Depth Functions on General Data Spaces, I.
Perspectives, with Consideration of “Density”
and “Local” Depths

Robert Serfling¹

Serfling & Thompson Statistical Consulting

May 2019

¹Serfling & Thompson Statistical Consulting, 1921 Sparrows Point Drive,
Plano, Texas 75023 USA. Email: rjserfling@gmail.com. Website:
www.utdallas.edu/~serfling.
Abstract

General perspectives on the depth approach are developed in order to aid the interpretation of developments and findings to date and to provide viewpoints apropos to continuing development. These treat data spaces, dimensionality, population versions, evaluation of depth functions, relevant inference settings, and depth functions versus density functions. A concept of “essential support” for probability distributions is introduced. Notions of “density depth” and “local depth” are examined. Some special results on dense and nowhere dense sets are obtained. This work is carried out in the setting of a general data space.

AMS 2000 Subject Classification: Primary 62G99 Secondary 62H99

Key words and phrases: Depth functions; High and infinite dimensionality; Inference models; Density functions; Local depth; Dense and nowhere dense sets.
1 Introduction

For nonparametric description and exploration of univariate data sets, the sample quantiles are fundamental and fruitful. In particular, besides locating the “middle half”, the quartiles yield the median as a measure of location and the interquartile range (IQR) as a measure of spread, and these artifacts yield the boxplot, a useful visual summary. Other inter-quantile ranges demark the central 75% or 90% or 95% regions, for example. There are useful quantile-based measures of skewness and kurtosis. Further, the quantiles generate order statistics and ranks, which not only have key roles themselves but also yield the important classes of L-statistic estimators and linear rank tests serving nonparametric inference goals.

A further key element in exploring univariate data is outlier detection. The boxplot incorporates a quantile-based approach the quartiles and IQR to define outlier detection thresholds (“fences”), in effect using a scaled-deviation type outlyingness function based on the median and IQR (as discussed in Wang and Serfling, 2015). Other scaled-deviation type outlyingness functions as well have long been in use in the univariate setting, for example with the median and MAD as robust location and spread measures. Note that in defining scaled-deviation outlyingness functions, the “center” or location measure is specified in advance and provided as an input.

Instrumentally underlying these univariate methods is the natural order of the real line \( \mathbb{R} \). However, in considering possible extensions of these univariate methods to data sets in higher dimensional Euclidean spaces and to still more complex data settings, a major conceptual challenge arises: no longer is such a natural order available, nor is a well-determined median-type notion of center. As a compensation, a natural and effective approach is to order points in some center-outward fashion relative to a “center” that itself is determined by the ordering process. This is the “depth function approach” conceived by Tukey (1975), where “deeper” corresponds to more central, and the deepest point defines a “center”.

As a key motif of the depth approach, “centrality” receives a stronger focus and a more explicit structural role than in the univariate setting. Thus, given a data set \( X_n \) of size \( n \) in a data space \( \mathcal{X} \), one identifies not only an \( X_n \)-based “center” in \( \mathcal{X} \), but also, for any point \( x \in \mathcal{X} \), the degree of its centrality relative to the data set \( X_n \), yielding a center-outward ordering of points in \( \mathcal{X} \). In particular, the Tukey (1975) formulation was in the simplest non-univariate case \( \mathcal{X} = \mathbb{R}^2 \), with a “depth function” based on counts of points in halfspaces. Its extension to \( \mathbb{R}^d \) for arbitrary \( d \), as well as to other data spaces, is immediate. Beginning notably with Liu (1988, 1990), many other versions of “depth” have now been formulated for Euclidean and Banach or Hilbert spaces, reflecting a diversity of points of view. Among depths on \( \mathbb{R}^d \),
the "Tukey" or "halfspace" depth remains a leading competitor as well as a prototypical structure for formulations of "depth" in other contexts such as parameter spaces (see Rousseeuw and Hubert, 1999, Mizera, 2002, and Zuo, 2018a, for example).

The "depth function" approach yields associated "order statistics" and associated "outlyingness", "quantile", and "rank" functions. These Depth, Outlyingness, Quantile, and Rank functions may be viewed coherently within a single framework as essentially equivalent yet distinct entities interrelated through a "DOQR paradigm", described in Serfling (2010), for example. In particular, depth and outlyingness functions are equivalent through an inverse relationship, and thus the long-standing use of outlyingness functions in the univariate setting represents already an implicit use of depth functions and center-outward ordering. Further, univariate scaled-deviation outlyingness functions have long been extended to multivariate versions, for example the Mahalanobis distance. Therefore, what is especially novel in the Tukey depth approach in $\mathbb{R}^d$ is not the center-outward ordering itself, but rather the focus on centrality and the method of accomplishing the ordering without necessarily specifying in advance a "center", which instead is given by the deepest point of the depth function. This, of course, leads to new outlyingness functions as well for the multivariate case. Further, although determination of a center and identification of outliers are technically equivalent algorithmically, these goals are very different both conceptually and in application. An emphasis on centrality is decidedly fundamental in extending univariate order statistics to the multivariate case.

Let us note that any notion of "centrality" comparing points in a space $\mathcal{X}$ relative to their representativeness of the bulk of the data points, must take into account not only the locations of the data points but also, when present, any other distinguishing features they might possess. Thus, in the case of functional data for example, points have not only locations but also shapes. In general, a suitable depth function must take account of all relevant features of points in the given data space, making the depth approach necessarily sensitive to the particular type of data space $\mathcal{X}$.

Broad methodological treatments of the depth approach on Euclidean data spaces are provided by Liu, Parelius, and Singh (1999) for general depths, by Mosler (2002) focusing on the zonoid depth, and by Oja (2010) focusing on the spatial depth. Theoretical developments emphasizing desired properties for depth functions have been developed by Zuo and Serfling (2000a,b) and Serfling (2006) for Euclidean data and by Nieto-Reyes and Battey (2016) and Mosler and Polyakova (2018) for functional data. The many depth functions that have been postulated range widely in their structural form, properties, robustness, computational ease, and asymptotic behavior. These aspects trade off against each other and have differing subjective priorities, so that no depth
exhibits overall dominance by any common agreement.

In practice, one chooses a depth function for practical use by evaluating candidates according to one’s adopted criteria. Concerning such evaluation of prospective depth functions, some fundamental questions arise. In what way is the particular type of data space $\mathcal{X}$ germane to the choice? How is the dimensionality of $\mathcal{X}$ relevant? Is the density function a particular example of depth function? Should depth contours mimic those of the density function? Does unfavorable performance found for a “population” depth in some special case carry substantive implications for practical use of the sample version? Are some populations pathologically unsuitable for the formulation of depth functions? From an inference standpoint, does the depth approach apply not only in the broadest nonparametric setting as usually formulated, but also in narrower semiparametric or parametric settings, and, if so, under what type of modifications if any? Are some basic principles involved in critically evaluating depth functions? Some of these questions have been addressed substantially in the depth function literature but still lack definitive answers, while some have been rather overlooked but also should be considered.

In Sect. 2, perspectives and “assertions” are provided which address the above issues and contribute toward a general view of the depth approach that is unencumbered by details specific to particular implementations for specific cases of $\mathcal{X}$. These concern aspects of data sets and data spaces, the nature of high or infinite dimensionality, sample depths versus population versions, relevant probability distributions for population versions, critical evaluation of depth functions, relevant inference settings for the depth approach, depth functions versus density functions, and the notion of “local depth”. While some of the expressed viewpoints (Assertions 2.1-2.10) may have advocates to the contrary, these considerations are important in clarifying and guiding the role and implementation of the depth approach.

