Aalto University
School of Science
Degree programme in Systems and Operations Research

# Functional depth

MS-E2108 Independent Research Projects in Systems Analysis

*Author:* Aleksi Seppänen

*Instructor & supervisor:* Prof. Pauliina Ilmonen

31.5.2016

# Contents

# 1 Introduction

In some rare settings, the center of an object is easy to define. For example, a center of a circle is easy to measure. In some cases, the order of the data is also clear. For example, it is easy to order a set of real numbers. However, most of the time the center of an object is very difficult to define, as there is no universal definiton of center. Also the order of the data is often ill-defined. Multivariate data is one of the most common settings, where the center and the order of the data are not trivial.

The concept of depth was developed to address both of the aforementioned issues in multivariate setting, by introducing a way to measure the centrality of an observation with respect to a given set of data or an underlying distribution. The term depth was first used by J.W. Tukey in 1975 [5]. In the same paper, Tukey introduced the concept of halfspace depth, which became widely used in many procedures. In multivariate case, the halfspace depth of point $x \in \mathbb{R}^d$ is the "minimal" probability of $x$ belonging to any closed halfspace. Since then, various propositions for multivariate depth has been introduced, but even as of today, there is no standardized way to define the depth for multivariate data.

In many fields, e.g. biology, pharmacy, economics etc., the data is often generated by a stochastic function. Therefore it is often beneficial to consider the data as functions. For example, if a progress of several similar phenomena is observed at different points, and the sample sizes vary for each phenomena, multivariate analysis is not applicable. One practical example of functional data is the growth curves of a group of children.

If the center and ordering of multivariate data are difficult to define, for functional data the tasks are even harder. Functional data can be considered as an infinite dimensional multivariate data. However, compared to multivariate data, functional data has some additional properties, e.g. continuity and partial observability. Therefore different methods are required to define depth for functional data.

The earliest attempt to generalize median for functional data was conducted already in 1987. However, it was only in the 00s, when the first actual propositions for functional depth were published. During the last few years, several constructions for functional depth have been proposed, but as in the multivariate case, there is no standard way to define depth for functional data. [6]

One reason for the various definitions of functional depth is the fact, that it is always not clear, what properties of a function should be considered when defining the depth. Two of the main properties of a function are location and shape. Then, which should be more deep: 1) a function with similar values compared to other functions in the data set, but different shape, e.g. values vary slightly, whereas other functions are constant or 2) a function, which also gets constant values, but the values are notably different to the values of other functions. However, there are some properties, which all functional depth constructions should fulfill. Nieto-Reyes and Battey [4] defined a set of six properties, which functional depth should satisfy.

In this study, one of the proposed functional depth constructions is selected, and it is evaluated against the set of properties defined by Nieto-Reyes and Battey [4]. In addition, a few simulations are conducted to examine the practical behaviour of the selected functional depth in various circumstances. The selected functional depth is the band depth, introduced by López-Pintado and Romo [3].

The rest of the study is constructed as follows. In Chapter 2, the functional depth is explained, by using the definition of Nieto-Reyes and Battey [4]. In Chapter 3, the band depth is introduced. The same chapter includes also the proofs of the band depth satisfying or not satisfying the aforementioned set of six properties. The proofs presented in Chaprer 3 follow closely to those presented in [4]. The simulations are presented in Chapter 4, and Chapter 5 concludes the study with results and discussion.

# 2 Functional depth

In functional setting, one data point is a realisation of the random function $\{X(v) : v \in \mathcal{V}\}$, where $\mathcal{V}$ is a compact subset of $\mathbb{R}^d$ for $d \geq 1$ [4].

Nieto-Reyes and Battey [4] defined a statistical functional depth through a set of six properties, which functional depth functions should fulfill. Those properties, i.e. P-1. to P-6., are presented below, but first, one preliminary definition, also from [4], is presented.

**Definition 2.1.** Let $(\mathfrak{F}, \mathcal{A}, P)$ be a probability space and $\mathcal{E}$ be the smallest set in $\sigma$-algebra $\mathcal{A}$ such that $P(\mathcal{E}) = P(\mathfrak{F})$. Then the convex hull of $\mathfrak{F}$ with respect to $P$ is defined as

$$\mathfrak{C}(\mathfrak{F}, P) := \{x \in \mathfrak{F} : x(v) = \alpha L(v) + (1-\alpha)U(v) : v \in \mathcal{V}, \alpha \in [0, 1]\},$$

where $U := \{\sup_{x\in\mathcal{E}} x(v) : v \in \mathcal{V}\}$ and $L := \{\inf_{x\in\mathcal{E}} x(v) : v \in \mathcal{V}\}$.

**Definition 2.2.** Let $(\mathfrak{F}, \mathcal{A}, P)$ be a probability space. Let $\mathcal{P}$ be the space of all probability measures on $\mathfrak{F}$. The mapping $D(\cdot,\cdot) : \mathfrak{F} \times \mathcal{P} \to \mathbb{R}$ is a statistical functional depth if it satisfies properties P-1. to P-6, below.

P-1. *Distance invariance.* $D(f(x), P_{f(X)}) = D(x, P_X)$ for any $x \in \mathfrak{F}$ and $f : \mathfrak{F} \to \mathfrak{F}$ such that for any $y \in \mathfrak{F}$, $d(f(x), f(y)) = a_f \cdot d(x, y)$, with $a_f \in \mathbb{R}\backslash\{0\}$. Here $d(\cdot,\cdot)$ denotes some metric on $\mathfrak{F}$.

P-2. *Maximality at centre.* For any $P \in \mathcal{P}$ with a unique centre of symmetry $\theta \in \mathfrak{F}$ w.r.t. some notion of functional symmetry, $D(\theta, P) = sup_{x\in\mathfrak{F}} D(x, P)$.

