Lectures on random sets and their applications in economics and finance

Ilya Molchanov*

18 June 2012

Abstract

This course introduces main concepts from the theory of random sets with emphasis on applications in economics and finance: most importantly inference for partially identified models and transaction costs modelling.

The main mathematical ideas introduced in this course are that of a random closed set, its distribution and main analytical tools to handle it, selections and expectations of random sets, laws of large numbers and limit theorems, and set-valued stochastic processes.

Theoretical material will be presented along with describing the following applications.

- Partially identified models appearing if the available data do not suffice to uniquely identify the parameter of interest, even if the sample size grows. Possible reasons for this are interval responses in regression models or multiple equilibria in games. Using random sets, it is possible to come up with an adequate mathematical framework that makes it possible to unify a number of special cases and come up with new results.
- In finance it is possible to represent the range of prices (which are always non-unique in case of transaction costs) as random sets. In the univariate case, this set is a segment with end-points being bid and ask prices. The no-arbitrage property of the dynamic model with discrete time is closely related to the existence of martingales that evolve inside the set-valued process.

1 Distributions of random sets

1.1 Basic examples

In many applications the available data come in the form of sets rather than points. For instance, in econometric applications respondents may report a salary bracket instead of the exact salary or the profit of a firm may be intentionally converted to an interval to ensure anonymity.

The most basic statistical example concerns the estimation of the mean where exact observations $x_1, \dots, x_n$ are not available, but instead it is known that x_i belongs to a set X_i for $i = 1, \dots, n$. The nature of the set X_i that replaces the point x_i may be explained by the reasons mentioned above, where there is no reason to prefer a single point from X_i over other points, or X_i may be a confidence region for the true (unknown) value x_i obtained from the initial stage of the data collection. Then it is necessary to deal with a sample of sets $X_1, \dots, X_n$, which can be considered realisations of a random set X.

*University of Bern, Institute of Mathematical Statistics and Actuarial Science, Sidlerstrasse 5, 3012 Bern, Switzerland, ilya@stat.unibe.ch. Currently visiting University Carlos III of Madrid supported by Santander Bank.

Example 1.1 (see [6])**.** Consider a variant of the classical linear regression model, where one observes the values of an explanatory variable x_i, while the only information about the response y_i is that y_i belongs to the interval $Y_i = [y_{iL}, y_{iU}]$, $i = 1, \dots, n$. This setting leads to consideration of random sets being intervals.

Example 1.2 (see [5])**.** Consider a random game of k players with pure strategies. The set of (say Nash) equilibria for such game is a random subset of the discrete cube $\{0, 1\}^k$.

Example 1.3 (see [5])**.** Consider a random element z of $\{0, \dots, n\}$ and vector $y_0, \dots, y_n$ such that $y_i = y$ is a random variable in $[0, 1]$ if $z = i$ and otherwise y_i is any element of $[0, 1]$. The corresponding random set can be represented as

$$Y = Y_0 \times \cdots \times Y_n ,$$

where $Y_i = \{y\}$ if $z = i$ and $Y_i = [0, 1]$ otherwise. This random set is useful to model response function in case of missing treatments.

Example 1.4 (see [15, 18])**.** In real life financial assets have two prices (bid and ask prices) and the interval between them represents all possible prices at which the asset can be traded. In case of several assets the range of prices is a parallelepiped, or a more general convex set if simultaneous transactions on several assets attract an extra discount and so “cut the corners” of the parallelepiped. These sets of prices depend on time and are random, so form a set-valued stochastic process.

Example 1.5 (Dangerous area)**.** Let $\xi(x)$, $x \in \mathbb{R}^2$, be a function that describes the danger level associated to a point x from the plane. Then $X = \{x : \xi(x) \geq a\}$ is the set of points with danger level at least a. The same construction can be applied if $\xi(x)$ is a confidence level at point x, so that X describes the set of points with high confidence.

These and numerous further examples coming from image analysis, material science, pattern recognition, microscopy, networks, astronomy, to name few areas, call for a rigorous definition of a random set.

1.2 Measurability of set-valued functions on probability space

The first step in defining a random element is to describe the family of its values. For a random set, the values will be subsets of a certain carrier space $\mathbb{E}$, which is often taken to be the Euclidean space $\mathbb{R}^d$, but may well be different, e.g. a cube in $\mathbb{R}^d$, a sphere, a general discrete set, or an infinite-dimensional space like the space of (say continuous) functions. It is always assumed that $\mathbb{E}$ has the structure of a topological space.

The family of *all* subsets of any reasonably rich space is immense, and it is impossible to define a non-trivial distribution on it. In view of this, one typically considers certain families of sets with particular topological properties, e.g. closed, compact or open sets, or further properties, most importantly convex sets. The conventional theory of random sets deals with random *closed* sets. An advantage of this approach is that random points (or random sets that consist of a sigle points, also called singletons) are closed and so the theory of random closed sets then includes the classical case of random points or random vectors.[1]

[1]By passing to complements it is also possible to work with open sets, then the classical case of random vectors would correspond to complements of singletons.

Denote by $\mathcal{F}$ the family of closed subsets of the carrier space $\mathbb{E}$. Recall that the empty set and the whole $\mathbb{E}$ are closed and so belong to $\mathcal{F}$. A random closed set is a map $X : \Omega \mapsto \mathcal{F}$, where Ω is the space of elementary events equipped with σ-algebra $\mathfrak{F}$ and probability measure $\mathbf{P}$.