A further fundamental challenge is to specify what requirements must be satisfied for a proposed depth function to be admissible as such. This entails additional key questions. What intrinsic features of a depth function should hold across all types of data space $\mathcal{X}$? What invariance under transformations of coordinates is desired? What kind of monotonicity as one moves outward from the center? A depth function generates a family of contours, but is the converse true in a practical sense? When the population satisfies some notion of symmetry about a center, is a “good” depth maximal at this center? How generally can nonparametric notions of symmetry be defined? These all are addressed in the companion paper Serfling (2019), also in the setting of general $\mathcal{X}$, drawing upon the perspectives of the present paper and developing a number of specialized technical results as well.
2 Some perspectives on the depth approach

In any given data setting, various depth functions that are specialized to the setting can be proposed, and these should be compared according to selected criteria that also are specialized to the particular data setting. However, there are certain perspectives and preliminaries that apply from a general standpoint and should be kept in mind as fundamental considerations.

2.1 Data sets and data spaces

In practice, computations with a data set \( X_n \) necessarily involve but finitely many components of any data point. Hence restriction to a data space \( \mathcal{X} \) of some finite dimension \( d \) is an essential feature of the depth approach as actually carried out in practice. We thus assert:

**Assertion 2.1.** The depth approach applies in practice to data sets \( X_n \) of finite size \( n > 1 \) in some data space \( \mathcal{X} \) of dimension \( d < \infty \).

Further, as discussed in Sect. 2.3 below, the stricter requirement that \( d < n \) is sometimes imposed in order to avoid computational degeneracy.

On the other hand, some settings conceptually involve data points in a nominal infinite-dimensional space \( \mathcal{X} \). In such cases, it is illuminating and productive to study the theoretical properties of a depth function using its formulation in this nominal space. Nevertheless, practical implementation is carried out for a closely associated data set \( X_n \) in a surrogate finite-dimensional space, typically within \( \mathbb{R}^d \) for some finite \( d \), obtained by suitable truncation and/or discretization of the points \( x \in \mathcal{X} \).

For example, if points \( x \) are represented as \( x = \sum_{j=1}^{\infty} a_j \phi_j \) in terms of a basis \( \{ \phi_j \} \) for \( \mathcal{X} \), then a truncated sum \( \sum_{j=1}^{J} a_j \phi_j \) is used in practice, for finite \( J \) possibly depending upon \( n \). Effectively, \( J \) can be quite small, as illustrated in Chakraborty and Chaudhuri (2014) and Cuevas (2014), for example. Also, if \( x \) is a function \( x(t), t \in T \), then discretized versions of \( x \) and any corresponding basis functions are used, for some \( \{ t_1, \ldots, t_m \} \subset T \) with finite \( m \) possibly depending upon \( n \). Further, for such truncated and discretized data, \( J \) and \( m \) must be sufficiently small that the total number of “grid points” \( nm^d \) remains computationally feasible. Likewise, if \( \mathcal{X} \) is already of finite dimension \( d \) but \( d \geq n \), such techniques of truncation and/or discretization may be carried out to obtain \( d < n \) if so desired.

An overarching assumption for typical depth approach applications is that \( \mathcal{X} \) be a locally compact and locally convex Hausdorff topological vector space over \( \mathbb{R} \). See Appendix A for definition and some typical examples.

For multivariate data, the relevant data space is \( \mathbb{R}^d \) for some finite \( d \geq 2 \). For directional data in \( \mathbb{R}^d \), the data space is a submanifold of lesser dimension.
For functional data (each point a curve), relevant nominal data spaces are infinite-dimensional Banach or Hilbert spaces such as \( L^p(V) \), \( C(V) \), and \( W^{k,p}(V) \), typically for \( V \) a finite interval \([a, b] \). For \( K \)-variate functional data (each point a vector of curves), a product space is used, for example \( \bigotimes_{k=1}^K L^2[a, b] \). Through truncation and/or discretization as described above, such data spaces are replaced in practice effectively by Euclidean spaces, but with each point accompanied by a representation as a curve or \( K \)-vector of curves. Here, then, not only the locations of the points but also the shape features of the curves are to be taken carefully into account, making the depth approach for functional data much more challenging and intricate than mere straightforward extension of that for ordinary multivariate data.

Single-point data sets are precluded in considering the depth approach. In typical stochastic process settings, for example, the data consists of a single realization, such as a time series \( X(t), t \in T \), in \( \mathbb{R}^T \), or a sequence \( \{X_i\} \) in \( \ell^2 \). Of key interest is the almost sure behavior of the process in terms of features such as path regularity, boundary crossings, and convergence. Although such a single data point may have infinitely many components, these do not, however, comprise a set of multiple data points to be described geometrically using the depth approach, which thus does not apply to such settings. Of course, for a data set consisting of multiple realizations of some stochastic process, the depth approach is indeed relevant and productive, as in the setting of functional data.

### 2.2 Considerations on high or infinite dimensionality

In high- or infinite-dimensional data space settings, besides the need to use lower dimensional surrogate spaces for practical representation and handling of data sets, two further “dimensionality” aspects are germane to the depth approach. Firstly, structural features of high- and infinite-dimensional spaces contrast sharply with our experience and thinking regarding low-dimensional Euclidean space. Secondly, geometric features of probability distributions over high- and infinite-dimensional spaces differ markedly from those of probability distributions over low-dimensional Euclidean spaces. Even the formulation of probability distributions can pose issues in the infinite-dimensional case. These closely-related aspects are apropos to our conceptual understanding of the depth approach and its role in applications.

It would be fortuitous if our intuition and understanding of high- and infinite-dimensional settings could be based on experience with low dimensions. In dimension \( d = 2 \) or 3, we are familiar with typical unimodal probability distributions having high probability in the central region and diminishing probability in the tails. Corresponding data sets are concentrated in a central region with relatively few “outliers”. In modeling, we use mixtures of such
unimodal distributions. We describe models in terms of density functions, for which nonparametric estimators include the histogram based on counts of observations in small intervals, squares, or cubes in the univariate, bivariate, and trivariate cases, respectively. In these low dimensions, the bulk of the data in such cells lies within the inscribed circular or spherical regions, respectively.

Unfortunately, this is not an accurate guide to the nature and handling of high- and infinite-dimensional distributions and data. In higher dimension, the central region contains little probability mass or data mass. Samples tend to spread heavily into the tails, leaving most “local” neighborhoods empty. In the infinite-dimensional settings, density functions are not even defined due to lack of a suitable Lebesgue measure. Within a high dimensional hypercube, the inscribed hypersphere occupies a miniscule fraction of the total volume. Indeed, infinite-dimensional hypercubes are unbounded. In Sects. 2.2.1-2.2.3 below, we examine in detail these aspects and their implications regarding the depth approach.

2.2.1 Hypercubes and hyperspheres in high and infinite dimension

Conceptually, perhaps the simplest geometric structure in \( d \)-dimensional space is the unit \( d \)-cube \([0, 1]^d\). In our low-dimensional experience, this is a relatively “small” object in \( \mathbb{R}^d \). No points within it are far apart, relative to the unit side length. As \( d \to \infty \), however, its diameters have length \( \sqrt{d} \) tending to \( \infty \), even though its volume remains fixed at \( 1^d = 1 \). In the limiting case, the infinite-dimensional unit cube remains defined with side length 1 and it is meaningful to note that it has infinite diameters\(^1\), but we cannot mention its “volume” since this is undefined due to lack of a Lebesgue measure on \( \mathbb{R}^\infty \) or \( \ell_2 \).