P-3. *Strictly decreasing with respect to the deepest point.* For any $P \in \mathcal{P}$ such that $D(z, P) = sup_{x\in\mathfrak{F}} D(x, P)$ exists, $D(x, P) < D(y, P) < D(z, P)$ holds for any $x, y \in \mathfrak{F}$ such that $min\{d(y, z), d(y, x)\} > 0$ and $max\{d(y, z), d(y, x)\} < d(x, z)$.

P-4. *Upper semi-continuity in $x$.* $D(x, P)$ is upper semi-continuous as a function of $x$, i.e., for all $x \in \mathfrak{F}$ and for all $\epsilon > 0$, there exists a $\delta > 0$ such that

$$\sup_{y:d(x,y)<\delta} D(y, P) \leq D(x, P) + \epsilon \tag{1}$$

P-5. *Receptivity to convex hull width across the domain.* $D(f(x), P_{f(X)}) > D(x, P_X)$ for any $x \in \mathfrak{C}(\mathfrak{F}, P)$ with $D(x, P) < sup_{y\in\mathfrak{F}} D(y, P)$ and $f : \mathfrak{F} \to \mathfrak{F}$ such that $f(y(v)) = \alpha(v)y(v)$, with $\alpha(v) \in (0, 1)$ for all $v \in L_\delta$ and $\alpha(v) = 1$ for all $v \in L_\delta^c$, where

$$L_\delta = \underset{H\subseteq\mathcal{V}}{\text{argsup}} \{ \sup_{x,y\in\mathfrak{C}(\mathfrak{F},P)} d(x(H), y(H)) < \delta\}$$

for any $\sigma \in [\inf_{v\in\mathfrak{V}} d(L(v), U(v)), d(L, U)]$ such that $\lambda(L_\delta) > 0$ and $\lambda(L_\delta^c) > 0$. Here $\lambda$ denotes Lebesgue measure on $\mathcal{V}$.

P-6. *Continuity in $P$.* For all $x \in \mathfrak{F}$, for all $P \in \mathcal{P}$ and for every $\epsilon > 0$, there exists a $\delta(\epsilon) > 0$ such that $|D(x, Q) - D(x, P)| < \epsilon$ $P$-almost surely for all $Q \in \mathcal{P}$ with $d_\mathcal{P}(Q, P) < \delta$ $P$-almost surely, where $d_\mathcal{P}$ metricises the topology of weak convergence.

Property P-1. states that the depth does not change after mapping from $\mathfrak{F}$ to $\mathfrak{F}$, that preserves, up to a scaling factor, the relative distances between elements in the $d$ metric. The property P-1. ensures, for example, that if the same data is presented in different scales, or different units, the depths are not affected.

Properties P-2. and P-3. together ensure the centre-outward ordering, which is one of the features, depth functions were originally conceived for. Property

P-2. itself is somewhat paradoxial, as there is no unique notion of symmetry in function spaces and depth itself is meant to give meaning to the concept of centre of symmetry. Therefore the deepest element according to the depth function is also a valid definition of a centre of symmetry, which means that any depth function fulfills property P-2. as long as $sup_{x \in \mathfrak{F}} D(x, P) = max_{x \in \mathfrak{F}} D(x, P)$. It is more insightful to consider some special case, with particular $P$ for which many notions of centre of symmetry coincide at $\theta$. Nieto-Reyes and Battey [4] used Gaussian process as such $P$. Thus, instead of property P-2., the following property is considered.

P-2G. *Maximality at Gaussian process mean.* For $P$ a zero-mean, stationary, almost surely continuous Gaussian process on $\mathcal{V}$,
$D(\theta, P) = sup_{x \in \mathfrak{F}} D(x, P) \neq inf_{x \in \mathfrak{F}} D(x, P)$, where $\theta$ is the zero mean function.

The upper-semicontinuity in $x$ (property P-4.) is required for depth to be able to reveal the features of the underlying distribution. Property P-5. can be used to decrease the importance of a particular subset of the domain $L \subset \mathcal{V}$, where all functions exhibit little variability and most likely overlap significantly. The property might be desirable, because in such subsets, even small noise or measurement error can cause significat changes in the order of the functions. If the importance of those subsets can be decreased when calculating depth, the effects of small measurement errors and noise are mitigated as well. The most important implication of property P-6. is, that the depth calculated from empirical distibution converges almost surely to its corresponding population distribution.

# 3 Band depth

Band depth is one of the proposed functional depth functions. It was introduced by López-Pintado and Romo [3]. The band depth has the same underlying idea as simplical depth [2] in multivariate case. That is, the depth of a point $x$ is determined by the probability of $x$ belonging to a random simplex, constructed from random $j$-tuple $X_1, ..., X_j$.

The band depth is defined in the space of continuous functions on $\mathcal{V}$, i.e. $\mathcal{C}(\mathcal{V})$. Therefore, from now on, if not otherwise stated, we use $(\mathfrak{F}, d) = (\mathcal{C}(\mathcal{V}), \|\cdot\|_\infty)$, where $\|\cdot\|_\infty$ is the supremum norm.

The random simplex in functional space is [4]

$$S_j(P) = \{y \in \mathfrak{F} : y(v) = \alpha_1 X_1(v) + \ldots + \alpha_j X_j(v) : (\alpha_k)_{k=1}^j \in \Delta^j \forall v \in \mathcal{V}, (X_k)_{k=1}^j \sim P\}$$

where $j \geq 2$ and $\Delta^j \subset \mathbb{R}^{j-1}$ is the unit $j$-simplex. The band depth is then defined as [3]

$$D_J(x, P) = \sum_{j=2}^{J} P_{S_j}(x \in S_j(P)), \tag{2}$$

where $P_{S_j}$ is the probability measure over all random simplices $S_j$.

