \dots, \mathbf{U}_{n,y}), \end{aligned} \quad (3.3)$$

where  $\mathbf{U}_{i,y}$  are iid from  $\text{MLS}(\mathbf{I}_d)$  with  $\mathbf{I}_d$  is the  $d$ -dimensional identity matrix. In showing (3.3), we use the invariance property proven in (ii) of Lemma 1 as in (3.2). The similar steps are applied to  $\text{chull}(\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_n)$  with the scale transformation  $\mathbf{Q}_Z = \{\text{diag}(\Sigma)\}^{-1/2}$ , and show that

$$\text{v.chull}(\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_n) = \left( \prod_{j=1}^d \sigma_{jj} \right)^{\frac{1}{2}} \text{v.chull}(\mathbf{U}_{1,z}, \mathbf{U}_{2,z}, \dots, \mathbf{U}_{n,z}), \quad (3.4)$$

where  $\sigma_{jj}$  is the  $j^{\text{th}}$  diagonal element of  $\Sigma$  and  $\mathbf{U}_{i,z}$  are iid from  $\text{MLS}(\mathbf{I}_d)$ .

Finally, we conclude the proof using the Hadamard's inequality (Cover and Gamal, 1983), which tells, for the covariance matrix  $\Sigma$ ,

$$\left( \prod_{j=1}^d \lambda_j \right) = \det(\Sigma) \leq \left( \prod_{j=1}^d \sigma_{jj} \right). \quad (3.5)$$

□

### 3.2. Gaussian random convex hull

The Gaussian random convex hull, the random convex hull for  $d$ -dimensional Gaussian random vectors, is studied by many researchers. Suppose  $\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_n$  are iid from the  $d$ -dimensional standard normal distribution. It is shown by Affentranger (1991) that

$$\begin{aligned} \mathbb{E}\{\text{v.chull}(\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_n)\} &= \kappa_d (2 \log n)^{\frac{d}{2}} (1 + o(1)) \\ &= \left\{ \frac{\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2} + 1)} \right\} (2 \log n)^{\frac{d}{2}} (1 + o(1)), \end{aligned} \quad (3.6)$$

where  $\kappa_d$  is the volume of the  $d$ -dimensional unit ball. It is also shown by Hug and Reitzner (2005) that, for  $d \geq 1$ ,

$$\text{var}\{\text{v.chull}(\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_n)\} = O\left((\log n)^{\frac{d-3}{2}}\right). \quad (3.7)$$

However, Hug (2013) points out that the explicit finite sample distribution function is still unknown. Instead, some of its asymptotic are known. For example, Bárány and Vu (2007) prove the central limit theorem for the volume of the Gaussian random convex hull.

Theorem 1 in Section 3 further shows that, for the general  $\Sigma$ , the volume has a constant multiplicative factor  $(\prod_{j=1}^d \lambda_j)^{1/2}$  which is canceled out from numerator and denominator of its standardized form. The standardized statistic

$$\mathbf{t}_{\text{v.chull}} = \frac{\text{v.chull}(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n) - E(\text{v.chull}(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n))}{\sqrt{\text{var}(\text{v.chull}(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n))}}$$

is invariant to the scale transformation and has the same distribution with

$$\mathbf{t}_{\text{v.chull}} = \frac{\text{v.chull}(\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_n) - E(\text{v.chull}(\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_n))}{\sqrt{\text{var}(\text{v.chull}(\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_n))}},$$

where  $E(\text{v.chull}(\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_n))$  and  $\text{var}(\text{v.chull}(\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_n))$  are those in (3.6) and (3.7), respectively.