The differing nature of the unit cube in the finite versus infinite dimensional cases is thus not well characterized by its volume. Also, it is even somewhat misleading to state that its volume remains fixed over finite dimensions \( d \), for in comparing the \( d \)-cube and \((d + 1)\)-cube one must take into account that the \( d \)-cube actually has volume 0 when considered in the \((d + 1)\)-dimensional setting.

More generally, it is misleading to compare volumes of geometric objects across dimensions, although it is a common exercise to do so. In general, a region \( R_1 \) in dimension \( d_1 \) having volume \( v_1 > 0 \) and a region \( R_2 \) in dimension \( d_2 > d_1 \) having volume \( v_2 > 0 \) cannot be compared in terms of \( v_1 \) and \( v_2 \). Rather, they are comparable only within the space of larger dimension, within which the relevant volume of \( R_1 \) is 0. Thus, for example, a square of side 2

---

\(^1\) Nevertheless, this so-called “Tikhonov cube” is topologically equivalent (homeomorphic) to the “Hilbert cube”, the subspace of \( \ell_2 \) consisting of all points \( x = (x_1, x_2, \ldots) \) with \( 0 \leq x_n \leq 1/2^n, \ n = 1, 2, \ldots \), having diameter \( 1/\sqrt{3} \). See van Mill (1988).
in dimension 2 has volume 4 in dimension 2 but volume 0 in dimension 3. It is not meaningful to compare these entities in terms of their volumes within their respective dimensions.

On the other hand, *distances are indeed comparable across dimensions.* Points on the unit circle in $\mathbb{R}^2$ retain the same distance 1 from the origin when considered as points on the unit sphere in $\mathbb{R}^3$. In terms of distances, there is no disconnect in passing from the high-dimensional case to the infinite-dimensional case, despite the disconnect when considering volumes.

Another fashionable exercise is to consider the fraction of volume of the unit $d$-cube occupied by the inscribed hypersphere. In our low-dimensional experience, this fraction is “substantial”, namely 1.00 for $d = 1$, 0.79 for $d = 2$, and 0.52 for $d = 3$ (Scott, 1992). However, this fraction $\rightarrow 0$ quite rapidly as $d \rightarrow \infty$, although the diameter of the inscribed hypersphere remains fixed at 1, even for the infinite-dimensional case.

Consequently, because volumes are not comparable across dimensions while distances are comparable, it is more illuminating to compare *diameters*. The inscribed hypersphere within the unit $d$-cube has *relative diameter* $1/\sqrt{d} \rightarrow 0$ as $d \rightarrow \infty$, and this limit characterizes the infinite-dimensional case as well, by virtue of $1/\infty = 0$. Thus, as $d \rightarrow \infty$, although the inscribed hypersphere within the unit cube does not vanish, its relative “weight” becomes zero by virtue of the growth of the surrounding unit cube.

For high and infinite dimension, the bulk of the unit cube is found in its “corners”, these becoming both increasingly large and increasingly numerous, comprising virtually the entire cube! This somewhat clarifies the comparative natures of hypercubes and hyperspheres in high and infinite dimension.

### 2.2.2 Geometric description of data sets in higher dimension

For geometric description of data sets by grouping using histograms, it is clear from the above discussion that cubes of common side length $\varepsilon$ are not suitable in the high- and infinite-dimensional cases. The reason is not because the inscribed hypersphere has negligible weight, but rather because the extent of the cube (length of diameters) is incommensurate with the small common length of the sides. *Classification into a cell with a very large possible distance between its points fails to properly "locate" a data point.*

Further, the computational burden of such histograms becomes prohibitive. In the case of a large finite dimension $d$, for $x$ given by $(x_1, \ldots, x_d)$ let the $i$th coordinate variable $x_i$ be grouped within a grid of $m_i$ sub-intervals each of width $\varepsilon$ over the interval $[t_i, t_i^*]$, namely with endpoints $t_i, t_i + \varepsilon, \ldots, t_i + m_i \varepsilon = t_i^*$, for $1 \leq i \leq d$. Then there are $m_1 \times \ldots \times m_d$ such cells in the “product grid”. For reasonable data binning, the cells should number $\leq \sqrt{n}$, say, but this requires $d \leq (\log n)/(2 \log m)$, forcing $n$ and the computational burden to
become prohibitively large.

One might consider constructing a “Hilbertian histogram” based on small Hilbert cubes (defined in Footnote 1). Here the $i$th coordinate variable $x_i$ is grouped within $m_i^*$ sub-intervals each of width $a/2^i$ over the interval $[t_i, t_i^*]$, namely with endpoints $t_i, t_i + a/2^i, \ldots, t_i + m_i^*a/2^i = t_i^*$, for $1 \leq i \leq d$. Then, in the “product grid”, each cell has $d$ sides of lengths $a/2, \ldots, a/2^d$, respectively, and now each cell has small common diameter $a/\sqrt{3}$, but there are now $m_1^* \times \ldots \times m_d^*$ such cells, even greater than the above number $m_1 \times \ldots \times m_d$. Thus this approach also fails.

In summary, in very high dimension $d$, points which are very “close” coordinate-wise, in terms of belonging to the same $d$-cube of side $\varepsilon$, can in fact be very distant from each other due to the large diameters of the cube. This defeats the histogram for practical purposes by rendering it unable to account effectively for the whereabouts of data points. This indeterminacy reflects the so-called “curse of dimensionality”, the property that points in a space can more easily be extremely spread out as the dimensionality increases. Unfortunately, this “sparsity” issue also impacts the depth approach, since it, even more than the histogram, is based on relative locations of data points.

### 2.2.3 Probability distributions in high and infinite dimension

The representation of a probability distribution by a density function $f(x)$ and of local probability mass by “$f(x)dx$” fails in infinite dimension, because, as noted above, the corresponding infinitesimal hypercubes “$dx$” have infinite diameters and no longer specify locality. Thus histograms and “geometric” description are no longer relevant. Rather, probability distributions must be formulated and described by analytical means, for example using characteristic functions.

In particular, even the simple uniform distribution on the unit $d$-cube has very different geometric properties in the high- and infinite-dimensional cases, as follows from the discussion in Sect. 2.2.1. And if this simple distribution can thus challenge geometric description in the high- and infinite-dimensional case, this must hold even more so for any other probability distribution. Indeed, even for the familiar normal distribution, this change of properties becomes quickly evident as $d$ increases. The bulk of the univariate normal distribution is in the central region with very little probability in the tails, whereas this reverses as the dimension increases. For example (Scott, 1992), the percentage of mass within the contour defined by density equal to 0.01 diminishes rapidly from 99.8% for $d = 1$ and 99% for $d = 2$ to 49% as soon as $d = 10$ and to 2% already for $d = 20$.

Description, modeling, and inference with high dimensional distributions should take into account the above-mentioned “heavy tailed” feature. Yet our
experience in the univariate case offers little guidance on how to think properly about high dimensional descriptive measures, even when formulated as natural extensions of familiar univariate versions that (we believe) we understand. How might we think of “kurtosis” for distributions consisting mostly of heavy tails? Especially, then, how might we think of “depth” in higher dimension, when all of the probability is in the tails and all of the data is “outlying”? The depth approach readily identifies a “center” but cannot be expected to inform as successfully about geometric structure that is either mostly in the tails (large \(d < \infty\)) or inherently undefined \((d = \infty)\).