The corresponding sample analogue is obtained by replacing probability distribution $P$ by sampling distribution $P_n$. In $n$ samples, there $n$ choose $j$ different $j$-combinations, which means there are as many distinct j-simplices on $\mathfrak{F}$. The sample analogue is thus

$$D_J(x, P_n) = \sum_{j=2}^{J} \binom{n}{j}^{-1} \sum_{1 \leq i_1 < \ldots < i_j \leq n} I\{x \in B_{i_j}\} \tag{3}$$

where $I\{A\} \in \{0, 1\}$ is one if and only if $A$ is true,

$$B_{i_j} = \{y \in \mathfrak{F} : y(v) = \alpha_1 X_{i_1}(v) + \ldots \alpha_j X_{i_j}(v) : (\alpha_k)_{k=1}^{j} \in \Delta^j, \forall v \in \mathcal{V}\}$$

and $\{(i_1, ..., i_j) : i = 1, ..., n\}$ is the set of all $j$-tuples from $X_1, ..., X_n$.

The band $B_{i_j}$ is illustrated in Figure 1, where the band of two functions and one function, belonging to the band are presented.

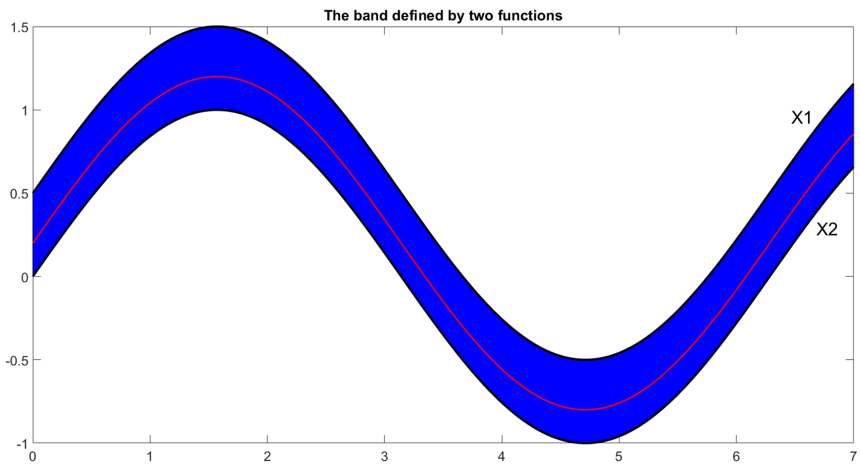


Figure 1: Illustration of the band formed by two functions: X1 and X2. A third function, belonging to the band, is marked by a red line.

Next, the band depth is evaluated against the six properties from definiton 2.2. All the proofs folllow the outline provided in [4], but some parts are clarified for making them easier to follow.

First, property P-1., i.e. distance invariance is considered. In the space of continuous functions, the set of functions satisfying $d(f(x), f(y)) = a_f \cdot d(x, y)$ for any $x, y \in \mathfrak{F}$ is

$$A_f = \{f : f(x(v)) = a(v)x(v) + b(v), |a(v)| = a_f > 0, \forall v \in \mathcal{V}\}.$$

Since $y(v) = \alpha_1 X_1(v) + \ldots + \alpha_j X_j(v)$ is equivalent to $a(v)y(v) + b(v) = \alpha_1(a(v)X_1(v) + b(v)) + \ldots + \alpha_j(a(v)X_j(v) + b(v))$ (because $\sum_{i=1}^{j} \alpha_i = 1$), the probabilities $P_{S_j}(x \in S_j(P_X))$ and $P_{S_j}(f(x) \in S_j(P_{f(X)}))$ are the same, when $f \in A_f$. Thus, $D_J(x, P_X) = D_J(f(x), P_{f(X)})$, so property P-1. is satisfied.

Instead of property P-2., property P-2G. is considered for the reasons described in the previous chapter. Thus, $P$ is now assumed to be a zero-mean, stationary, almost surely continuous Gaussian process on $\mathcal{V}$. From the definition of the band depth we get

$$\sup_{x \in \mathfrak{F}} D_J(x, P) \leq \sum_{j=2}^{J} \sup_{x \in \mathfrak{F}} P_{S_j}(x \in S_j(P)).$$

The simplices in the definition of band depth are defined through separate random draws from distribution $P$, which has mean $\theta = \mathbb{E}[X]$, and since $P$ is continuous, $P_{S_j}$ is a continuous distribution over simplices. Therefore it is clear, that the $x$, which maximizes the probability of $x$ belonging to a random $j$-simplex, is $x = \theta$. That yields to $\sup_{x \in \mathfrak{F}} D_J(x, P) = D_J(\theta, P)$, meaning the band depth satisfies the property P-2G.

