Journal of Multivariate Analysis 107 (2012) 306–318

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

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

Trimmed regions induced by parameters of a probability

Ignacio Cascos [a,*], Miguel López-Díaz [b]

[a] Department of Statistics, Universidad Carlos III de Madrid, Av. Universidad 30, E-28911 Leganés, Madrid, Spain

[b] Departamento de Estadística e I.O. y D.M., Universidad de Oviedo, c/ Calvo Sotelo s/n, E-33007 Oviedo, Spain

ARTICLE INFO

Article history:
Received 6 April 2011
Available online 2 February 2012

AMS subject classifications:
62H05
60D05
62G20

Keywords:
α-trimming of a probability
Depth function
Location-scale depth
Quantile trimming
Zonoid trimming

ABSTRACT

Consider any kind of parameter for a probability distribution and a fixed distribution. We study the subsets of the parameter space constituted by all the parameters of the probabilities in the α-trimming of the fixed distribution. These sets will be referred to as *parameter trimmed regions*. They are composed of all parameter candidates whose degree of suitability as such a parameter for the distribution is, at least, a specific value α.

In particular, we analyze location, scale, and location-scale parameters and study the properties of the trimmed regions induced by them. Several specific examples of parameter trimmed regions are studied. Among them, we should mention the zonoid trimmed regions obtained when the chosen parameter is the mean value and the location-scale regions of a univariate distribution obtained when the parameter is the pair given by the mean and the standard deviation.

1. Introduction

In the classical setting, depth functions assign the degree of centrality to each point with respect to a multivariate probability distribution or a data cloud. From this perspective, depth-trimmed regions are the level-sets of the depth functions, that is, they are sets constituted by all points whose depth is, at least, a given value. It is thus possible to consider these classical depth-trimmed regions as sets of location parameters (see [10,19] for the classical definition of a depth function, Liu et al. [11] for its statistical applications, and Zuo and Serfling [20] for depth-trimmed regions).

In this paper, the concept of depth is extended to parameter spaces. Any notion of *parameter depth* is a mapping from a parameter space into $\mathbb{R}$ and the depth of a candidate for a parameter of a probability distribution should be interpreted as its degree of suitability as such a parameter for the distribution. Associated with each parameter depth, there is a family of *parameter trimmed regions*, which are subsets of the parameter space formed by all elements whose parameter depth is, at least, some given value. The parameters considered in the present work are location, scale, and location-scale parameters. Notice that for location parameters, the parameter space is the Euclidean space of the same dimension as the data and we obtain a classical depth function, that is, a notion of *data depth*.

The first generalization of the notion of depth to a parameter space is due to Rousseeuw and Hubert [16], who worked on the linear regression model and defined the *regression depth*. Their idea was generalized by Mizera [12], emphasizing its applications to the location model, and developed in the location-scale model by Mizera and Müller [13], who built location-scale depths based on probabilistic models. Throughout those papers the concept of *nonfit* is essential. A nonfit is

* Corresponding author.
E-mail addresses: ignacio.cascos@uc3m.es, icascos@est-econ.uc3m.es (I. Cascos), mld@uniovi.es (M. López-Díaz).

0047-259X/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2012.01.024

a non-suitable candidate for a parameter with respect to a data set; and any of the former notions of depth is defined as the fraction of points that must be deleted from the data set so that the parameter is a nonfit.

The term *parameter depth*, referring to the suitability of a parameter, was coined by He and Portnoy in their discussion of [13].

Throughout the paper, we will place stronger emphasis on the families of parameter trimmed regions than on the parameter depths associated with them. For a given parameter and a probability distribution, the parameter trimmed region of a fixed level α is the set obtained by applying such a parameter to all the probabilities in the α-trimming of the original probability. If the starting point is a location parameter, we obtain *location trimmed regions*, if it is a scale parameter, *scale trimmed regions*, and finally, if it is a location-scale parameter, *location-scale trimmed regions*.

According to this terminology, the location trimmed regions are the classical depth-trimmed regions. In the new framework based on the α-trimming of a probability, we will obtain the quantile trimming and the zonoid trimming of Koshevoy and Mosler [9] as particular instances. However, other classical notions of data depth, like the halfspace depth (see [18]) or the simplicial depth (see [10]), are based on geometrical considerations about the data cloud and fall out of our scope.