Consequently, in conceiving, developing, and applying the depth approach, it is perhaps best, therefore, not to require absolute consistency between its features and performance in lower dimensional settings and in high- or infinite-dimensional settings, which differ so greatly in geometric characteristics. In this regard, performance of the depth approach in the lower dimensional case has a determining priority. This is not too generous or too permissive, as data in \(\mathbb{R}^d\) with high \(d\) often has structure of lower dimension, which can be found approximately by dimension reduction techniques such as projection pursuit.

The foregoing discussion has implicitly concerned probability distributions on data defined as random elements of \(\mathbb{R}^d\) for \(d \leq \infty\) with no \textit{a priori} order imposed on the components, any particular ordering of them being equally acceptable. If, on the other hand, coordinates meaningfully follow a single specified order, as with a stochastic process unfolding in time, then each point is endowed with an associated structure (a curve in a suitable curve space) and this should be taken into account in describing such a data set. Accordingly, probability distributions may be considered on a relevant function space and described in terms of probabilities of properties that possibly are satisfied by a sample curve, such as convergence, path smoothness, level crossings, etc. Also, in this situation, regardless of dimension \(d\), data points may be visualized on a two-dimensional plot. In this particular (nominally) infinite-dimensional setting, the depth approach and notions of centrality and outlyingness may be formulated quite usefully and constructively, with “unusual” or “outlying” data points being those whose curves differ from the bulk of the curves in either location or shape (or both). Of course, as noted earlier, in practice the dimension becomes finite.

2.3 Sample depth functions versus population depth functions

The depth approach for description of the basic geometric structure of a data set \(X_n\) in a space \(\mathcal{X}\) employs a depth \(D(\cdot, X_n)\) implemented \textit{algorithmically} on \(X_n\). The computational burden increases with both the sample size \(n\) of \(X_n\) and the dimension \(d\) of \(\mathcal{X}\), and, inevitably, practical limits hold even with the best
algorithms that can be developed using computational geometry. This leads to the use of heuristic approximations to the actual $D(\cdot, X_n)$, and thus in practice a sample depth function $D(\cdot, X_n)$ might have various possible implementations.

A further technical consideration concerns the relative sizes of $n$ and $d$. Note that if $X_n$ is a sample from a continuous probability distribution $P$ on $\mathcal{X}$, and $n < d$, then the convex hull $\text{ch}(X_n)$ of the data set is a lower-dimensional subset of $\mathcal{X}$ with $P$-probability 0. In this case, when also the depth $D(\cdot, X_n) = 0$ outside $\text{ch}(X_n)$ as holds for some depths, it follows that the sample depth $D(\cdot, X_n)$ can be nonzero only on a set of $P$-probability 0, rendering it of limited or only specialized practical use. One resolution when $n < d$ is to "smooth" the depth function $D(\cdot, X_n)$ to be nonzero everywhere in $\mathcal{X}$. Or, one can discretize the nominal data points in $\mathcal{X}$ to points in a selected space of dimension $d < n$. Another motivation for achieving $n > d$ is that, typically, asymptotic results developed for $n \to \infty$ are applicable in practice only when the given $n$ is greater than $d$. Although we do not strictly require $d < n$, this condition is desirable and productive when convenient. We thus assert:

**Assertion 2.2.** The depth approach applies in practice to data sets $X_n$ of finite size $n > 1$ in some data space $\mathcal{X}$ of dimension $d < \infty$, with $d < n$ preferred whenever feasible.

The intrinsic nature of a given depth function $D(\cdot, X_n)$ can be obscured by the associated computational details and heuristics. In order to explore and understand a depth function conceptually and with greater simplicity, it is helpful to introduce an associated “population” version $D(x, P)$, $x \in \mathcal{X}$, defined for $P$ any probability distribution on $\mathcal{X}$. Here $D(x, P)$ measures the depth of $x$ relative to $P$. A sample version $D(\cdot, X_n)$ is recovered by taking $P = \hat{P}_n$, the empirical distribution placing probability $n^{-1}$ on each observation in $X_n$, or alternatively a smoothed version of $\hat{P}_n$.

Despite their close association, there are key differences in how sample and population depth functions should be viewed and used in their respective contexts of a data set $X_n$ in $\mathcal{X}$ versus a distribution $P$ on $\mathcal{X}$. While one can formulate and study $D(\cdot, P)$ over the very broad scope of all possible distributions $P$ on $\mathcal{X}$, the study of the associated sample version $D(\cdot, X_n)$ concerns only the simpler and narrower scope of the empirical distributions $\hat{P}_n$ on $\mathcal{X}$. This opens the possibility that a particular depth function $D(\cdot, P)$ might exhibit anomalous behavior for certain choices of $P$ on certain $\mathcal{X}$ and yet be well-behaved for empirical distributions $\hat{P}_n$. Should such a finding for $D(\cdot, P)$ imply a concern with the practical use of the sample version $D(\cdot, X_n)$? In some

---

2It is implicitly understood throughout that all probability distributions $P$ “on $\mathcal{X}$” are defined over some fixed $\sigma$-algebra $\mathcal{A}$ of sets in $\mathcal{X}$, for example the usual Borel sets when $\mathcal{X} = \mathbb{R}^d$. When necessary, this will be made explicit.
cases the anomalous behavior for a depth \( D(\cdot, P) \) is seen only for \( P \) having geometric features that are somewhat ambiguous relative to formulation of notions of “depth”. We thus assert:

**Assertion 2.3.** A population version \( D(\cdot, P) \) should not be considered a guide regarding the practical merits of the data-based version \( D(\cdot, X_n) \).

Indeed, certain classes of \( P \) may be disregarded for study of \( D(x, P) \), as seen in Sect. 2.4 below. The above assertion is formalized more fully in Sect. 2.5.

### 2.4 Considerations on relevant probability distributions for the formulation of population depth functions

The goal in considering population depths \( D(\cdot, P) \) relative to distributions \( P \) on a space \( \mathcal{X} \) is insight into data-based versions \( D(\cdot, X_n) \) apart from algorithmic details. Due to the breadth of possibilities for \( P \), however, there can be found particular examples of space \( \mathcal{X} \), depth function \( D(\cdot, \cdot) \), and distribution \( P \) on \( \mathcal{X} \) for which the associated population depth \( D(\cdot, P) \) exhibits behavior quite inconsistent with the goals of the depth approach, for example a property such as \( D(X, P) = 0 \) with \( P \)-probability 1. The relevance of such a finding depends upon whether the particular \( P \) in question is or is not itself anomalous in some sense, with respect to potential formulation of notions of centrality. Let us explore, therefore, criteria for a probability distribution \( P \) on a space \( \mathcal{X} \) to be amenable to formulation of depth functions on \( \mathcal{X} \). This hinges on geometric features of the support\(^3\) \( \text{supp} P \), which we now discuss. It turns out that, for any given \( \mathcal{X} \), there are two classes of distribution \( P \) on \( \mathcal{X} \) which we choose to set aside as unhelpful for study of centrality notions relative to \( P \) on \( \mathcal{X} \).