In order to define the measurability of this map X we specify the family of functionals of X that are random variables. A possible idea would be to require that indicator function $\mathbf{1}_X(x)$ (which is one if $x \in X$ and otherwise is zero) is a stochastic process, i.e. each its value is a random variable. However this does not work well for random sets X that are "thin", e.g. for $X = \{\xi\}$ being a random singleton. For instance, if ξ is a point in the Euclidean space with an absolutely continuous distribution, then $\{x \in X\} = \{x = \xi\}$ has probability zero and so the measurability condition of the indicator $\mathbf{1}_X(x) = \mathbf{1}_{\xi=x}$ does not impose any extra requirement on ξ (if the underlying σ-algebra is complete). The same problem arises if X is a segment or a curve in the plane.[2]

So the definition of the measurability based on indicators of points does not work well. Note that too strict measurability conditions unnecessarily restrict the possible examples of random sets. On the other hand, too weak measurability conditions do not ensure that important functionals of a random set become random variables. The measurability of a random closed set is defined in the following way, which essentially replaces indicator $\mathbf{1}_X(x) = \mathbf{1}_{x\in X}$ by the indicator of the event $\{X \cap K \neq \emptyset\}$ for some test sets K.

Definition 1.6. A map X from a probability space $(\Omega, \mathfrak{F}, \mathbf{P})$ to the family $\mathcal{F}$ of closed subsets of a second countable locally compact Hausdorff space[3] $\mathbb{E}$ is called a *random closed set* if

$$X^-(K) = \{\omega : \ X(\omega) \cap K \neq \emptyset\}$$

belongs to the σ-algebra $\mathfrak{F}$ on Ω for each compact set $K \subset \mathbb{E}$.

In other words, a random closed set is a measurable map from the given probability space to the family of closed sets equipped with σ-algebra generated by the families of closed sets $\{F \in \mathcal{F} : \ F \cap K \neq \emptyset\}$ for all $K \in \mathcal{K}$, where $\mathcal{K}$ denotes the family of compact subsets of $\mathbb{E}$.

The function $X : \Omega \mapsto \mathcal{F}$ is an example of set-valued function, and X^- defined above for K being a singleton is said to be the inverse of X. In the same way it it is possible to define measurability of any set-valued function with closed values. Such function does not have to be defined on a probability space.

Many spaces of interest, e.g. the Euclidean space and discrete spaces, are second countable locally compact and Hausdorff. The local compactness condition fails for infinite-dimensional spaces, e.g. if $\mathbb{E}$ is the space of, say continuous, functions with the uniform metric and X is a random set of functions. In these spaces the family of compact sets is not sufficient to ensure good measurability properties of X and it has to be replaced with the family of closed (alternatively open, or Borel) sets.

A random *compact* set is defined as random closed set which is compact with probability one, so that almost all values of X are compact sets. A *convex* random set is defined similarly.

Example 1.7 (Random sets defined from random points)**.**

1. The singleton $X = \{\xi\}$ is a random closed set.

[2] This indicator definition however works well if X is regular closed, i.e. coincides with the closure of its interior, see [20].

[3] A topological space is called second countable if its topology has a countable base, i.e. there exists a countable family of open sets such that each open set is the union of sets from this family. A topological space is locally compact if each its point has a neighbourhood with a compact closure. A topological space is Hausdorff if each two its points have disjoint open neighbourhoods.

2. A ball $X = B_\xi(\eta)$ with η and radius ξ is a random closed set if η is a random vector and ξ is a non-negative random variable. If the joint distribution of (ξ, η) depends on a certain parameter, we obtain a parametric family of distributions for random balls.

3. A random triangle obtained as the convex hull of $\{\xi_1, \xi_2, \xi_3\}$ is a random closed set. Similarly, it is possible to consider random polytopes that appear as convex hulls of any (fixed or random) number of points in the Euclidean space.

Example 1.8 (Random sets related to deterministic and random functions). 1. Let $f : \mathbb{R}^d \mapsto \mathbb{R}$ be a deterministic function, and let ξ be a random variable. If f is continuous, then $X = \{x : f(x) = \xi\}$ is a random set. If f is upper semicontinuous, i.e.

$$f(x) \geq \limsup_{y \to x} f(y)$$

for all y, then $X = \{x : f(x) \geq \xi\}$ is closed and so also defines a random closed set. Its distribution is determined by the distribution of ξ and the choice of f. In these both case X can be obtained as the inverse image of a random set, e.g. as $f^{-1}(\{\xi\})$ or $f^{-1}((-\infty, \xi])$.

2. Let $\xi(x)$, $x \in \mathbb{R}^d$, be a real-valued stochastic process. If ξ has continuous sample paths, then $\{x : \xi(x) = t\}$ is a random closed set. If ξ has almost surely upper semicontinuous sample paths, then the excursion set $\{x : \xi(x) \leq t\}$ and the hypograph $\{(x, s) \in \mathbb{R}^d \times \mathbb{R} : \xi(x) \geq s\}$ are random closed sets.

Exercise 1.9. Let X be a random closed set in the Euclidean space $\mathbb{E} = \mathbb{R}^d$, as specified in Definition 1.6. Prove that

1. $\{x \in X\}$ is a random event, i.e. belongs to the σ-algebra $\mathfrak{F}$, for all $x \in \mathbb{R}^d$, so that $\mathbf{1}_X(x)$, $x \in \mathbb{R}^d$, is a stochastic process on $\mathbb{R}^d$;

2. $\{X \cap G \neq \emptyset\}$ is a random event for each open set G;

3. $\|X\| = \sup\{\|x\| : x \in X\}$ is a random variable (with possibly infinite values);

4. $\rho(y, X) = \inf\{\rho(x, y) : x \in X\}$, i.e. the distance from y to the nearest point of X, is a random variable, and so $\rho(y, X)$, $y \in \mathbb{R}^d$, is a stochastic process.