### 3.3. Numerical illustration

We now numerically illustrates the findings in the previous subsection. The identity (3.3) tells that

$$\text{v.chull}(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n) = \left( \prod_{j=1}^d \lambda_j \right)^{\frac{1}{2}} \text{v.chull}(\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_n),$$

where  $\mathbf{X}_i$ s are from the multivariate normal distribution with mean 0 and variance  $\Sigma$ ,  $\mathbf{Z}_i$ s are iid  $d$ -dimensional standard normal vector, and  $\lambda_j$ ,  $j = 1, 2, \dots, d$ , are the eigenvalues of  $\Sigma$ . Thus,

$$\Delta = \log \text{v.chull}(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n) - \frac{1}{2} \log \left( \prod_{j=1}^d \lambda_j \right)$$

has the same distribution with

$$\log \text{v.chull}(\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_n),$$

and is invariant to the choice of  $\rho$  for fixed  $d$  and  $n$ . We numerically investigate this identity.

We generate samples from  $d$ -dimensional multivariate  $t$ -distribution with degrees of freedom  $\nu = 3, 5, 10, \infty$  ( $\infty$  corresponds to the multivariate normal distribution),  $\mu = \mathbf{0}_d$ , and two types of  $\Sigma$  defined below with the dimension  $d = 2, 4$ . Two covariance matrices we consider are: (i) the compound symmetry (CS) matrix, notated as  $\text{CS}(\rho)$ , is defined as

$$\Sigma_{\text{cs}} = (1 - \rho)\mathbf{I}_d + \rho\mathbf{1}_d\mathbf{1}_d^T$$

and its log-determinant is  $\log \det(\text{CS}(\rho)) = \log\{1 + (d - 1)\rho\} + (d - 1)\log(1 - \rho)$ , (ii) the first order auto-regressive (AR) model, notated as  $\text{AR1}(\rho)$ , is defined as

$$\Sigma_{\text{ar}} = (\sigma_{ij} = \rho^{|i-j|}, i, j = 1, 2, \dots, (d - 1)),$$

and its log-determinant is  $\log \det(\text{AR1}(\rho)) = (d - 1)\log(1 - \rho^2)$ . The sample size is set as  $n = 50, 100, 300, 500$ . We simulate  $B = 1000$  data sets and, in each data, we compute the area of the