The first class consists of non-atomic distributions \( P \) with support a proper subset of \( \mathcal{X} \) with empty interior,

\[
\mathcal{P}_1(\mathcal{X}) = \{ P \text{ on } \mathcal{X} : P \text{ is non-atomic, } \text{supp } P \neq \mathcal{X}, (\text{supp } P)^\circ = \emptyset \}.
\]

Since the assumption \( \text{supp } P \neq \mathcal{X} \) implies that also \( \partial(\text{supp } P) \neq \mathcal{X} \), it follows by Corollary A.1(b) that, for \( P \in \mathcal{P}_1(\mathcal{X}) \), \( \text{supp } P \) is nowhere dense in \( \mathcal{X} \). That is, the distribution \( P \) lives on a proper subspace \( \mathcal{Y} \) of \( \mathcal{X} \) with empty interior. For example, \( \mathcal{Y} \) might be a line in \( \mathcal{Y} \) or the surface of a ball in \( \mathcal{X} \). For such cases, the depth approach relative to \( P \) is more properly formulated with \( P \) restricted to the space \( \mathcal{Y} \), with a depth function \( D(y, P) \), \( y \in \mathcal{Y} \), measuring

\(^3\)The support \( \text{supp} P \) of a probability distribution \( P \) on a topological or normed space \( \mathcal{X} \) is the closed set \( S \) of points \( x \) with \( P(U) > 0 \) for each open \( U \) containing \( x \), where \( P(S^c) = 0 \). Equivalently, \( \text{supp} P \) is the largest set \( S \subseteq \mathcal{X} \) such that \( P(U \cap S) > 0 \) for all open sets \( U \) with \( U \cap S \neq \emptyset \). For any \( \mathcal{X} \), \( \text{supp} P \) may be taken as \( \{ x \in \mathcal{X} : P(\{x\}) > 0 \} \) if \( P \) is discrete, and as the closure of \( \{ x \in \mathcal{X} : p(x) > 0 \} \) if \( P \) is continuous with density \( p(\cdot) \), or as the union of these for a mixture distribution.
depth within \( \mathcal{Y} \), not within \( \mathcal{X} \). In some such cases, one can transform \( \mathcal{Y} \) to a more convenient equivalent space \( \mathcal{Y}' \). The assumption of non-atomic \( P \) means that \( P_1(\mathcal{X}) \) does not include discrete distributions, even though these too have support nowhere dense in \( \mathcal{X} \).

In order to define the second class of \( P \) that we choose to set aside, we need a more refined notion of support of \( P \), which we now introduce.

**Definition 2.1 (Essential Support).** For a probability distribution \( P \) on a topological or normed space \( \mathcal{X} \), an essential support \( \text{esupp} \ P \) is any subset \( E \) of \( \text{supp} \ P \) for which \( P(E) = 1 \) and \( P(E^c) = 0 \), that is, for which \( P(X \in E) = 1 \) and \( P(X \in E^c) = 0 \) for \( X \) having distribution \( P \). □

Note that \( \overline{\text{esupp} P} = \text{supp} P \) always holds. Clearly, if \( P \) is discrete, then the only \( \text{esupp} \ P \) is \( \text{supp} P \) itself.

Let us now consider the case that \( P \) places probability 1 on a proper subset \( E \) of \( \mathcal{X} \) that is dense in \( \mathcal{X} \). Then \( E = \text{esupp} P \) with both \( E \) and \( E^c \) nonempty and with \( \partial E = \partial(E^c) = \mathcal{X} \). In this case, \( \overline{E} = \text{supp} P = \mathcal{X} \), which however masks the information about the geometric features of \( P \) as a distribution on \( \mathcal{X} \) given by the fact that \( E = \text{esupp} P \). By Corollary B.1(a) in Appendix B, we have in the present case that both \( E \) and \( E^c \) have nonempty interior and are dense in \( \mathcal{X} \). Then every locale or region within \( \mathcal{X} \) is associated equally with each of \( E \) and \( E^c \), in the geometric sense. Thus the geometric structure of \( \text{esupp} P \) coincides with that of \( (\text{esupp} P)^c \). This somewhat obscures the idea of a defining geometric structure attached to the distribution \( P \) and, therefore, somewhat confounds the formulation of a geometric notion of centrality relative to the distribution \( P \) on \( \mathcal{X} \). To avoid such ambiguity when studying population depth functions on \( \mathcal{X} \), we choose to bypass such \( P \). Applying Corollary B.1(a), the class to be bypassed may be represented as

\[
\mathcal{P}_2(\mathcal{X}) = \{ P : \exists E = \text{esupp} P \text{ satisfying } E \neq \partial E \text{ and } \partial E = \mathcal{X} \}.
\]

Regarding the study of population depth functions, we thus assert:

**Assertion 2.4.** Properties desired for a population depth function \( D(\cdot, P) \) need not be satisfied for \( P \in \mathcal{P}_1(\mathcal{X}) \cup \mathcal{P}_2(\mathcal{X}) \).

Consequently, for study of any population depth function \( D(\cdot, P) \) defined on \( \mathcal{X} \), we confine to distributions \( P \) on \( \mathcal{X} \) not in either \( \mathcal{P}_1(\mathcal{X}) \) or \( \mathcal{P}_2(\mathcal{X}) \). This class, denoted by \( \mathcal{P}_0(\mathcal{X}) \), conveniently includes all empirical distributions \( \hat{P}_n \) associated with potential data sets \( \mathcal{X}_n \). Although \( \mathcal{P}_0(\mathcal{X}) \) most likely includes some further \( P \) of questionable appeal for defining depth functions, we leave these to be specified only should they arise in a particular context. For \( \mathcal{X} = \ell_2 \), examples of \( P \) in \( \mathcal{P}_0(\mathcal{X}) \), \( \mathcal{P}_1(\mathcal{X}) \), and \( \mathcal{P}_2(\mathcal{X}) \), respectively, are provided in Appendix C.
2.5 Evaluation principle for depth functions

In evaluating and comparing depth functions, three relevant considerations are as follows.

(a) As noted earlier, implementation of $D(\cdot, X_n)$ is *algorithmic*. Therefore, trust in $D(\cdot, X_n)$ is based on its performance characteristics and on its technical properties, relative to various criteria. Regarding technical properties, it is conceptually helpful to formulate and study these also for the version $D(\cdot, P)$, as the inherent defining entity that underlies the various possible algorithmic implementations of $D(\cdot, X_n)$. While one can also learn from visualizing $D(\cdot, X_n)$ with bivariate data, this case is not truly representative of the setting of high-dimensional data (as discussed in Sect. 2.2 above). Overall, the evaluation and comparison of competing depth functions $D(\cdot, \cdot)$ entails the *performance properties* of the sample version $D(\cdot, X_n)$ and the *technical properties* of the population version $D(\cdot, P)$.

(b) For the fullest understanding of a population depth $D(\cdot, P)$, one studies it in as broad a setting as possible. Here one allows various possible infinite-dimensional $X$. Nevertheless, any consideration of $D(\cdot, P)$ on an infinite-dimensional version $D(\cdot, X_n)$, because the data set $X_n$ itself in such a case is derived through some discretization of nominal data points in the infinite-dimensional $X$. While the scope of $D(\cdot, P)$ may include infinite-dimensional $X$, the sample versions $D(\cdot, X_n)$ are applied only in finite-dimensional settings.

(c) In view of (b), technical issues that might arise with a depth $D(\cdot, P)$ for certain $P$ on certain infinite-dimensional $X$ do not necessarily imply any practical limitations regarding the use of the data-based version $D(\cdot, X_n)$ in finite-dimensional settings. This must be evaluated judiciously.

The preceding considerations yield a methodological principle regarding the question, *does a technical shortcoming of $D(\cdot, P)$ in some sense necessarily obviate the practical use of $D(\cdot, X_n)$?* We assert

**Assertion 2.5. Evaluation Principle.** The practical merits of any depth function should be evaluated solely on the basis of the performance of its sample version $D(\cdot, X_n)$ as a statistical procedure. The relevance of its population version $D(\cdot, P)$ is only for insights.