This paper is structured as follows: in Section 2 we set our framework and define the new trimmed regions. In Section 3 we present particular families of location, scale, and location-scale trimmed regions obtained in the framework of the α-trimming of a probability. Section 4 contains the main properties of these location, scale, and location-scale trimmed regions, which are obtained as a consequence of the properties of the parameters and probabilities, except for the consistency of their empirical estimates, which is postponed to Section 5. Finally, Section 6 is devoted to some concluding remarks.

2. Preliminaries

We will start introducing the α-trimming of a probability and afterwards outline some characteristics of the sets we will be working with.

2.1. Main definitions and trimming of a probability

Let $\mathbb{P}$ denote the set of probability measures on $(\mathbb{R}^d, \mathcal{B}_d)$, with $\mathcal{B}_d$ the Borel σ-algebra on $\mathbb{R}^d$. The class $\mathbb{P}$ will be endowed with the topology of the weak convergence. For any $k \in \mathbb{N}$, let $\mathbb{P}_k$ be the subclass of $\mathbb{P}$ composed of all probabilities P which satisfy $\int \|x\|^k \, \mathrm{d}P(x) < \infty$.

For any probability $P \in \mathbb{P}$, we define a family of sets of probabilities, namely its α-*trimming*; see [4,5,1,2].

Definition 2.1. Given $\alpha \in (0, 1]$ and $P \in \mathbb{P}$, we define the α-*trimming of* P as

$$P^\alpha := \{Q \in \mathbb{P} : Q \leq \alpha^{-1}P\},$$

where, by $Q \leq \alpha^{-1}P$, we mean that for any $A \in \mathcal{B}_d$, it holds that $Q(A) \leq \alpha^{-1}P(A)$.

The α-trimming of a probability was already used by Gordaliza [8], who considered the probability $Q \in P^\alpha$ minimizing a penalty function, and named it *impartial trimming*. The reason for such a terminology is that the data to be trimmed are self-determined by the whole data set and do not necessarily correspond to the tail of the distribution, which makes the procedure particularly useful in the multivariate setting.

The sets P^α are convex and grow larger from $P^1 = \{P\}$ as α decreases, i.e., if $\alpha \leq \beta$, then $P^\alpha \supset P^\beta$. Further, P^α is tight and compact for the weak topology; see [5, Lemma 3]. Trivially, any $Q \in P^\alpha$ is absolutely continuous with respect to P. Moreover, any Radon–Nikodym derivative of Q with respect to P is bounded by α^{-1} a.s. $[P]$.

Since throughout this work we will be working with sets of probabilities and sets of points, we need some notion of convergence for sequence of sets.

The *superior limit* of a sequence of sets $\{A_n\}_n$ of certain topology space is defined as the set of limits of convergent subsequences in the corresponding topology, namely, $\limsup_n A_n = \{\lim_k a_{n_k} : a_n \in A_n\}$.

The *inferior limit* of the sequence $\{A_n\}_n$ is the set of limits of convergent sequences in the corresponding topology, that is, $\liminf_n A_n = \{\lim_n a_n : a_n \in A_n\}$.

If $\limsup_n A_n$ and $\liminf_n A_n$ coincide, then $\{A_n\}_n$ is said to converge and $\lim_n A_n = \liminf_n A_n = \limsup_n A_n$.

If $\{\hat{P}_n\}_n$ is a sequence of empirical probabilities of P, then it holds *a.s.* that

$$\lim_n \hat{P}_n^\alpha = P^\alpha \quad \text{a.s.}$$

for all $\alpha \in (0, 1]$; see [5, Theorem 12]. Álvarez-Esteban [2] proved that the above result holds when we consider a general sequence converging in the weak topology.

We denote by $T(P)$ any set of parameters associated with a probability distribution P, that is, $T(P) \in \mathcal{P}(\mathbb{R}^k)$, where $\mathcal{P}(\mathbb{R}^k)$ is the power set of $\mathbb{R}^k$ excluding the empty set.