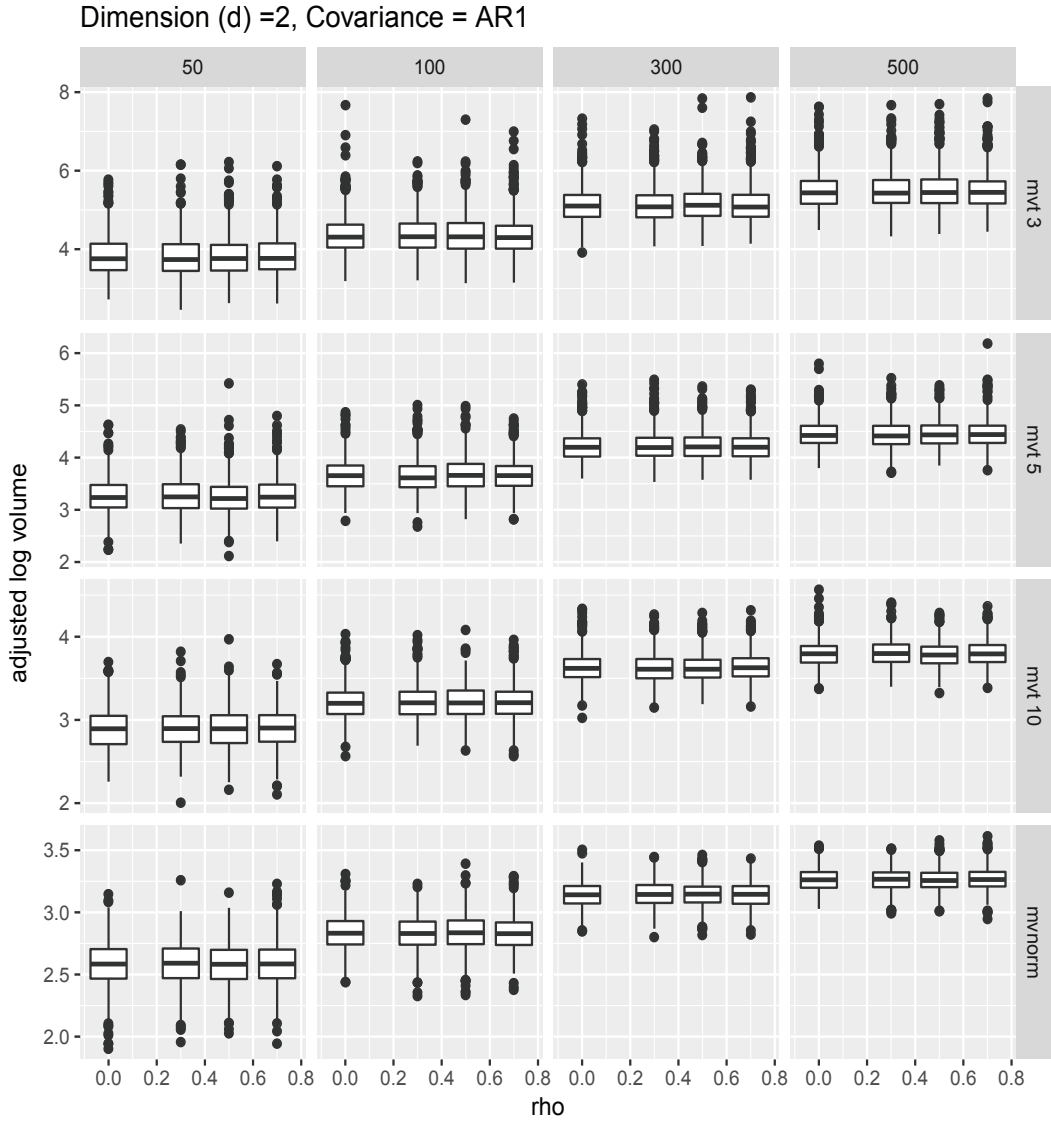


Figure 1:  $d = 2$  and  $\Sigma = \Sigma_{ar}$ : Box plots of  $\Delta = \log v.\text{chull}(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n) - \sum_{j=1}^d \log \lambda_j$  for different choices of  $\rho$ ,  $v$ , and  $n$ . AR = auto-regressive.

convex hull. To compute the area of the convex hull, we use the “convhulln” function in the R-package “geometry”, which implements the Quickhull algorithm (Barber *et al.*, 1996).

The box plots of  $\Delta = \log v.\text{chull}(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n) - (1/2) \log(\prod_{j=1}^d \lambda_j)$  versus  $\rho$  for fixed  $v, n$  are presented in Figures 1–4 for each combination of  $d = 2, 4$  and  $\Sigma = \Sigma_{cs}, \Sigma_{ar}$ . The figures show that the distribution of  $\Delta$  is invariant to the choice of  $\rho$  for given  $v$  and  $n$  in every combination of  $d = 2, 4$  and  $\Sigma = \Sigma_{cs}, \Sigma_{ar}$ . In each figure, the four box plots in each panel has the same distribution as the distribution of  $\log v.\text{chull}(\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_n)$ . The mean of  $\log v.\text{chull}(\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_n)$  therefore varies