2.6 Considerations on relevant inference settings

Besides its descriptive role, a depth function might also be used for inference relative to an underlying statistical model. Here key considerations are as follows.
(a) The formulation of \( D(\cdot, X_n) \) as \( D(\cdot, \hat{P}_n) \) does not require \( X_n \) to be a sample from some unknown distribution \( P \) on \( \mathcal{X} \). It is \( X_n \)-based but not model-based. Were also a model assumed, however, then estimating \( P \) by the usual empirical distribution \( \hat{P}_n \) is tantamount to adopting the broadest possible statistical inference model, which states simply that \( P \) belongs to the class \( \mathcal{P}_0(\mathcal{X}) \). This model poses no restrictions on the data \( X_n \). Therefore, we assert:

**Assertion 2.6.** Without loss of generality, the “depth approach” may be viewed as a statistical procedure within the broadest nonparametric inference setting, the model \( \mathcal{P}_0(\mathcal{X}) \).

In this broadest nonparametric inference setting based on \( \mathcal{P}_0(\mathcal{X}) \), the sample depth function \( D(\cdot, X_n) \) is computed without explicit reference to the underlying model.

(b) A population depth function \( D(\cdot, P) \) constitutes an important parameter of \( P \), as such of interest not only in itself, but also for developing useful \( D(\cdot, P) \)-based tools for inference about features of \( P \). Note that the parameter \( D(\cdot, P) \) need not determine \( P \), a property that holds in some cases but is purely incidental to practical application.

(c) From a very general perspective, the depth approach may be formulated and applied in the setting of any statistical inference model \( \{(\mathcal{X}, P), P \in \mathcal{P}\} \) given by specifying some subclass \( \mathcal{P} \) of \( \mathcal{P}_0(\mathcal{X}) \). Here \( P \) should be estimated not by the usual empirical distribution, but rather by an optimal estimator \( \hat{P}_n(\mathcal{P}) \) that takes account of the specified family \( \mathcal{P} \). In such a setting, there are two provisos:

- Conceptually, the population version \( D(\cdot, P) \) must be formulated as a fully nonparametric functional, thus one which is meaningful for all \( P \in \mathcal{P}_0(\mathcal{X}) \), the broadest nonparametric model.

- The appropriate sample version \( D(\cdot, X_n) \), however, is no longer \( D(\cdot, \hat{P}_n(\mathcal{P}_0)) \), but rather is given by \( D(\cdot, \hat{P}_n(\mathcal{P})) \), which is not only \( X_n \)-based but also “\( \mathcal{P} \)-adapted”.

Thus we assert

**Assertion 2.7.** The “depth approach” is very general. While always formulated for \( P \in \mathcal{P}_0(\mathcal{X}) \), it is in principle applicable relative to any chosen inference setting – nonparametric, semiparametric, or parametric – through specification of the subfamily \( \mathcal{P} \).

The development and treatment of the depth approach in subfamily settings \( \mathcal{P} \) is similar to that for the broad setting \( P \in \mathcal{P}_0(\mathcal{X}) \), except for the possibly challenging step of deriving the associated \( \mathcal{P} \)-adapted estimator \( \hat{P}_n(\mathcal{P}) \) of \( P \).
2.7 Depth functions versus density functions

2.7.1 General considerations

Nonparametric density estimation is another key aspect of exploration of a data set and compares with the depth approach as follows. Whereas depth functions are formulated from a global orientation reflecting degree of centrality or outlyingness, density functions reflect local probability mass. A maximal depth point represents a center in the sense of generalizing the univariate median, whereas maximal density points or modes do not necessarily represent central locations and multimodality can be present as well. Thus density and depth functions serve complementary goals quite different in nature.

For unimodal finite-dimensional distributions \( P \) satisfying certain notions of symmetry, the family of depth function contours coincides with that of the density function contours, an appealing circumstance. The corresponding labelings have different meanings and interpretations, however, reflecting local and global perspectives, respectively.

In multimodal settings, however, the density contours are nonconvex, while typically the depth contours are convex or nearly convex. The center that is determined by the depth contours can even lie well afield of the region of relatively high probability mass, although still within the convex hull of this region. Although in the univariate bimodal case the usual median can lie in a region of very low probability mass and such a feature is generally accepted, its acceptability in the multivariate setting has been questioned by some. For example, some have proposed considering the density function itself as a depth function (see Sect. 2.7.2). Also, certain “local” versions of depth have been formulated whose contours mimic those of the density function (see Sect. 2.7.3).

Thus, in developing the depth approach, one must address the question, should depth function contours necessarily follow or mimic density function contours? Our answer is “No”. It is unproductive to make a depth function perform like a density function, since then nothing new is accomplished. In fact, this would compromise the role of the depth function, which under any circumstances is to assess centrality of data points from a bona fide notion of centrality.

Two extremal cases add further perspective. For a uniform distribution in any finite dimension, every point is a mode and the density function is constant with no density contours at all, whereas typical depth functions have distinctive convex contours that enclose a well-defined center. Also, in typical

---

4When “density” is mentioned, it is assumed that the data space \( \mathcal{X} \) possesses the so-called Radon-Nikodym property and that the “density” of a probability distribution \( P \) on \( \mathcal{X} \) denotes a version of the Radon-Nikodym derivative of \( P \) with respect to a designated \( \sigma \)-finite dominating measure.
infinite-dimensional settings, density functions are not even defined, whereas depth functions can still be formulated and, minimally, determine a center.

We thus assert

**Assertion 2.8.** Depth function contours need not mimic density contours, nor must the maximal depth point lie within a region of high probability mass.

Depth functions complement density functions by adding information of a different nature. The two functions are conceptually quite different and have distinctly different roles, neither of which should be suppressed or conscripted to substitute for the other. Density functions exhibit probability structure as well as multimodality structure when present, whereas depth functions provide center-outward ordering of points along with connected level sets. The most powerful approach is to use these functions individually, hand in hand.

We elaborate the above considerations in Sects. 2.7.2-2.7.3 below. The related question of whether a given family of contours generates a proper depth function is treated in Serfling (2019).

### 2.7.2 Nonvalidity of density function as depth function

It might be tempting to let the density function itself serve as a depth function. The most compelling case for this would be that of a unimodal density, for which the density values may also be interpreted as measuring depth, because as the density at a point decreases so does its degree of centrality. Indeed, a unimodal density function fulfills defining properties for depth functions with the mode as the center: decreasing along rays out from the center, invariance, nondegeneracy, and nullness at infinity (Properties (M), (I), (Z), and (N) in Serfling, 2019). On this basis, a unimodal density function might technically serve an additional role as a depth function. However, this is only within the restricted setting of $P$ with unimodal densities rather than in the unrestricted nonparametric setting of $P \in \mathcal{P}_0(\mathcal{X})$ as discussed above in Sect. 2.6. That is, such a “depth function” fails to constitute a nonparametric functional $D(\cdot, P)$ defined over all of $\mathcal{P}_0(\mathcal{X})$. From a practical standpoint, the assumption of unimodal $P$ is rather restrictive as a nonparametric setting and yet is quite too broad for parametric modeling. We assert:

**Assertion 2.9.** Density functions should not masquerade as depth functions.

Let us emphasize this by viewing density and depth results for a bivariate uniform distribution in Figure 1 and for bivariate standard normal in Figure 2. These illustrate how depth and density functions are different conceptually and thus complement each other, sometimes with similar appearance, sometimes not.
Figure 1: Plots for distribution $P = \text{Uniform}([0, 1]^2)$. Upper left: population halfspace depth contours. For comparison, the density function is constant with no contours. Upper right: a sample of size 100 from $P$. Lower: sample halfspace depth contours (left) and sample nonparametric density estimator contours (right). (Sample density via \texttt{kde} in \texttt{ks}, sample halfspace depth via \texttt{perspdepth} in \texttt{depth}, in \texttt{R}.)
In general, depth concerns *centrality* from a *global* perspective, whereas density concerns *local probability mass*. Depth functions *ignore nonconvexity and multimodality*, leaving these to be captured by the density function.