The band depth do not satisfy the property P-3., as can be seen from the following counterexample. Let $P \in \mathcal{P}$ be a discrete distribution with $P(\{x_1\}) = P(\{x_2\}) = 1/2$, where $x_1$ and $x_2$ are both constant functions, with $x_1(v) = -c$ for all $c \in \mathcal{V}$ and $x_2(v) = c$ for all $c \in \mathcal{V}$, where $c \in \{c \in \mathbb{R} : c > 0\}$. Let $j = J = 2$. Then $P_{S_j}$ is also discrete, with $P_{S_j}(S_{j,1}) = P_{S_j}(S_{j,2}) = 1/4$ and $P_{S_j}(S_{j,3}) = 1/2$, where $S_{j,1} = \{x_1\}$, $S_{j,2} = \{x_2\}$ and $S_{j,3} = \{[x_1(v), x_2(v)] : v \in \mathcal{V}\}$. Then the band depth $D_J(z, P)$ has two global maxima; one at $z = x_1$ and another at $z = x_2$, both with $D_J(z, P) = 3/4$. Now we can set $z = x_1$ without loss of generality. Then, any functions $x, y \in \mathfrak{F} = \mathcal{C}(\mathcal{V})$, that are between $x_1$ and $x_2$, i.e. for which $|x(v)| < c$ and $|x(v)| < c$ for all $v \in \mathcal{V}$, and satisfy $\max\{d(y, z), d(y, x)\} < d(x, z)$, have band depth value of $D_J(x, P) = D_J(y, P) = 1/2$, which violates the property P-3.

Regarding the property P-4., the case $D_J(y, P) \leq D_J(x, P)$ is trivial. Let us consider the case $D_J(y, P) > D_J(x, P)$. The condition in property P-4. can be expressed as

$$\sup_{y:d(x,y)<\delta} \sum_{j=2}^{J} P_{S_j}(y \in S_j(P)) - \sum_{j=2}^{J} P_{S_j}(x \in S_j(P)) \leq \epsilon$$

For the left hand side of the inequality

$$\begin{aligned}
&\sup_{y:d(x,y)<\delta} \sum_{j=2}^{J} P_{S_j}(y \in S_j(P)) - \sum_{j=2}^{J} P_{S_j}(x \in S_j(P)) \\
&\leq \sum_{j=2}^{J} [\sup_{y:d(x,y)<\delta} P_{S_j}(y \in S_j(P)) - P_{S_j}(x \in S_j(P))] \\
&\leq \sup_{y:d(x,y)<\delta} P_{S_j}(y \in S_j(P) \cap x \notin S_j(P) - \inf_{y:d(x,y)<\delta} P_{S_j}(y \notin S_j(P) \cap x \in S_j(P) \\
&\leq \sup_{y:d(x,y)<\delta} P_{S_j}(y \in S_j(P) \cap x \notin S_j(P)
\end{aligned}$$

For

$$\sup_{y:d(x,y)<\delta} P_{S_j}(y \in S_j(P) \cap x \notin S_j(P) \leq \epsilon$$

to hold, we can take $\delta < \sup\{\eta > 0 : P(x \notin S_j(P) \cap \min_{v \in \mathcal{V}}(|x(v) - L_j(v)|, |x(v) - U_j(v)|) < \eta) \leq \epsilon\}$, which ensures that the property P-4. is satisfied. Here $L_j(v) := \min_{y \in \mathfrak{X}_j} y(v)$ and $U_j(v) := \max_{y \in \mathfrak{X}_j} y(v)$, where $\mathfrak{X}_j = (X_1, ..., X_j)$ and $X_1, ..., X_j \sim P$.

Property P-5. is another one, which band depth fails to satisfy. Again, this can be proofed by a counterexample. Let $P$ be discrete, with probabilities $P[x_i] = 1/3$ for $i = 1, 2, 3$ and $x_1(v) > 0, x_2(v) = 0$ and $x_3(v) < 0$ for all $v \in \mathcal{V}$, where $x_1$ and $x_2$ are non-constant functions. When function $f$ is as defined in property P-5., also $f(x_1)(v) > 0, f(x_2)(v) = 0$ and $f(x_3)(v) < 0$ for all $v \in \mathcal{V}$. When $j \in \{2, 3\}$, the transformation just shrinks the convex hull of any simplex in region $L_\delta$, but the probabilities of transformed points belonging to the transformed random simplices remain the same as before the transformation. Thus, $D_J(x, P_X) = D_J(f(X), P_{f(X)}$, which violates the property P-5.

The band depth does not satisfy the property P-6. for all $\mathfrak{F}$, but do satisfy, when $\mathfrak{F}$ is restricted to the space of equicontinuous functions on $\mathcal{V} \subset \mathbb{R}$. Since $d_{\mathcal{P}}(P, Q)$ in P-6. metrices the weak topology, $d_{\mathcal{P}}(P, Q) < \delta \to 0$ can be written as $X_\delta \rightsquigarrow Y$ as $\delta \to 0$, where $\rightsquigarrow$ denotes weak convergence and $X_\delta : \Omega \to \mathfrak{F}$ and $Y : \Omega \to \mathfrak{F}$ are random variables, such that for any $A \in \mathcal{A}$, $P(A) = \mathbb{P}(X_\delta^{-1}(A))$ and $Q(A) = \mathbb{P}(Y^{-1}(A))$, where $\mathbb{P}$ is a probability on the underlying sample space $\Omega$. The Portmanteau theorem (e.g. Th. 11.3.3, [1]) states that the probability measures $V_N$ and $V$ satisfy $V_N \rightsquigarrow V$ if and only if $\mathbb{E}f(V_N) \to \mathbb{E}f(V)$ for all bounded Lipschitz functions $f$. Let $X_{\delta,1}, ..., X_{\delta,J}$ to be i.i.d. realizations of random variable $X_\delta$ and $Y_1, ..., Y_J$ to be i.i.d. realizations of random variable $Y$.