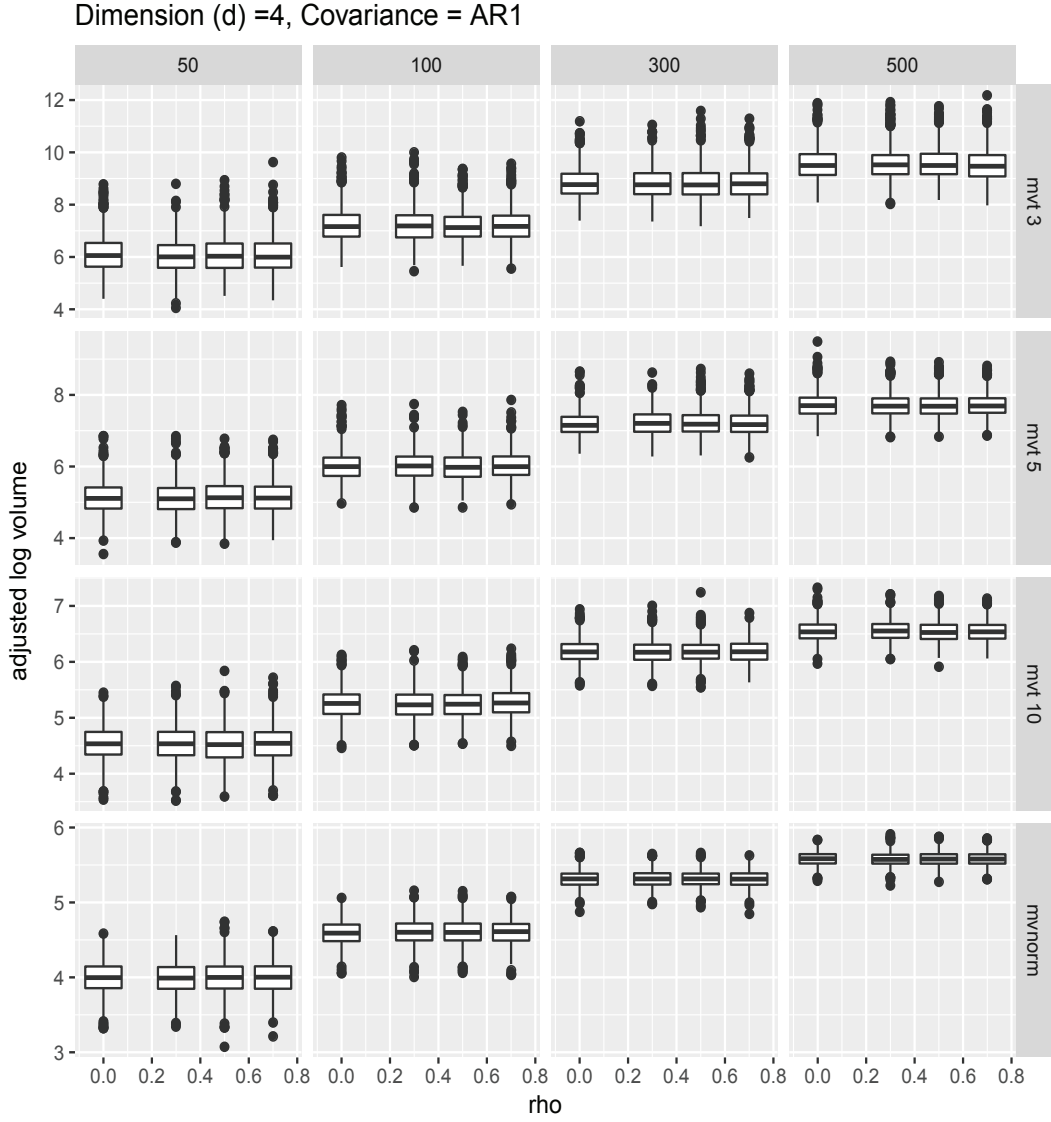


Figure 2:  $d = 4$  and  $\Sigma = \Sigma_{ar}$ : Box plots of  $\Delta = \log v.\text{chull}(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n) - \sum_{j=1}^d \log \lambda_j$  for different choices of  $\rho, \nu$  and  $n$ . AR = auto-regressive.

according to the changes of  $\nu$  and  $n$ ; it tends to increase as either  $\nu$  decreases or  $n$  increases.