### 2.7.3 Considerations on “local” depth functions

Typical depth functions on $\mathbb{R}^d$ have convex or nearly convex contours even if the population or sample exhibits nonconvex support and/or multimodality. Some researchers argue, as above, that density and depth functions can and should have complementary roles. Some others, who want the depth contours to follow the density contours, have developed notions of local depth that possess this feature. See Chen, Bart, Dang, and Peng (2007), Agostinelli and Romanazzi (2011), and Paindaveine and Van Bever (2013).
Let us gain some perspective on this issue by looking at the halfspace and local depths for the following mixture of two bivariate normal distributions,

$$0.5 \times \text{Normal}\left(\left(\frac{3}{0}\right), \left(\frac{2}{0.5}\right)\right) + 0.5 \times \text{Normal}\left(\left(\frac{0}{3}\right), \left(\frac{0.5}{0.2}\right)\right),$$

which places high probability on a region of “lazy L” shape, a quite representative illustration. From Figure 3 it is clear that density and depth contours offer complementary pieces of information, each appealing in its own way, and that the local depth contours cannot substitute for either.

Figure 3: Sample of size 100 from mixture of two standard normals (upper left), sample nonparametric density contours (upper right), sample halfspace depth contours (lower left), and sample local depth contours (lower right). (Sample density via \texttt{kde} in \texttt{ks}, sample halfspace depth via \texttt{perspdepth} in \texttt{depth}, sample local depth via \texttt{depthLocal} in \texttt{DepthProc}, in R.)

We assert:

**Assertion 2.10.** Density and depth functions should be used in tandem rather than merged into a single function.
3 Concluding remarks

General perspectives on the depth approach as in Sect. 2 are intended to be instrumental, even if some of these are controversial. One purpose is to offer some foundational viewpoints relative to the continuing development of the depth approach. Another is to provide some clarifications regarding notions of “density depth” and “local depth”.

Another application (Serfling, 2019) is to better interpret the findings of Dutta, Ghosh and Chaudhuri (2011) such as the anomalous behavior of the population Tukey depth for certain $P$ on certain infinite-dimensional $\mathcal{X}$, which also holds for the population projection depth. Since these classical depths have straightforward formulations in quite general data settings, as well as incarnations in the setting of depth functions on parameter spaces (see Rousseeuw and Hubert, 1999, and Mizera, 2002, for use of the halfspace notion, and Zuo, 2018a,b and 2019, for a broad treatment including use of the projection notion), their potential roles are broad and it is productive to deepen our understanding of them.

Acknowledgements

The author gratefully acknowledges useful input from G. L. Thompson.

A Some common sample spaces

Here the data space $\mathcal{X}$ is assumed to be a vector space. If equipped with a topology, $\mathcal{X}$ is called a topological vector space. In particular, here we suppose $\mathcal{X}$ to be a locally compact and locally convex Hausdorff space\(^5\). Some typical examples are as follows.

$\mathbb{R}^d$ Euclidean space of dimension $d \geq 1$: vectors $\vec{x} = (x_1, \ldots, x_d)$, with norm $\|\vec{x}\| = (\sum_{i=1}^{d} x_i^2)^{1/2}$.

$L_p(V, \mu)$ Space of $p$th power $\mu$-integrable functions on compact $V \subset \mathbb{R}^d$, with norm $\|f\|_p = (\int |f(x)|^p d\mu(x))^{1/p}$, for $1 \leq p < \infty$.

$L_\infty(V, \mu)$ Space of essentially bounded functions, with $\|f\|_\infty = \text{ess sup}_x |f(x)|$ with respect to measure $\mu$.

\(^5\)A Hausdorff space is a topological space for which any distinct points can be separated by disjoint neighborhoods. It is locally convex if any neighborhood of the zero element contains a convex neighborhood of the zero element. It is compact if closed and bounded and, if infinite-dimensional, also flat: for every $\varepsilon > 0$, contained in the $\varepsilon$-neighborhood of a finite-dimensional space.
Space of sequences $x = (x_1, x_2, \ldots)$ in $\mathbb{R}$, with $\|x\|_p = \left(\sum |x_i|^p\right)^{1/p} < \infty$ for $1 \leq p < \infty$, and sequences with $\|x\|_\infty = \sup |x_i| < \infty$ for $p = \infty$.

c₀ Space of sequences $x = (x_1, x_2, \ldots)$ in $\mathbb{R}$ with $\lim_{n \to \infty} x_n = 0$ and $\|x\| = \sup |x_i|$.

$C(V)$ Space of continuous functions on compact $V \subset \mathbb{R}$, with norm $\|f(x)\| = \sup |f(x)|$ or with inner product $\langle f, g \rangle = \int f g$.

$C_b(V)$ Space of bounded and continuous functions on $V \subset \mathbb{R}$.

$C_\infty(V)$ Space of functions on $V \subset \mathbb{R}$ with continuous derivatives of all orders.

$W^{k,p}(V, \mu)$ Sobolev space of functions whose weak derivatives $f^{(j)}$ of order $j \leq k$ belong to $L_p(V, \mu)$, with $\|f\|_{k,p} = \left(\sum_{j=0}^{k} \int f^{(j)}(x)d\mu(x)\right)^{1/p}$.

### B Some results on dense and nowhere dense sets

Some useful facts concerning dense and nowhere dense sets are given in the following special lemma. (For convenient reference, we recall that $\mathcal{X}^o = \mathcal{X}$ and $\partial \mathcal{X} = \partial \emptyset = \emptyset$.)

**Lemma B.1.** Let $\mathcal{X}$ be a normed or topological space and let $E$ be a nonempty proper subset of $\mathcal{X}$.

(a) $\partial E = \mathcal{X}$ if and only if both $E$ and $E^c$ are dense in $\mathcal{X}$.

(b) $\overline{E}^o = \emptyset$ implies $E^o = \emptyset$. Conversely, $E^o = \emptyset$ implies $\overline{E}^o = \emptyset$ unless $\overline{E} = \mathcal{X}$.

**Proof.** (a) (i) Suppose $\partial E = \mathcal{X}$. Then $\overline{E} = \mathcal{X}$ and so $E$ is dense in $\mathcal{X}$. Since $E^c$ is also nonempty, then $\partial E = \partial (E^c)$ and so $E^c$ also is dense in $\mathcal{X}$. (ii) Suppose that both $E$ and $E^c$ are dense in $\mathcal{X}$. Then, for every $x \in \mathcal{X}$, every open neighborhood of $x$ contains points of both $E$ and $E^c$, and thus every $x \in \mathcal{X}$ belongs to $\partial E = \partial (E^c)$, yielding $\mathcal{X} = \partial E = \partial (E^c)$.

