Statistics and Probability Letters xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

A note on weak convergence of general halfspace depth trimmed means

Jin Wang

Department of Mathematics and Statistics, Northern Arizona University, Flagstaff, AZ 86011-5717, USA

HIGHLIGHTS

- We establish the weak convergence of general halfspace depth trimmed means, which extends the result of Massé (2009) for dimensions one and two to any dimension.
- The asymptotic distribution of the Donoho (1982) halfspace depth trimmed mean is obtained as a special case and concretized for elliptically symmetric distributions.

ARTICLE INFO

Article history:
Received 9 March 2018
Received in revised form 28 June 2018
Accepted 3 July 2018
Available online xxxx

MSC:
primary 62G05
secondary 62H05

Keywords:
Halfspace depth
Multivariate trimmed mean
Weak convergence
Multivariate analysis

ABSTRACT

In this note, we restudy the general halfspace depth trimmed means and establish the weak convergence of their sample versions, which extends the result of Massé (2009) for dimensions one and two to any dimension. The asymptotic distribution of the Donoho (1982) halfspace depth trimmed mean is obtained as a special case and concretized for elliptically symmetric distributions.

1. Introduction

In the univariate case, the trimmed mean is a very popular location estimator. Both mean and median can be regarded as a special case of the trimmed mean. A trimmed mean can be not only efficient but also robust, and one can balance efficiency and robustness by trimming percentage in practical applications. Donoho (1982) extended the univariate trimmed mean to the multivariate case via the halfspace depth function.

Generally, a depth function $D_F(\boldsymbol{x})$ is a nonnegative real-valued mapping which provides a distribution-based center-outward ordering of points $\boldsymbol{x}$ in $\mathbb{R}^d$. Given a depth function $D_F(\boldsymbol{x})$, the $\alpha(\geq 0)$ *depth trimmed region* (or α *depth inner region*) is defined as $T_{D_F}(\alpha) = \{\boldsymbol{x} \in \mathbb{R}^d : D_F(\boldsymbol{x}) \geq \alpha\}$ and its boundary $\partial T_{D_F}(\alpha)$ is called the α *depth contour*. Furthermore an α *depth trimmed mean* can be defined as

$$L(F) = \int_{T_{D_F}(\alpha)} \boldsymbol{x} w(D_F(\boldsymbol{x})) dF(\boldsymbol{x}) \Big/ \int_{T_{D_F}(\alpha)} w(D_F(\boldsymbol{x})) dF(\boldsymbol{x}),$$

E-mail address: jin.wang@nau.edu.

https://doi.org/10.1016/j.spl.2018.07.005
0167-7152/

where $w(\cdot)$ is a suitable weight function on $[0, \infty)$ such that $L(F)$ is well defined. Given a sample $\boldsymbol{X}^n = \{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n\}$, the sample version $L(F_n)$ of $L(F)$ is obtained by replacing F with F_n, the empirical distribution function of $\boldsymbol{X}^n$. To study the weak convergence of $L(F_n)$, $w(\cdot)$ is typically assumed to be continuously differentiable on $[0, \infty)$. This general depth trimmed mean has attracted many statisticians' interest. Dümbgen (1992) studied the simplicial depth trimmed mean, Massé (2004) investigated the halfspace depth trimmed mean, and Zuo et al. (2004) studied the projection depth trimmed mean. However, when studying the weak convergence of their sample versions, all of them invoked the condition that $w(t) = 0$ for $t \in [0, \alpha)$, which along with the continuous differentiability assumption excludes some important cases such as the unweighted depth trimmed means. The important fact is that if we can release the condition and thus work with a larger class of weight functions, we can improve efficiency and robustness of the depth trimmed means. To my knowledge, Zuo (2006) first treated the projection depth trimmed mean with the condition released. While the projection depth-based trimming is a very important trimming procedure, it is quite different from the trimming based on the halfspace depth, which is a natural extension of the usual univariate trimming. Thus Massé (2009) further studied the halfspace depth trimmed means. However, his weak convergence result for the general case (Theorem 3.3 (a)) was still restricted to dimensions one and two. To make the theory complete, we here restudy the general halfspace depth trimmed means, especially weak convergence of their sample versions. Basic properties of the trimmed means are summarized in Section 2. The weak convergence results are established in Section 3 and detailed proofs are given in the Appendix.

Throughout this paper, we use uppercase letters to denote distribution functions and their lowercase counterparts to denote density functions if they exist. For example, we denote by $F_{\boldsymbol{X}}$ and $f_{\boldsymbol{X}}$ the *cdf* and density of a random vector $\boldsymbol{X}$ in $\mathbb{R}^d$. When X is a random variable, the quantile function of X is denoted by F_X^{-1}. Without confusion, we will omit the subscript. The indicator function of a set A is denoted by $I_A(\cdot)$ and $S^{d-1} = \{\boldsymbol{u} \in \mathbb{R}^d : \|\boldsymbol{u}\| = 1\}$.