#### 4. Conclusion

In this paper, the maximal property of the volume of the convex hull of  $d$ -dimensional independent random vectors is investigated. In stochastic sense, the volume of the convex hull is maximized when the covariance matrix  $\Sigma$  of the underlying probability distribution is diagonal. This results is true for the distribution from the multivariate location-scale family that includes skewed/unskewed



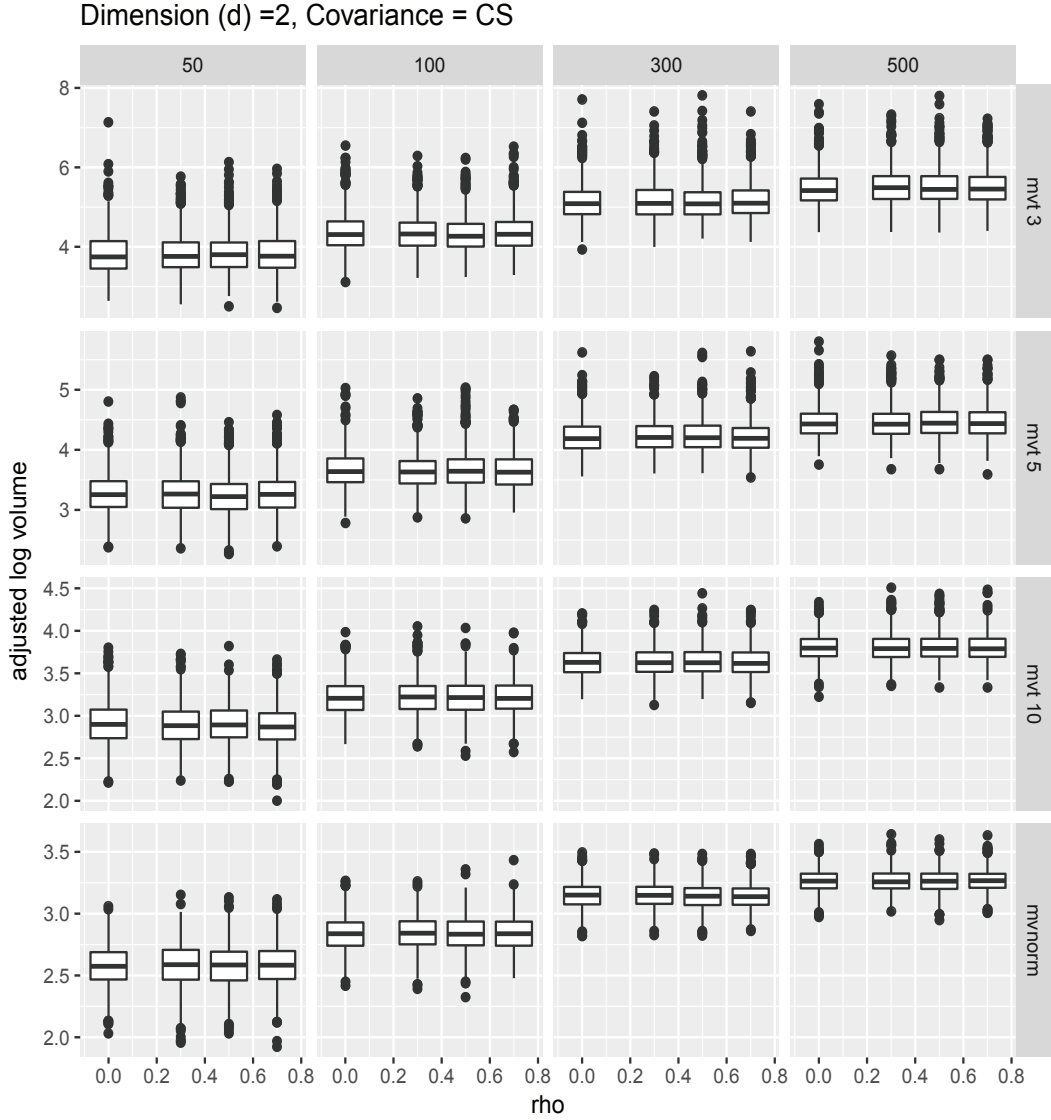


Figure 3:  $d = 2$  and  $\Sigma = \Sigma_{cs}$ : Box plots of  $\Delta = \log v.\text{chull}(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n) - \sum_{j=1}^d \log \lambda_j$  for different choices of  $\rho, v$  and  $n$ . CS = compound symmetry.

multivariate  $t$ -distribution and elliptical distribution. Thus, the volume of the convex hull can be used for testing the independence of a  $d$ -dimensional vector as in Ng *et al.* (2014).