Exercise 1.10. Let X be a random set in a finite space $\mathbb{E}$, so that X is automatically closed. Prove that the number of points in X is a random variable.

Exercise 1.11. Using the definition check that random singleton $\{\xi\}$ is a random closed set. Check that $\{\xi_1, \ldots, \xi_n\}$ is also a random closed set, where $\xi_1, \ldots, \xi_n$ are random singletons.

Exercise 1.12. By modifying Definition 1.6, suggest definitions of random open sets or random Borel sets, see also [20].

1.3 Hitting probabilities and capacity functional

Definition 1.6 means that X is explored by its hitting events, meaning that X hits a compact set K. The corresponding hitting probabilities

$$T(K) = \mathbf{P}\{X \cap K \neq \emptyset\}\,, \quad K \in \mathcal{K}\,,$$

become a functional of K, which is called capacity (or hitting) functional of X. Sometimes we add X as a subscript and denote it T_X. Note that T is a probability measure if X is a random singleton. In general T is a monotonic map from the family $\mathcal{K}$ of compact sets to $[0,1]$, and T is called a capacity.

Exercise 1.13. Describe a random closed set whose capacity functional is a sub-probability measure, i.e. T is a measure with total mass less than or equal to one.

Exercise 1.14. Find $T(K)$ for $X = \{\xi, \eta\}$, where ξ and η are two independent random vectors in $\mathbb{R}^d$.

Exercise 1.15. Find a random closed set such that $T(K) = p$ for all non-empty compact sets K.

Example 1.16 (Random interval)**.** Let $X = [\xi, \eta]$ be a random interval on $\mathbb{R}$, where ξ and η are two (dependent) random variables so that $\xi \leq \eta$ almost surely. Then $T(\{x\}) = \mathbf{P}\{\xi \leq x \leq \eta\}$ and $T([a,b]) = \mathbf{P}\{\xi < a, \eta \geq a\} + \mathbf{P}\{\xi \in [a,b]\}$.

Example 1.17 (Random ball)**.** If $X = B_r(\eta)$ is the ball of radius r centred at random point $\eta \in \mathbb{R}^d$, then $T(K) = \mathbf{P}\{\eta \in K^r\}$, where $K^r = \{x : \rho(x,K) \leq r\}$ is r-envelope of K.

Even if two compact sets K_1 and K_2 are disjoint, non-singleton X may hit them both and so the functional T is generally not additive, but only subadditive meaning that

$$T(K_1 \cup K_2) \leq T(K_1) + T(K_2)\,. \tag{1.1}$$

This property can be strengthened to show that

$$T(K_1 \cup K_2 \cup K) + T(K) \leq T(K_1 \cup K) + T(K_2 \cup K) \tag{1.2}$$

for all compact sets $K, K_1, K_2 \in \mathcal{K}$. For the proof use

$$\mathbf{P}\{X \cap K = \emptyset, X \cap K_1 \neq \emptyset, X \cap K_2 \neq \emptyset\} \geq 0\,.$$

Note that (1.2) immediately implies (1.1) by taking $K = \emptyset$ and noticing that $T(\emptyset) = 0$.

It is natural to expect that the hitting probabilities $T(K)$ for all $K \in \mathcal{K}$ determine uniquely the distribution of X, which is indeed the case, since events $\{X \cap K \neq \emptyset\}$, $K \in \mathcal{K}$, generate the σ-algebra on the family of closed sets.

It remains to identify the properties of a functional $T(K)$ defined on a family of compact sets that guarantee the existence of a random closed set X having T as its capacity functional. This is done in the following result called the Choquet theorem formulated in the current form by G. Matheron and proved in a slightly different formulation by D.G. Kendall.

Theorem 1.18 (Choquet-Kendall-Matheron)**.** *A functional $T : \mathcal{K} \mapsto [0,1]$ defined on the family of compact subsets of a locally compact second countable Hausdorff space $\mathbb{E}$ is the capacity functional of a random closed set X in $\mathbb{E}$ if and only if $T(\emptyset) = 0$ and*

1. T *is* upper semicontinuous*, i.e.* $T(K_n) \downarrow T(K)$ *whenever* $K_n \downarrow K$ *as* $n \to \infty$ *with* K, K_n*,* $n \geq 1$*, being compact sets.*

2. T *is* completely alternating*, i.e. the following successive differences*

$$\begin{aligned} \Delta_{K_1} T(K) &= T(K) - T(K \cup K_1)\,, \\ \Delta_{K_n} \cdots \Delta_{K_1} T(K) &= \Delta_{K_{n-1}} \cdots \Delta_{K_1} T(K) \\ &\quad - \Delta_{K_{n-1}} \cdots \Delta_{K_1} T(K \cup K_n)\,, \quad n \geq 2\,. \end{aligned}$$

   are all non-positive for all compact sets $K, K_1, \ldots, K_n$*.*

Exercise 1.19. Consider $\mathbb{E} = \mathbb{R}$ and let $X = (-\infty, \xi]$ for a random variable ξ. Check that X is indeed a random closed set, express its capacity functional by means of the cumulative distribution function F of ξ and check that the upper semicontinuity condition corresponds to the right continuity of F, while the complete alternation condition is the monotonicity of F.