As $X_\delta \rightsquigarrow Y$, also $\sum_{k=1}^{j} \alpha_k X_{\delta,k} \rightsquigarrow [\sum_{k\neq l} \alpha_k X_{\delta,k}] + \alpha_l Y_l$, for any $l \in \{1, ..., j\}$, where $j \in \{2, ..., J\}$ and for any $(\alpha_1, ..., \alpha_j) \in \Delta^j$. Then, by the Portmanteau theorem, there exist a $\delta < \delta_l$ such that

$$|\mathbb{E}[f(\sum_{k=1}^{j} \alpha_k X_{\delta,k})] - \mathbb{E}[f([\sum_{k\neq l} \alpha_k X_{\delta,k}] + \alpha_l Y_l)]| < \delta/j.$$

Hence

$$\begin{aligned} &|\mathbb{E}[f(\sum_{k=1}^{j} \alpha_k X_{\delta,k})] - \mathbb{E}[f(\sum_{k=1}^{j} \alpha_k Y_k)]| \\ \leq &\sum_{l=1}^{j} |\mathbb{E}[f(\sum_{k=1}^{j} \alpha_k X_{\delta,k})] - \mathbb{E}[f([\sum_{k\neq l} \alpha_k X_{\delta,k}] + \alpha_l Y_l)]| < \delta \end{aligned}$$

for all $\delta < \min\{\delta_l :\in \{1, ..., j\}\}$. Let us define $Z_{X(\delta),j}(\boldsymbol{\alpha}) := \sum_{k=1}^{j} \alpha_k X_{\delta,k}$ and $Z_{Y,j}(\boldsymbol{\alpha}) := \sum_{k=1}^{j} \alpha_k Y_k$. Then, by the Portmanteau theorem, it follows that $Z_{X(\delta),j}(\boldsymbol{\alpha}) \rightsquigarrow Z_{Y,j}(\boldsymbol{\alpha})$, for any $j \in \{2, ..., J\}$ and any $\boldsymbol{\alpha} \in \Delta^j$. Thereby, for every finite subset $\boldsymbol{\alpha}_1, ..., \boldsymbol{\alpha}_l$, where $\boldsymbol{\alpha}_k \in \Delta^j$ for each $k \in \{1, ..., l\}$, holds $(Z_{X(\delta),j}(\boldsymbol{\alpha}_1), ..., Z_{X(\delta),j}(\boldsymbol{\alpha}_l)) \rightsquigarrow (Z_{Y,j}(\boldsymbol{\alpha}_1), ..., Z_{Y,j}(\boldsymbol{\alpha}_l))$.

Theorem 1.5.4 of van der Vaart and Wellner [7] state that for an arbitrary $X_\alpha : \Omega_\alpha \to \mathbb{L}^\infty(T)$, $X_\alpha$ converges weakly to a tight limit if and only if $X_\alpha$ is asymptotically tight and the marginals $(X_\alpha(t_1), ..., X_\alpha(t_l))$ converge weakly to a limit for every finite subset $t_1, ..., t_l$ of $T$. It also states, that if $X_\alpha$ is asymptotically tight, and its marginals converge weakly to the marginals of a stochastic process $X$, i.e. $(X(t_i), ..., X(t_l))$, then there is version of $X$ such that $X_\alpha \rightsquigarrow X$. Here $(Z_{X(\delta),j}(\boldsymbol{\alpha}_1), ..., Z_{X(\delta),j}(\boldsymbol{\alpha}_l))$ is a finite set of arbitrary marginals of the stochastic process $Z_{X(\delta),j} := \{Z_{X(\delta),j}(\boldsymbol{\alpha}) : \boldsymbol{\alpha} \in \Delta^j\}$, which is the map $Z_{X(\delta),j} : \Omega^j \to \mathfrak{F}(\Delta^j) = \mathcal{C}(\mathcal{V}, \Delta^j) \subset \mathbb{L}^\infty(\mathcal{V} \times \Delta^j)$, where $\mathbb{L}^\infty(\mathcal{V} \times \Delta^j)$ is the space of bounded functions from $(\mathcal{V} \times \Delta^j)$ to $\mathbb{R}$. In the same way, $(Z_{Y,j}(\boldsymbol{\alpha}_1), ..., Z_{Y,j}(\boldsymbol{\alpha}_l))$ is an arbitrary finite set of marginals of the stochastic process $Z_{Y,j} := \{Z_{Y,j}(\boldsymbol{\alpha}) : \boldsymbol{\alpha} \in \Delta^j\}$. Therefore, by the aforementioned definition, $Z_{X(\delta),j} \rightsquigarrow Z_{Y,j}$ for every $j \in \{2, ..., J\}$, if $Z_{X(\delta),j}$ is asymptotically tight for any $j \in \{2, ..., J\}$. $Z_{X(\delta),j}$ is asymptotically tight, if for every $\xi > 0$ there exist a compact set $K$ such that $\liminf_{\delta\to 0} P_{Z(\delta),j}(Z_{X(\delta),j} \in K^\eta) \leq 1-\xi$

for every $\eta > 0$, where $P_{Z(\delta),j}$ is defined at every $A \in \mathcal{A}$ by $P_{Z(\delta),j}(A) = \mathbb{P}^j(Z^{-1}_{X(\delta),j}(A))$.