Possible future research direction is to extend this conclusion to the random convex hull from dependent samples including time series data. In time series data, the lagged plot, the plot  $(y_t, y_{t-1}, \dots, y_{t-d+1}), \dots, (y_{t-d+1}, y_{t-d}, \dots, y_{t-2d})$ , plays an important role in exploring the data. In particular, the convex hull of the data points in the lagged plot is a key tool to find outliers and influential points. However, its theoretical property is rarely understood and further study on it is demanded.

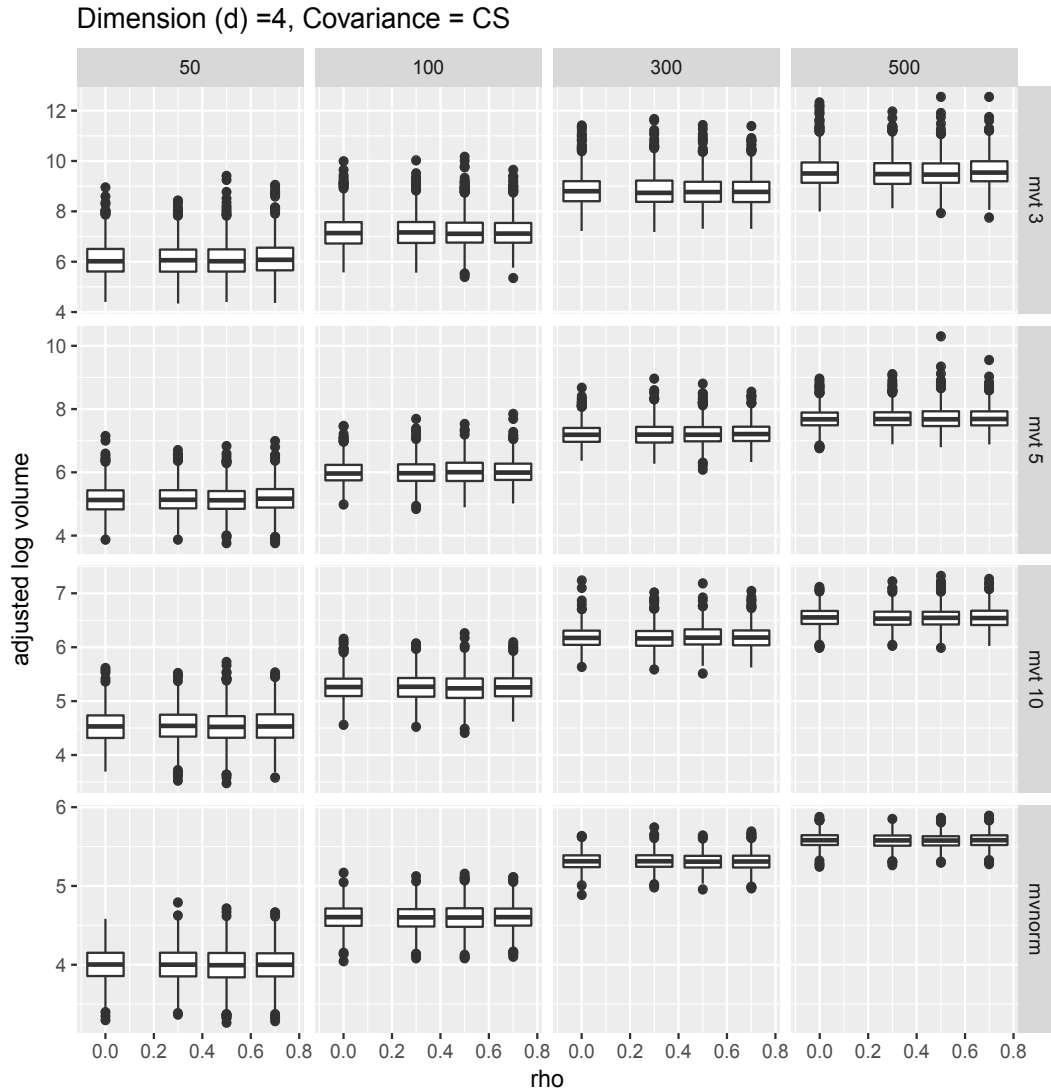


Figure 4:  $d = 4$  and  $\Sigma = \Sigma_{cs}$ : Box plots of  $\Delta = \log v.\text{chull}(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n) - \sum_{j=1}^d \log \lambda_j$  for different choices of  $\rho, v$  and  $n$ . CS = compound symmetry.