Exercise 1.20. Check what does the complete alternation condition imposes for $n = 1$ and that the complete alternation condition for $n = 2$ becomes (1.2).

There are three standard proofs of the Choquet theorem. One derives it from the first principles of extension of measures from algebras to σ-algebras. For this, one notices that the events of the form $\{X \cap V = \emptyset, X \cap W_1 \neq \emptyset, \ldots, X \cap W_k \neq \emptyset\}$ form an algebra, where $V, W_1, \ldots, W_k$ are obtained by taking finite unions of open and compact sets and $k \geq 0$. The probabilities of these events are given by

$$\begin{aligned} \Delta_{W_1} T(V) &= \mathbf{P}\{X \cap V \neq \emptyset\} - \mathbf{P}\{X \cap (V \cup W_1) \neq \emptyset\} \\ &= -\mathbf{P}\{X \cap W_1 \neq \emptyset,\ X \cap V = \emptyset\} \end{aligned}$$

and further by induction

$$-\Delta_{W_k} \cdots \Delta_{W_1} T(V) = \mathbf{P}\{X \cap V = \emptyset,\ X \cap W_i \neq \emptyset,\ i = 1, \ldots, k\}\,,$$

so that the non-positivity of the successive differences corresponds to the non-negativity of the probabilities. Note that the missing events can be combined, i.e. $\{X \cap V_1 = \emptyset, X \cap V_2 = \emptyset\} = \{X \cap (V_1 \cup V_2) = \emptyset\}$, while the hitting events cannot, i.e. it is not possible to simplify $\{X \cap W_1 \neq \emptyset, X \cap W_2 \neq \emptyset\}$.

Since T determines uniquely the distribution of X, properties of X can be expressed as properties of T. For instance, X is stationary (i.e. $X + a$ coincides in distribution with X for all translations a) if and only if the capacity functional of X is translation invariant.

Exercise 1.21. Prove that a random closed set is convex if and only if its capacity functional satisfies

$$T(K_1 \cup K_2) + T(K_1 \cap K_2) = T(K_1) + T(K_2)$$

for all convex compact sets K_1 and K_2 such that $K_1 \cup K_2$ is also convex.

The capacity functional can be properly extended to a functional on all (even non-measurable!) subsets of $\mathbb{R}^d$, see [20]. This is done by approximation, for example $T(\mathbb{R}^d)$ becomes the limit of $T(K)$ as $K \uparrow \mathbb{R}^d$. Note that $T(\mathbb{R}^d)$ can be strictly smaller than one, since the random set X may be empty.

A closely related functional is the *avoidance functional* $Q(K) = 1 - T(K)$ that gives the probability that X misses compact set K. The avoidance functional can be written as

$$Q(K) = \mathbf{P}\{X \cap K \neq \emptyset\} = \mathbf{P}\{X \subset K^c\}\,,$$

where K^c is the complement to K. The right-hand side corresponds to the value of the capacity functional on the open set K^c. Thus, the avoidance functional is related to the containment functional

$$C(F) = \mathbf{P}\{X \subset F\}\,.$$

Another useful functional related to X is the inclusion functional

$$I(K) = \mathbf{P}\{K \subset X\}\,.$$

The inclusion functional vanishes if X is a singleton. The distribution of X is possible to describe by the inclusion functional if an only if the indicator function $\mathbf{1}_X$ is a separable stochastic process, and in this case X itself is also called separable.

Example 1.22 (Random intervals)**.** Let $Y = [y_L, y_U]$ be a random segment, where y_L and y_U are two dependent random variables such that $y_L \leq u_U$ almost surely. Then the capacity functional of Y for $K = [a, b]$ is

$$T_Y(K) = \mathbf{P}\{y_L \in [a, b] \ \text{ or } \ y_U \in [a, b]\}$$

and so determines the joint distribution of y_L and y_U. In this case the containment functional has an simpler expression

$$C([a, b]) = \mathbf{P}\{[y_L, y_U] \subset [a, b]\} = \mathbf{P}\{y_L \geq a,\ y_U \leq b\}\,.$$

Example 1.23 (Measure of a random set)**.** If μ is a locally finite measure, then $\mu(X)$ is a random variable. Indeed, Fubini's theorem applies to the integral of $\mathbf{1}_X(x)$ with respect to $\mu(dx)$ and leads to

$$\begin{aligned} \mathbf{E}\mu(X) &= \mathbf{E}\int \mathbf{1}_X(x)\,\mu(dx) \\ &= \int \mathbf{E}\,\mathbf{1}_X(x)\,\mu(dx) \\ &= \int \mathbf{P}\{x \in X\}\,\mu(dx)\,. \end{aligned}$$

The fact that the expected value of $\mu(X)$ for a locally finite μ equals the integral of the probability $\mathbf{P}\{x \in X\}$ is known under the name of the Robbins theorem formulated by A.N. Kolmogorov in 1933 and then independently by H.E. Robbins in 1944-45. It should be noted that this fact does not hold for a general measure μ. [4]

Exercise 1.24. Find an expression for $\mathbf{E}(\mu(X)^n)$ for a locally finite measure μ.

Example 1.25 (Distance function)**.** If X is a random closed set, then the distance function $\xi(y) = \rho(y, X)$ is a stochastic process. Then

$$\mathbf{P}\{\xi(y) \leq t\} = \mathbf{P}\{X \cap B_t(y) \neq \emptyset\} = T(B_t(y))\,.$$

Thus,

$$\mathbf{E}\rho(y, X) = \int_0^\infty (1 - T(B_t(y)))dt\,.$$

[4] For instance, if X is a singleton with an absolutely continuous distribution and μ is the counting measure, then $\mathbf{E}\mu(X) = 1$, while $\mathbf{P}\{x \in X\}$ vanishes identically.