According to another theorem by van der Vaart and Wellner (Th. 1.5.7, [7]), $P_{Z(\delta),j}$ is asymptotically tight if and only if $P_{Z(\delta),j}(v, \boldsymbol{\alpha})$ is tight in $\mathbb{R}$ for every $w = (v, \boldsymbol{\alpha})$, and there exist a semimetric $d_w$ on $\mathcal{W} = (\mathcal{V} \times \Delta^j)$ such that $(\mathcal{W}, d_w)$ is totally bounded and $P_{Z(\delta),j}$ is asymptotically uniformly $d_w$-equicontinuous in probability, i.e. for every $\kappa, \varsigma > 0$ there exists a $\gamma$ such that

$$\limsup_{\delta \to 0} P_{Z(\delta),j}\left(\sup_{w,w':d_w(w,w')<\gamma} |Z_{X(\delta),j}(w) - Z_{X(\delta),j}(w')| > \kappa\right) < \varsigma.$$

$Z_{X(\delta),j}(v, \boldsymbol{\alpha})$ is tight, because $\mathfrak{F}$ is copmlete. Since $\mathcal{V}$ is compact, $\mathcal{W}$ is also compact. Thereby $(\mathcal{W}, d_w)$ is totally bounded, when we select $d_w$ to be $l_1$ norm. Then, it remains to show that $P_{Z(\delta),j}$ is asymptotically uniformly $d_w$-equicontinuous in probability. Now we have

$$\begin{aligned}
&Pr\bigg(\sup_{w,w':d_w(w,w')<\gamma} |Z_{X(\delta),j}(w) - Z_{X(\delta),j}(w')| > \kappa\bigg) \\
&\leq Pr\bigg(\sup_{w,w':d_w(w,w')<\gamma} |Z_{X(\delta),j}(v, \boldsymbol{\alpha}) - Z_{X(\delta),j}(v', \boldsymbol{\alpha})| > \kappa/2\bigg) \\
&\quad + Pr\bigg(\sup_{w,w':d_w(w,w')<\gamma} |Z_{X(\delta),j}(v', \boldsymbol{\alpha}) - Z_{X(\delta),j}(v', \boldsymbol{\alpha}')| > \kappa/2\bigg) = I + II.
\end{aligned}$$

Since $\mathfrak{F}$ is now defined to be the space of $d_w$-equicontinuous functions over $\mathcal{V}$, and since $d_w$-equicontinuity is preserved under convex combinations, $Z_{X(\delta),j}(\cdot, \boldsymbol{\alpha})$ is $d_w$-equicontinuous with probability 1. Hence, for every $\kappa, \varsigma > 0$, there exists a $\gamma > 0$ such that $I < \varsigma/2$. As $v' \in \mathcal{V}$ is fixed in $II$, it is obvious, that when $\gamma$ is small enough, it follows that $II < \varsigma/2$. As the bound on $I$ and $II$ hold independently of $\delta$, $P_{Z(\delta),j}$ is asymptotically uniformly $d_w$-equicontinuous in probability.

Thus, by Theorem 1.5.4 of van der Vaart and Wellner [7], we now know that $Z_{X(\delta),j} \rightsquigarrow Z_{Y,j}$ for every $j \in \{2, ..., J\}$. Now, by the Portmanteau theorem (Th. 11.3.3, [1]), we get that there exists a $\eta \searrow 0$ as $\delta \searrow 0$ such that $\rho(P_{Z(\delta),j}, Q_{Z(Y),j}) = M < \eta(\delta)$, where $Q_{Z(Y),j}(a) = \mathbb{P}^j(Z^{-1}_{Y,j}(A))$ and $\rho$ denotes the Lévy-Prokhorov metric. That means, that for all $A \in \mathcal{A}$, $P_{Z(\delta),j}(A) \leq Q_{Z(Y),j}(A^\xi) + \xi$ for all $\xi \in [M, \eta(\delta))$, where $A^\xi := \{p \in \mathfrak{F} | \exists q \in A, d(p,q) < \xi\}$ is the $\xi$-neighborhood of $A$. By setting $B(x) = \cup\{A \in \mathcal{A} : x \in A\}$, we get $P_{Z(\delta),j}(B(x)) \leq Q_{Z(Y),j}(B(x)^\xi) + \xi$ for all $\xi \in [M, \eta(\delta))$. By the symmetry of Lévy-Prokhorov metric and since $B(x) \subset B(x)^{\xi}$ for $\xi > 0$, we get that $|P_{Z(\delta),j}(B(x)) - Q_{Z(Y),j}(B(x))| \leq \xi < \eta(\delta)$. Thus,

$$|D_J(x,P) - D_J(x,Q)| \leq \sum_{j=2}^{J} |P_{Z(\delta),j}(B(x)) - Q_{Z(Y),j}(B(x))| < (J-1)\eta(\delta).$$

By setting $\epsilon = (J-1)\eta(\delta)$ and replacing every $\delta_l$ in the above derivations with $\delta = \eta^{-1}(\epsilon/(J-1))$, property P-6. is satisfied with the additional constraint, i.e. $\mathfrak{F}$ restricted to be the space of equicontinuous functions on $\mathcal{V} \subset \mathbb{R}$.

To sum it up, from the six desired properties for the functional depth, the band depth satisfies four, i.e. P-1., P-2G., P-4. and P-6. Altough for P-6. additional constraint is required. The band depth fails to satisfy properties P-3. and P-5.

# 4 Simulations

Then, the properties of the band depth are examined in practice through simulation data. In this study, two simulation scenarios are considered. The first case is to examine, what kind of effect different noises in the measurements cause in the band depth values. In the second case, the receptivity of the band depth to shape differences is considered.

In the first simulation scenario, the data consist of random constant functions, with additional noise, i.e. $X_i(t) = Y_i + Z_i(t)$, where $Y_i$ for every $i \in \{1, ..., N\}$ are independent realizations of random variable $Y \sim \mathcal{N}(0,1)$, $Z_i(t)$ for every $i \in \{1, ..., N\}$ and for every $t \in \{1, ..., T\}$ are independent realizations of random variable $Z \sim \mathcal{N}(0, \sigma_z^2)$. $N$ is the number of simulated functions and $T$ is the number of observed points in each function. The parameters used in every simulation are $N = 30$ and $T = 100$. When calculating the band depth values, the maximum number of functions used for constructing the simplices, were set to $J = 3$.