## References

- Affentranger F (1991). The convex hull of random points with spherically symmetric distributions, *Rendiconti del Seminario Matematico Università e Politecnico di Torino*, **49**, 359–383.
- Asimit AV, Furman E, and Vernic R (2010). On a multivariate Pareto distribution, *Insurance: Mathematics and Economics*, **46**, 308–316.
- Azzalini A and Capitanio A (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew  $t$ -distribution, *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, **65**, 367–389.

- Bárány I and Vu V (2007). Central limit theorems for Gaussian polytopes, *The Annals of Probability*, **35**, 1593–1621.
- Barber CB, Dobkin DP, and Huhdanpaa H (1996). The Quickhull algorithm for convex hulls, *ACM Transactions on Mathematical Software (TOMS)*, **22**, 469–483.
- Barnett V (1976). The ordering of multivariate data, *Journal of the Royal Statistical Society. Series A (General)*, **139**, 318–355.
- Carpenter M and Diawara N (2007). A multivariate gamma distribution and its characterizations, *American Journal of Mathematical and Management Sciences*, **27**, 499–507.
- Cook RD (1979). Influential observations in linear regression, *Journal of the American Statistical Association*, **74**, 169–174.
- Cover T and Gamal AE (1983). An information - theoretic proof of Hadamard's inequality (Corresp.), *IEEE Transactions on Information Theory*, **29**, 930–931.
- Fang KT, Kotz S, and Ng KW (1990). *Symmetric Multivariate and Related Distributions*, MA Springer US, Boston.
- Fawcett T and Niculescu-Mizil A (2007). PAV and the ROC convex hull, *Machine Learning*, **68**, 97–106.
- Hug D (2013). Random polytopes. In Spodarev E (eds) *Stochastic Geometry, Spatial Statistics and Random Fields, Lecture Notes in Mathematics*, (Vol. 2068, pp. 205–238). Springer-Verlag, Berlin-Heidelberg.
- Hug D and Reitzner M (2005). Gaussian polytopes: variances and limit theorems, *Advances in Applied Probability*, **37**, 297–320.
- Lim J and Won JH (2012). ROC convex hull and nonparametric maximum likelihood estimation, *Machine Learning*, **88**, 433–444.
- Ng CT, Lim J, Lee KE, Yu D, and Choi S (2014). A fast algorithm to sample the number of vertexes and the area of the random convex hull on the unit square, *Computational Statistics*, **29**, 1187–1205.
- Ollila E, Oja H, and Croux C (2003). The affine equivariant sign covariance matrix: asymptotic behavior and efficiencies, *Journal of Multivariate Analysis*, **87**, 328–355.
- Rényi A and Sulanke R (1963). Über die konvexe Hülle von  $n$  zufällig gewählten Punkten, *Zeitschrift Für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **2**, 75–84.
- Rényi A and Sulanke R (1964). Über die konvexe Hülle von  $n$  zufällig gewählten Punkten, *Zeitschrift Für Wahrscheinlichkeitstheorie Und Verwandte Gebiete*, **3**, 138–147.
- Shao J (2003). *Mathematical Statistics* (2nd ed), Springer, New York.
- Son W, Ng CT, and Lim J (2015). A new integral representation of the coverage probability of a random convex hull, *Communications of Statistical Applications and Methods*, **22**, 69–80.
- Zhao J and Kim HM (2016). Power  $t$  distribution, *Communications for Statistical Applications and Methods*, **23**, 321–334.