Example 1.26 (Point processes)**.** A simple point process ϕ can be viewed as the random closed set X which is locally finite, i.e. such that each compact set K contains only a finite number of points from X. Since $\mathbf{P}\{X \cap K = \emptyset\} = \mathbf{P}\{\phi(K) = 0\}$, the distribution of a simple point process is identically determined by its avoidance probabilities (i.e. probabilities that a given compact set contains no point of the process). For instance, a random closed set with the capacity functional

$$T(K) = 1 - e^{-\Lambda(K)}, \quad K \in \mathcal{K},$$

with Λ being a locally finite measure on $\mathbb{R}^d$ is the Poisson process with intensity measure Λ.

It is possible to integrate with respect to the hitting functional. The Choquet integral of a non-negative function is defined as

$$\int f dT = \int_0^\infty T(\{x : \ f(x) \geq t\})dt .$$

2 Selections and measurability issues

2.1 Existence of measurable selections

A random point ξ is said to be a *selection* of random set X if $\xi \in X$ almost surely. In order to emphasise the fact that ξ is measurable itself, one often calls it a *measurable* selection. A possibly empty random set clearly does not have a selection. Otherwise, the fundamental selection theorem establishes the existence of a selection of a random closed set under rather weak conditions. It is formulated below for random closed sets in $\mathbb{R}^d$.

Theorem 2.1 (Fundamental selection theorem, see [20])**.** *If $X : \Omega \mapsto \mathcal{F}$ is an almost surely non-empty random closed set in $\mathbb{R}^d$, then X has a measurable selection.*

Remark 2.2 (Different selections for identically distributed sets)**.** Since the family of selections depends on the underlying σ-algebra, two identically distributed random closed sets might have different families of selections. For instance, consider random closed set X which always takes the value $\{0, 1\}$. If the underlying probability space Ω is trivial, then the only selections of X are $\xi = 0$ and $\xi = 1$, while if the probability space is rich, e.g. $\Omega = [0, 1]$, then a random point taking values either 0 or 1 is also a selection of X. However, it is known that the weak closures of the families of selections coincide if the random closed sets are identically distributed.

The following result by C. Himmelberg establishes equivalences of several measurability concepts. It is formulated in a bit restrictive form for Polish spaces.[5]

Theorem 2.3 (Fundamental Measurability Theorem)**.** *Let E be a Polish space and let $X : \Omega \mapsto \mathcal{F}$ be a function defined on a complete probability space $(\Omega, \mathfrak{F}, \mathbf{P})$ with values being non-empty closed subsets of E. Then the following statements are equivalent.*

1. *$X^-(B) = \{\omega : \ X(\omega \cap B \neq \emptyset\} \in \mathfrak{F}$ for every Borel set $B \subset E$.*
2. *$X^-(F) \in \mathfrak{F}$ for every $F \in \mathcal{F}$.*
3. *$X^-(G) \in \mathfrak{F}$ for every open set $G \subset E$ (in this case X is said to be* Effros measurable.
4. *The distance function $\rho(y, X) = \inf\{\rho(y, x) : \ x \in X\}$ is a random variable for each $y \in E$.*

[5] Each complete separable metric space is Polish.

5. *There exists a sequence $\{\xi_n, n \geq 1\}$ of measurable selections of X such that*

$$X = \mathrm{cl}\{\xi_n,\ n \geq 1\}\,.$$

6. *The graph of X*

$$\mathrm{graph}(X) = \{(\omega, x) \in \Omega \times E:\ x \in X(\omega)\}$$

is measurable in the product σ-algebra of $\mathfrak{F}$ and the Borel σ-algebra on E.

Note that the family of compact sets does not appear in the Fundamental Measurability Theorem. Indeed, the family of compact sets in a general Polish space can be rather poor. If $\mathbb{E} = \mathbb{R}^d$ (or more generally if E is locally compact), then all above measurability conditions are equivalent to $X^-(K) \in \mathfrak{F}$ for all compact sets K.

Exercise 2.4 (see [20])**.** Let X be *regular closed*, i.e. X almost surely coincides with the closure of its interior. Show that all measurability properties of X are equivalent to $\{x \in X\} \in \mathfrak{F}$ for all $x \in E$.

In particular, Statement 5 of Theorem 2.3 means that X can be obtained as the closure of a countable family of random singletons, known as the Castaign representation of X. This is a useful tool to extend concepts defined for points to their analogues for random sets.

The fundamental measurability theorem helps to establish measurability of set-theoretic operations with random sets.

Theorem 2.5 (Measurability of set-theoretic operations)**.** *If X is a random closed set in a Polish space E, then the following multifunctions are random closed sets:*

1. *the closed convex hull of X;*
2. *αX if α is a random variable;*
3. *the closed complement to X, the closure of the interior of X, and ∂X, the boundary of X.*

If X and Y are two random closed sets, then

1. *$X \cup Y$ and $X \cap Y$ are random closed sets;*
2. *the closure of $X + Y = \{x + y:\ x \in X,\ y \in Y\}$ is a random closed set (if E is a Banach space);*
3. *if both X and Y are bounded, then the Hausdorff distance[6] $\rho_H(X, Y)$ is a random variable.*

If $\{X_n,\ n \geq 1\}$ is a sequence of random closed sets, then

1. *$\mathrm{cl}(\bigcup_{n\geq 1} X_n)$ and $\bigcap_{n\geq 1} X_n$ are random closed sets;*
2. *$\limsup_{n\to\infty} X_n$ and $\liminf_{n\to\infty} X_n$ are random closed sets.*

[6] The Hausdorff distance between two bounded sets X and Y in a metric space is defined as the infimum of all ε such that X is contained in the ε-neighbourhood of Y and Y is contained in the ε-neighbourhood of X.