Similar simulations are run with four different noise, by using different variances $\sigma_z^3$. The used variances are $\sigma_{z,1}^2 = 0$, $\sigma_{z,2}^2 = 0.05$, $\sigma_{z,2}^2 = 0.2$ and $\sigma_{z,3}^2 = 0.5$. One realization from each simulation with different $\sigma_z^2$, but same $Y_i$ for every $i \in \{1, ..., N\}$, are shown in Figures 2 to 5. In the figures, the function with the highest band depth, is marked by red line. The colors of other functions are also determined by their corresponding band depth values, such that darker color indicate higher depth value compared to other functions in the same figure.

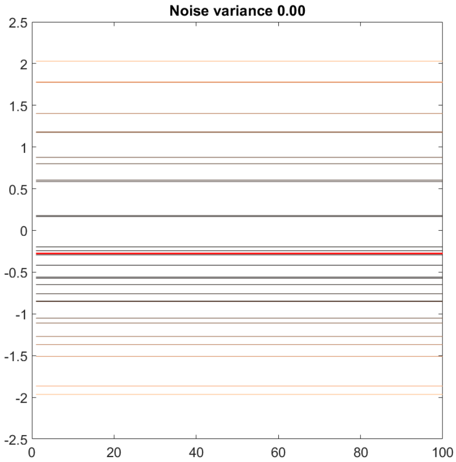


Figure 2: One realization of simulation without additional noise.

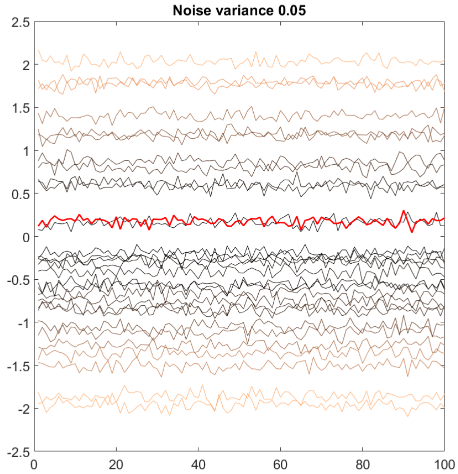


Figure 3: One realization of simulation with noise variance equal to 0.05.

From Figure 2, it can be seen that when there is no noise, the band depth gives the highest depth value to the function, corresponding to the sample median. Based on Figure 2 and Figure 3, it seems that the band depth works better as some noise is added, since the function getting the highest depth value is closer to zero compared to the deepest function in Figure 2. In both of the figures, the depth values decrease smoothly as the distance to the deepest function increases. The highest band depth value in the first case, i.e. without noise, is 0.774 and in the second case, with $\sigma_{z,2}^2 = 0.05$, the maximum depth is 0.697. In all of the simulations, the smallest achieved band depth value is 0.097, which means that the function belonged only to those simplices, which were defined by using the function itself. Thereby that is the lowest possible band depth value, when parameters $N$ and $J$ are fixed. Every function, that for any $t$, obtains higher or lower value than any other sample, gets the lowest possible band depth value regardless of the rest of the function.

Examples of simulation results with variances $\sigma_{z,2} = 0.2$ and $\sigma_{z,3} = 0.5$ are presented in Figures 4 and 5, respectively. There are still some differences in the band depth values in Figure 4, as the highest depth value is 0.463. In Figure 5, there is not much differences in the colors of the functions, and the highest depth value in that case is only 0.145. There are also great differences in the amount of functions obtaining the lowest possible depth value. Without additional noise, the number of functions with minimum band dept is obviously two. When the noise is increased to 0.05, the number

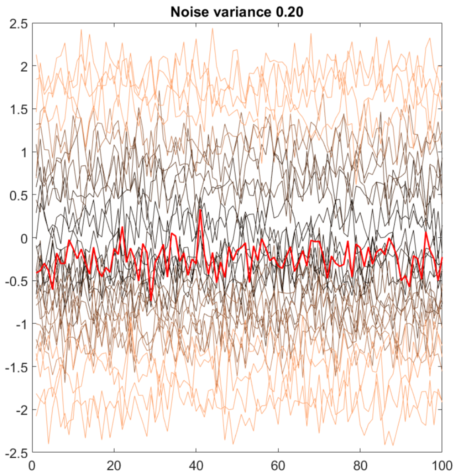


Figure 4: One realization of simulation with noise variance equal to 0.20.

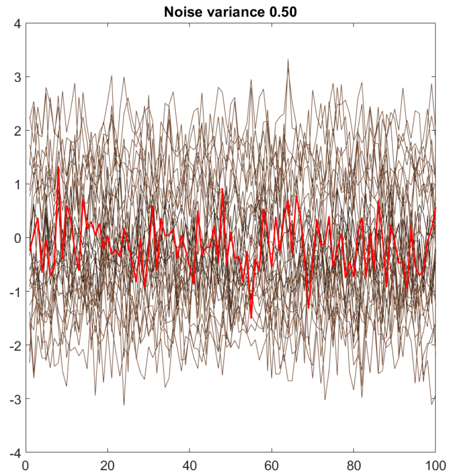


Figure 5: One realization of simulation with noise variance equal to 0.50.

of minimum band depth functions is increased by only one. With $\sigma^2_{z,2} = 0.2$, the corresponding amount is already seven, but in the last case, over half of the 30 functions, i.e. 18 functions, get the same, lowest possible depth value.

The second simulation scenario considered the band depth of functions with differences in shapes. For that purpose, simulation model, similar to the noiseless case in the last scenario, is used. The difference in this case is, that if $Y_i < \beta$, then $X_i(t) = Y_i + 0.1sin(t/2)$. Otherwise $X_i(t) = Y_i$, for every $t \in \{1, ..., T\}$. One realization of such simulation is presented in Figure 6. In this simulation, parameter $\beta = 0.16$ is used. The colors in the figure has the same meanings as in the previous figures.