(b) The first statement holds trivially. For the converse, suppose $E^o = \emptyset$. Then $E \subset \partial E = \partial \overline{E}$ and so $\overline{E} \subset \partial E$. Then $\overline{E}^c \subset \partial E = \partial \overline{E}$. Now suppose that $\overline{E}^c \neq \emptyset$ and let $x \in \overline{E}^c$. Then $\exists$ a neighborhood $N_\varepsilon(x) \subset \overline{E}^c$ with $N_\varepsilon(x) \cap \overline{E} = \emptyset$ and $N_\varepsilon(x) \subset \partial \overline{E}$. However, the latter inclusion implies, contradictorily, that $N_\varepsilon(x) \cap \overline{E}^c \neq \emptyset$, unless $\overline{E}^c = \emptyset$, which is ruled out by the assumption that $\overline{E} \neq \mathcal{X}$. We thus indeed have a contradiction to the supposition that $\overline{E}^c \neq \emptyset$. \qedsymbol
The following corollary of Lemma B.1 is immediate and provides a useful classification of the possibilities regarding whether or not a set $E$ is dense or nowhere dense in $\mathcal{X}$.

**Corollary B.1.** Let $\mathcal{X}$ be a normed or topological space and let $E$ be a nonempty proper subset of $\mathcal{X}$.

(a) If $\partial E = \mathcal{X}$, then $E^\circ = \emptyset$, $(E^c)^\circ = \emptyset$, and both $E$ and $E^c$ are dense in $\mathcal{X}$.

(b) If $\partial E \neq \mathcal{X}$ and $E^\circ = \emptyset$, then $E \neq \mathcal{X}$ and $E$ is nowhere dense in $\mathcal{X}$.

(c) If $\partial E \neq \mathcal{X}$ and $E^\circ \neq \emptyset$, then $E$ is not dense in $\mathcal{X}$ nor is $E$ nowhere dense in $\mathcal{X}$.

### C  Complement to Section 2.4

For $\mathcal{X} = \ell_2$, we provide examples of $P$ for each of the classes $\mathcal{P}_0(\mathcal{X})$, $\mathcal{P}_1(\mathcal{X})$, and $\mathcal{P}_2(\mathcal{X})$ formulated in Section 2.4.

**Example C.1.** Let $\mathcal{X} = \ell_2$ with the usual norm $\|x\| = (\sum_{i=1}^{\infty} x_i^2)^{1/2}$ for $x = (x_1, x_2, \ldots)$ and the usual Borel $\sigma$-field. Let $P$ be the uniform distribution on the ball of radius $r$ centered at 0, $B(0, r) = \{x \in \mathcal{X} : \|x\| \leq r\}$. Clearly, this $P$ belongs to $\mathcal{P}_0(\ell_2)$.

**Example C.2.** Let $\mathcal{X} = \ell_2$ and let $P$ be uniform on the sphere of radius $r$, $S(0, r) = \{x \in \mathcal{X} : \|x\| = r\}$. Clearly, this $P$ belongs to $\mathcal{P}_1(\ell_2)$.

For the next example, the following preliminary probability lemma is used.

**Lemma C.1.** Let $\{X_i\}$ be independent real-valued random variables with $EX_i \equiv 0$ and $EX_i^2 \equiv 1$.

(a) If $\inf_{i \geq 1} i P(|X_i| > \varepsilon) > 0$ for some $\varepsilon > 0$, then $P(\sum_{i \geq 1} X_i^2 = \infty) = 1$.

(b) If $\sum_{i \geq 1} i^{-2} EX_i^4 < \infty$, then $P(\sum_{i \geq 1} X_i^2 = \infty) = 1$.

**Proof.** (a) The assumed condition implies that $\sum_{i \geq 1} P(X_i^2 > \varepsilon^2) \geq \Delta \sum i^{-1} = \infty$, for some $\Delta > 0$. By the Kolmogorov Three Series Theorem [4, p. 117, Thm. 2], it follows that $\sum_{i \geq 1} X_i^2$ converges (finite) with probability $\neq 1$. Then, by the Kolmogorov 0-1 Law [4, p. 64, Thm. 3], this event has probability zero.

(b) It suffices to show that $n^{-1} \sum_{i \geq 1} X_i^2 \to 1$ with probability 1. Put $Z_i = X_i^2 - 1$. Then $EZ_i = 0$ and $EZ_i^2 = E(X_i^2 - 1)^2 = EX_i^4 - 1 \leq EX_i^4$. Then the assumed condition implies that $i^{-2} \sum EX_i^4 < \infty$, which yields $n^{-1} \sum_{i \geq 1} Z_i \to 0$ with probability 1, by a classical SLLN [4, p. 125, Cor. 1].
Example C.3. Let $X = \ell_2$ and let $P$ be the distribution of a random element $X = (X_1, X_2, \ldots)$ of $\ell_2$ with $\{X_i\}$ independent, $EX_i \equiv 0$, $EX_i^2 = \sigma_i^2$, and $\sum_{i \geq 1} \sigma_i^2 < \infty$. If, further, either

(a) $\inf_{i \geq 1} i P(|X_i| > \varepsilon \sigma_i) > 0$ for some $\varepsilon > 0$,

or

(b) $\sum_{i \geq 1} E(X_i^4)/i^2 \sigma_i^4 < \infty$,

then $\sum_{i \geq 1} X_i^2/\sigma_i^2 = \infty$ with $P$-probability 1. (Simply apply Lemma C.1 with $\{X_i/\sigma_i\}$ substituted for $\{X_i\}$.) It then follows that $E = \{x \in \ell_2 : \sum_{i \geq 1} x_i^2/\sigma_i^2 = \infty\}$ is an esupp for $P$. We now show that $E$ satisfies the criteria for $P$ to belong to $P_2(\ell_2)$.

Proof. Clearly, $E$ and $E^c$ are both nonempty. Then it suffices to show that their common boundary $\partial E = \partial(E^c)$ is $X$.

1. We show that every $x \in E$ is a limit point of $E^c$, as follows. Given $x = (x_1, x_2, \ldots)$, let $y_n = x(n) = (x_1, \ldots, x_n, 0, 0, \ldots)$. Then $\sum_{i \geq 1} y_n^2/\sigma_i^2 = \sum_{i=1}^n x_i^2/\sigma_i^2 < \infty$, so $\{y_n\} \in E^c$. Now $\|y_n - x\| = \sum_{i=n+1}^\infty x_i^2 \to 0$ since $\|x\| < \infty$. That is, $y_n \to x$, $n \to \infty$.

2. We now show that every $x \in E^c$ is a limit point of $E$. Let $\varepsilon > 0$ be given and put $\Delta_\varepsilon = \varepsilon (\sum_{i \geq 1} \sigma_i^2)^{-1/2}$. Given $x \in E^c$, take $y = x + \Delta_\varepsilon \sigma$, where $\sigma = (\sigma_1, \sigma_2, \ldots)$. Then $\|y - x\| = \Delta_\varepsilon \|\sigma\| = \varepsilon$, so that $y \in N_\varepsilon(x) \subset \ell_2$. Now $y_i/\sigma_i = x_i/\sigma_i + \Delta_\varepsilon$, and $x_i/\sigma_i \to 0$ since $x \in E^c$, so we have $|y_i/\sigma_i| \geq \Delta_\varepsilon/2$ for all large $i$, establishing that $y \in E$.

3. Therefore, $E \subset \partial(E^c)$ and $E^c \subset \partial(E^c)$. Since $\partial E = \partial(E^c)$, it follows that $X \subset \partial E$ and hence that $X = \partial E$, and hence also that $E \neq \partial E$. Therefore, $P \in P_2(\ell_2)$.

References