2.2 Characterisation of selections

Probability distributions of selections can be characterised by the following domination condition.

Theorem 2.6 (Artstein)**.** *A probability distribution μ is the distribution of a selection of random closed set X if and only if*

$$\mu(K) \le T(K) = \mathbf{P}\{X \cap K \neq \emptyset\} \tag{2.1}$$

for all compact sets K.

It is important to note that if μ from Theorem 2.6 is the distribution of some random vector ξ, then it is not guaranteed that $\xi \in X$ a.s., e.g. ξ can be independent of X. Theorem 2.6 means that for each such μ it is possible to construct ξ with distribution μ that belongs to X almost surely, in other words one couples ξ and X on the same probability space. If X is a singleton, then the inequality in (2.1) turns into an equality and means that $X = \{\xi\}$ for a random vector ξ coupled with X on the same probability space.

Proof of Theorem 2.6. The necessity is evident. The proof of sufficiency is based on the domination condition for probability distributions in partially ordered spaces from [17], which implies that two random sets Y and X can be coupled on the same probability space so that $Y \subset X$ a.s. if and only if

$$\mathbf{P}\{Y \cap K_1 \neq \emptyset, \dots, Y \cap K_n \neq \emptyset\} \le \mathbf{P}\{X \cap K_1 \neq \emptyset, \dots, X \cap K_n \neq \emptyset\} \tag{2.2}$$

for all compact sets $K_1, \dots, K_n$ and $n \ge 1$. In the special case of $Y = \{\xi\}$ being a singleton condition (2.1) implies that

$$\begin{aligned}
\mathbf{P}\{Y \cap K_1 \neq \emptyset, \dots, Y \cap K_n \neq \emptyset\} &= \mathbf{P}\{\xi \in \cap_{i=1}^n K_i\} \\
&\le \mathbf{P}\{X \cap \left(\cap_{i=1}^n K_i\right)\} \\
&\le \mathbf{P}\{X \cap K_1 \neq \emptyset, \dots, X \cap K_n \neq \emptyset\}\,.
\end{aligned}$$

□

The realisability of μ as a selection of X can be interpreted as the fact that μ is stochastically smaller than the distribution of X. This concept can be extended to compare two random sets Y and X. However, the domination of the capacity functionals $T_Y(K) \le T_X(K)$ is substantially weaker than (2.2) and does not suffice to ensure that X and Y can be realised on the same probability space (coupled) so that $Y \subset X$ a.s.

The family of compact sets in Theorem 2.6 can be replaced by all closed or all open sets. By passing from the capacity functional to the containment functional we arrive at the equivalent criterion requiring that

$$\mu(F) \ge C(F) = \mathbf{P}\{X \subset F\} \tag{2.3}$$

for all closed sets F.

2.3 Reduction of the family of test compact sets

An important issue in relation to distributions of random sets is the possibility to reduce the family of all compact sets required to describe the distribution or characterise a selection of a random closed set. The first such reduction is possible if the family of all compact sets is replaced by its subfamily, which is dense in a certain sense, see [20]. For instance, in the Euclidean space, it suffices to consider compact sets obtained as finite unions of closed balls with rational centres and radii.

For a further reduction one should impose further restrictions on the family of realisations of X. Assume that X is almost surely convex. It would be natural to expect that the probabilities of the type $C(F) = \mathbf{P}\{X \subset F\}$ for all convex closed sets F determine uniquely the distribution of X.

Theorem 2.7 (see [20])**.** *The distribution of a convex compact random set X in the Euclidean space is uniquely determined by its containment functional $C(F) = \mathbf{P}\{X \subset F\}$ on the family of all compact convex sets F (even on the family of all compact convex polytopes).*

Proof. A random compact convex set X can be viewed as its *support function*

$$h_X(u) = \sup\{\langle x, u\rangle : x \in X\}$$

being a conventional sample continuous stochastic process on the unit sphere. Then the probabilities of the events $\{h_X(u_1) \le a_1, \dots, h_X(u_k) \le a_k\}$ determine the finite-dimensional distributions of the support function. It suffices to note that these probabilities are exactly the values of the containment functional on a polytope F whose faces have normal vectors $u_1, \dots, u_k$. □

Theorem 2.7 does not hold for non-compact X even if F is any convex closed set. A counterexample is a random half-space touching the unit ball at a uniformly distributed point. Indeed, in this case, the containment functional $C(F)$ vanishes for each convex F.

The reduction of the family of compact sets required for the characterisation of selections is even more complicated. For instance, it is not sufficient to check (2.3) for convex sets F even if X is almost surely compact and convex. For the following result, note that $X \subset K$ if and only if $X \subset K_X$, where

$$K_X = \bigcup_{\omega \in \Omega',\, X(\omega) \subset K} X(\omega)$$

for any set Ω' of the full probability.

Proposition 2.8. *Let X be a random compact set. Then μ is the distribution of a selection of X if and only if*

$$\mu(K) \ge \mathbf{P}\{X \subset K_X\} \tag{2.4}$$

for all compact sets K.

Exercise 2.9. Assume that X is a segment in the plane. Suggest a reduction of the family of compact sets in order to characterise its distribution and describe all selections? Explore what Proposition 2.8 brings in this case.