2. Basic properties

The halfspace depth is first introduced by Tukey (1975) and is defined as

$$HD_F(\boldsymbol{x}) = \inf\{P(H) : \boldsymbol{x} \in H \in \mathcal{H}\}, \boldsymbol{x} \in \mathbb{R}^d,$$

where $\mathcal{H} = \{\text{all closed halfspaces}\}$. It is a leading depth function and has been very popular. Its important properties were studied by Donoho (1982), and Donoho and Gasko (1992). Romanazzi (2001) derived the influence function of $HD_F(\boldsymbol{y})$. The asymptotic distribution of its sample version $HD_{F_n}(\boldsymbol{y})$ was established by Massé (2004). Nolan (1992) studied the asymptotics of the sample halfspace depth trimmed regions.

With $T_{HD_F}(\alpha) = \{\boldsymbol{x} \in \mathbb{R}^d : HD_F(\boldsymbol{x}) \geq \alpha\}$, a general α halfspace depth trimmed mean can be defined as

$$\boldsymbol{\mu}_F(\alpha) = \int_{T_{HD_F}(\alpha)} \boldsymbol{x}\, w(HD_F(\boldsymbol{x}))dF(\boldsymbol{x}) \Big/ \int_{T_{HD_F}(\alpha)} w(HD_F(\boldsymbol{x}))dF(\boldsymbol{x}),$$

where $w(\cdot)$ is a suitable weight function on [0,1] such that $\boldsymbol{\mu}_F(\alpha)$ is well defined. $\boldsymbol{\mu}_F(\alpha)$ reduces to the Donoho (1982) halfspace depth trimmed mean when $w(t) = I_{[0,1]}(t)$. Because $T_{HD_F}(\alpha)$ is bounded, $\boldsymbol{\mu}_F(\alpha)$ is defined much more widely than the mean vector. The sample version of $\boldsymbol{\mu}_F(\alpha)$ is

$$\boldsymbol{\mu}_{F_n}(\alpha) = \int_{T_{HD_{F_n}}(\alpha)} \boldsymbol{x}\, w(HD_{F_n}(\boldsymbol{x}))dF_n(\boldsymbol{x}) \Big/ \int_{T_{HD_{F_n}}(\alpha)} w(HD_{F_n}(\boldsymbol{x}))dF_n(\boldsymbol{x}).$$

Remark 2.1. The important part of the weight function $w(t)$ is on $[\alpha, \alpha^*(F)]$, where $\alpha^*(F) = \sup_{\boldsymbol{x}\in\mathbb{R}^d}(HD_F(\boldsymbol{x}))$. Rousseeuw and Ruts (1999) showed that $HD_F(\boldsymbol{x})$ attains its supremum and thus $\alpha^*(F) = \max_{\boldsymbol{x}\in\mathbb{R}^d}(HD_F(\boldsymbol{x}))$. $\alpha^*(F) = 1$ if F is a degenerate distribution and it is not difficult to construct a distribution with $\alpha^*(F)$ being any value between 0.5 and 1. $\alpha^*(F)$ can also be less than 0.5. See Rousseeuw and Ruts (1999) and Massé (2004) for details. Without loss of generality, we consider the weight function on $[0, 1]$. For practical applications, we work with the sample version $\boldsymbol{\mu}_{F_n}(\alpha)$ and can choose a weight function based on $\alpha^*(F_n) = \max_{\boldsymbol{x}\in\mathbb{R}^d}(HD_{F_n}(\boldsymbol{x}))$.

Since the halfspace depth is affine invariant, $\boldsymbol{\mu}_F(\alpha)$ and $\boldsymbol{\mu}_{Fn}(\alpha)$ possess the following basic and important properties. The proof is relatively straightforward and thus omitted.

Proposition 2.1 *(1). $\boldsymbol{\mu}_F(\alpha)$ and $\boldsymbol{\mu}_{F_n}(\alpha)$ are affine equivariant.*
(2) If F is centrally symmetric about $\boldsymbol{\theta}$, then $\boldsymbol{\mu}_F(\alpha) = \boldsymbol{\theta}$ and $\boldsymbol{\mu}_{F_n}(\alpha)$ is an unbiased estimator of $\boldsymbol{\theta}$ for any α.

3. Weak convergence

Massé (2004) established the asymptotic distribution of $\boldsymbol{\mu}_{Fn}(\alpha)$ for the special case that $w(t) = 0$ for $t \in [0, \alpha)$. Massé (2009) further studied the general halfspace depth trimmed means, in which the condition that $w(t) = 0$ for $t \in [0, \alpha)$ is released, and obtained the asymptotic distribution of $\boldsymbol{\mu}_{Fn}(\alpha)$ for $d = 1$ and 2. What if $d > 2$? That is the focus of this section. To establish the asymptotic distribution of $\boldsymbol{\mu}_{Fn}(\alpha)$ in general, we invoke the following conditions.

Please cite this article in press as: Wang J., A note on weak convergence of general halfspace depth trimmed means. Statistics and Probability Letters (2018), https://doi.org/10.1016/j.spl.2018.07.005.