From Figure 6, it can be seen that the band depth do not differentiate the functions based on the shape, as the function with the largest depth value is actually one of the functions with the shape distortion.

# 5 Results and discussion

As showed in Chapter 3, the band depth fail to satisfy two of the six properties. The properties, which the band depth fail to satisfy, are P-3., i.e. strictly decreasing with respect to the deepest point, and P-5. i.e. receptivity to convex hull width across the domain. However, the failure to fulfill all of the properties does not mean that the band depth is especially bad as a

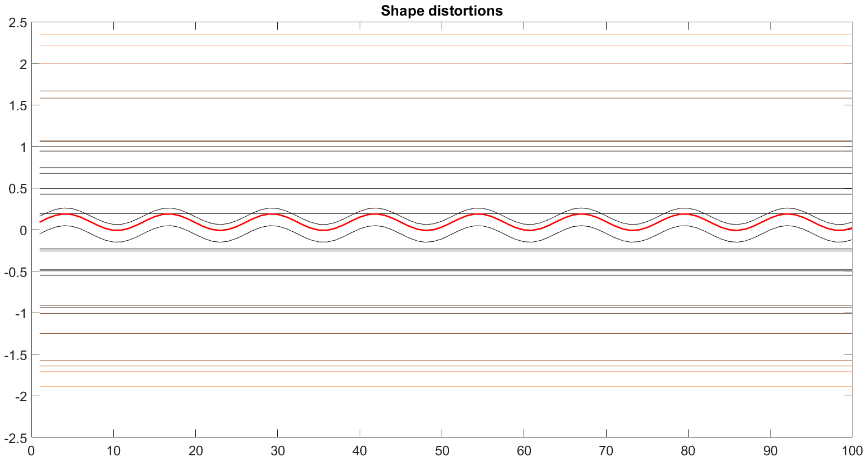


Figure 6: One realization of simulation with functions with different shape.

functional depth. As a comparison, in the study of Nieto-Reyes and Battey [4], six different functional depth constructions were considered, and none of them were able to satisfy all the properties.

In addition, not all of the properties are desired in every occasion. For example, the property P-5., which is related to the robustness of the depth function, might not always be desirable. The property state, that the regions, where the convex hull of the functions is small, should contribute less to the depth value compared to other regions, with wider convex hull. In some cases however, it might be that a seemingly normal function exhibit slight abnormality at the region, where all the function values are close to each other. Then it might be appropriate for the functional depth to identify the abnormality, by giving the function smaller depth value, which might not happen, if the impact of such region is neglected.

One of the main tasks for functional depth is to provide a center outward ordering of the data. Therefore the failure to satisfy property P-3. might be considered as a more significant drawback of the band depth. This should be taken into account especially if the underlying distribution is examined by using the order statistics of sample functions based on the band depth values.

From the simulations in Chapter 4, it can be seen that the band depth works well and as expected, when the noise variance is relatively low. When comparing the two simulations with the smallest variances, it looks like the one with higher variance has a deepest function closer to zero. However, this is only due to the small sample size, and the deepest function in the constant case would get closer to zero, if the sample size is increased.

In the two simulations with higher noise variance, the range of the depth values grew smaller and smaller as the noise increased. In the simulation with the highest noise, 60% of the samples received the same depth value. That is also not desirable, if the functions should be ordered based on the depth values. Due to the inherent properties of the band depth, even one abnormal value can determine the band depth of the whole function. In some applications, that is highly sought after property (for example in health care, even the smallest events can be significant), whereas in other applications it might make the band depth totally useless. When this property is combined with the failure to satisfy property P-5. and the observation from the high noise simulation case, the following example can be concluded. Consider a set of functions, which are 99.9% of the time the same as in the first simulation case, and during the rest 0.1%, they get values corresponding to the high noise simulation case, multiplied of very small $\epsilon > 0$. Now, by using the band depth, those functions would get depth values only based on the 0.1% region, and the depths would be the same as in the high noise simulation case. Again, it depends on the application, wether this is a good or bad feature. Nevertheless, some of the shortcomings of the band depth are corrected in other functional depth constructions, closest example being modified band depth by López-Pintado and Romo [3], but they have then their own drawbacks.

Another lack of the band depth is the negligence of the shape of the functions, which can be clearly noticed in the last simulation case. However, if the function values are close enough, functions with different shapes might easily cross over many functions with 'normal' shape. In that case, the band depth values are highly affected by the shape differences, even though the shape is not explicitly considered.

# References

[1] R. M. Dudley. *Real Analysis and Probability*. Cambridge University Press, second edition, 2002. Cambridge Books Online.

[2] R. Y. Liu. On a Notion of Data Depth Based on Random Simplices. *The Annals of Statistics*, 18(1):405–414, 1990.

[3] S. López-Pintado and J. Romo. On the Concept of Depth for Functional

Data. *Journal of the American Statistical Association*, 104(486):718–734, 2009.

[4] A. Nieto-Reyes and H. Battey. A topologically valid definition of depth for functional data. 2014.

[5] J. W. Tukey. Mathematics and the Picturing of Data. In R. D. James, editor, *International Congress of Mathematicians 1974*, volume 2, pages 523–532, 1974.

[6] G. Van Bever and D. Paindaveine. Contributions to nonparametric and semiparametric inference based on statistical depth. 2013.

[7] A. W. van der Vaart and J. A. Wellner. *Weak Convergence and Empirical Processes*. Springer Series in Statistics. Springer New York, New York, NY, 1996.