Exercise 2.10. Assume that X is the union of two convex sets. What could be the smallest class to characterise the distribution of X and to identify its selections?

Example 2.11 (Games with pure equilibria)**.** Consider a random game of n players which is assumed to possess with probability one at least one equilibrium in pure strategies. The pure strategies of all players can be represented as n-vectors composed of 0 and 1, so a point from the discrete cube $\{0,1\}^n$. Let X be random set of equilibria, which is a subset of the discrete cube $\{0,1\}^n$, so the carrier space is finite and consists of 2^n points. The observed equilibrium ξ is a selection of X. The probability distribution of ξ, which is vector of length 2^n describes the theoretical frequencies of all possible equilibria. By Theorem 2.6, it can be characterised by a finite (however large) set of inequalities. For estimation purposes, the distribution of X depends on a parameter which can be estimated as solution of these inequalities constructed with the empirical counterpart of μ.

Example 2.12 (Treatment effects)**.** Consider random set from Example 1.3. Its selections represent all possible treatment responses that are compatible with the observation in presence of missing treatments. Thus, the distributions of possible responses are exactly those which satisfy the inequalities (2.1).

Exercise 2.13. Let $X = (-\infty, \xi]$. Which family of compact sets is necessary to consider in order to ensure that the domination condition $\mu(K) \leq T_X(K)$ implies that μ is the distribution of a selection of X? How to generalise this for X being a random half-space in $\mathbb{R}^d$?

2.4 Weak convergence

The weak convergence of random closed sets is defined by specialising the general weak convergence definition for random elements or, equivalently, the weak convergence of probability measures on the space $\mathcal{F}$ of closed sets.

Theorem 2.14. *A sequence $\{X_n, n \geq 1\}$ of random closed sets in $\mathbb{R}^d$ converges weakly to X if and only if $T_{X_n}(K) \to T_X(K)$ for all compact sets K such that $T_X(K) = T_X(\mathrm{Int}K)$.*

The condition on K in Theorem 2.14 is akin to the conventional requirement for the weak convergence of random variables which is equivalent to the convergence of their cumulative distribution functions at all points of continuity of the limit. Condition $T_X(K) = T_X(\mathrm{Int}K)$ means that X "touches" K with probability zero.

Exercise 2.15. Prove that the weak convergence of random singletons $X_n = \{\xi_n\}$ corresponds exactly to the weak convergence of random vectors ξ_n in the classical sense.

Exercise 2.16. Figure out what the weak convergence means for intervals $X_n = [\xi_n, \eta_n]$ on the real line.

It is possible to metrise the weak convergence, i.e. to define the distance between distributions of random sets that metrises the convergence in probability, see [20].

3 Empirical distributions

3.1 Glivenko–Cantelli theorem for empirical capacity functionals

Assume that a sample $X_1, \dots, X_n$ of values of a random closed set X is given. A natural estimator for the capacity functional of X is its empirical variant

$$\hat{T}_n(K) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{K \cap X_i \neq \emptyset} \ .$$

The conventional law of large numbers implies that $\hat{T}_n(K)$ converges to $T(K)$ almost surely for each given K. It is often desirable to come up with a uniform convergence for K from a certain family $\mathcal{M}$ of compact sets, so that, for each compact set K_0,

$$\sup_{K \in \mathcal{M},\, K \subset K_0} |\hat{T}_n(K) - T_X(K)| \to 0 \quad \text{a.s. as } n \to \infty \,. \tag{3.1}$$

Exercise 3.1. Show that it is possible to get rid of condition $K \subset K_0$ in (3.1) if X is almost surely compact.

The following example shows that even the family of all singletons as $\mathcal{M}$ may fail to the Glivenko–Cantelli theorem (3.1).

Example 3.2. Let $\mathcal{M} = \{\{x_n\},\ n \geq 1\}$ be a countable family of distinct singletons in $[0,1]$. Passing to subclasses if necessary, assume that $x_n \to x_0$ as $n \to \infty$, and $\{x_0\} \in \mathcal{M}$. Let $\{\xi_n,\ n \geq 1\}$ be a sequence of independent identically distributed random variables taking the values 0 and 1 with probability 1/2. Define

$$X = \{x_n :\ n \geq 1,\ \xi_n = 1\} \cup \{x_0\}\,.$$

Then, for all its independent realisations $X_1, \dots, X_n$, there exists a point x_n (with a possibly random n) such that $x_n \notin \cup_{i=1}^n X_i$ almost surely. Hence, $\hat{T}_n(\{x_n\}) = 0$, while $T(\{x_n\}) = 1/2$, i.e. (3.1) does not hold.

The reason for this effect lies in possible realisations of X and similar examples can be constructed for other nowhere dense random sets of positive Lebesgue measure.

Recall that $\operatorname{Int} X$ denotes the interior of X and ∂X is the boundary of X. The random set X is said to be regular closed if X almost surely coincides with the closure of its interior.

Theorem 3.3. *Let $\mathcal{M}$ be a class of compact sets that is closed in the Hausdorff metric*[7]*. If a random closed set X satisfies the conditions:*

(A1) *X is almost surely regular closed;*

(A2) *for each $K \in \mathcal{M}$, $T_X(K) = T_{\operatorname{Int} X}(K)$, i.e. $T_X(K) = \mathbf{P}\{\operatorname{Int} X \cap K \neq \emptyset\}$;*

then (3.1) *holds. Conditions **(A1)** and **(A2)** are also necessary if $\mathcal{M} = \mathcal{K}$ and X is a.s. continuous, i.e. $\mathbf{P}\{x \in \partial X\} = 0$ for each $x \in \mathbb{R}^d$.*

The next corollary follows from the fact that the boundary of an a.s. regular closed set in $\mathbb{R}^1$ contains at most countably many points.

Corollary 3.4. *If X is an a.s. continuous random closed subset of the line $\mathbb{R}$, then* (3.1) *holds if and only if X is a.s. regular closed.*

Proof. It suffices to prove that $\mathbf{P}\{X \cap K \neq \emptyset,\ \operatorname{Int} X \cap K = \emptyset\} = 0$ for every compact set $K \subset \mathbb{R}$. Note that $\operatorname{Int} X = \cup_{i=1}^{\infty} (\alpha_i, \beta_i)$, where α_i, β_i, $i \geq 1$, are selections of X. Then

$$\mathbf{P}\{X \cap K \neq \emptyset,\ \operatorname{Int} X \cap K = \emptyset\} \leq \sum_{i=1}^{\infty} \mathbf{P}\{\alpha_i \in K,\ \operatorname{Int} X \cap K = \emptyset\} + \sum_{i=1}^{\infty} \mathbf{P}\{\beta_i \in K,\ \operatorname{Int} X \cap K = \emptyset\}\,.$$

Consider one of the summands in the first sum. Let K_1 be the set of $x \in K$ such that $K \cap (x, x+\varepsilon) = \emptyset$ for some $\varepsilon > 0$. Since every $x \in K_1$ corresponds to an interval from a family of disjoint intervals on $\mathbb{R}$, the set K_1 is at most countable. Then

$$\mathbf{P}\{\alpha_i \in K,\ \operatorname{Int} X \cap K = \emptyset\} \leq \sum_{x \in K_1} \mathbf{P}\{\alpha_i = x,\ (\alpha_i, \beta_i) \cap K = \emptyset\} \leq \sum_{x \in K_1} \mathbf{P}\{x \in \partial X\} = 0\,.$$

□

[7] The family $\mathcal{M}$ is closed in the Hausdorf metric if $\rho_H(K_n, K) \to 0$ for a sequence $K_n \in \mathcal{M}$ means that the limit K also belongs to $\mathcal{M}$.

It is well known that the family of all half-lines is a universal class (also called the Vapnik–Chervonenkis class) for empirical distributions of random variables, i.e. the empirical distribution of each random variable converges uniformly on this class to the theoretical distribution. In contrast to this, the universal classes for random closed sets are very poor, see [21].

3.2 Functional limit theorem

It follows from the standard central limit theorem that the finite dimensional distributions of the random field

$$Z_n(K) = \sqrt{n}\Big(T_n^*(K) - T_X(K)\Big)\,, \quad K \in \mathcal{K}\,, \tag{3.2}$$

converge as $n \to \infty$ to the finite-dimensional distributions of the Gaussian field $Z(K), K \in \mathcal{K}$, with zero mean and covariance

$$\begin{aligned} \sigma(K_1, K_2) &= \mathbf{E}\,[Z(K_1)Z(K_2)] \\ &= T_X(K_1) + T_X(K_2) - T_X(K_1 \cup K_2) - T_X(K_1)T_X(K_2)\,. \end{aligned} \tag{3.3}$$

However, to ensure the functional convergence of $Z(K)$ for $K \in \mathcal{M}$, the class $\mathcal{M}$ and the functional $T_X(\cdot)$ must satisfy additional conditions, which are similar to those used in the theory of empirical measures and set-indexed random functions [7, 25].

Theorem 3.5. *Suppose that $K \subset K_0$ for all $K \in \mathcal{M}$ and that*

(C1) $\log \nu(\varepsilon) = \mathcal{O}(\varepsilon^{-\beta})$ *for some $\beta \in (0,1)$, where $\nu(\varepsilon)$ is the cardinality of the minimum ε-net of $\mathcal{M}$ in the Hausdorff metric*[8]*;*

(C2) *there exists $\gamma > \beta$, such that*

$$\sup_{\rho_H(K_1,K_2)<\varepsilon,\; K_1,K_2\in\mathcal{M}} |T_X(K_1) - T_X(K_2)| = \mathcal{O}(\varepsilon^{\gamma})\,.$$

Then $Z_n(\cdot)$ converges weakly in the uniform metric to the Gaussian random field $Z(\cdot)$ on $\mathcal{M}$ with the covariance (3.3)*, i.e. every continuous in the uniform metric functional of Z_n converges in distribution to its value on Z.*

Note that **(C1)** and **(C2)** imply that Z is a.s. continuous in the Hausdorff metric on $\mathcal{M}$.

The empirical capacity functionals can be used to produce minimum contrast estimators for parameters of random sets. Assume that $T(\cdot;\theta)$ is the capacity functional that depends on unknown parameter θ. Then θ can be estimated by minimising the uniform distance between $T(K;\theta)$ and $\hat{T}_n(K)$ for K from a certain family of compact sets, see [10] for an application of this method for some stationary random sets called the Boolean models.

[8]The ε-net is the family $K_1, \dots, K_{\nu(\varepsilon)}$ such that each set $K \in \mathcal{M}$ lies within ε-neighbourhood of a set from the net.

3.3 Stationary random sets

A random closed set X in $\mathbb{R}^d$ is stationary if X coincides in distribution with $X+a$ for each $a \in \mathbb{R}^d$. In this case the indicator function $\mathbf{1}_X(x)$ is a stationary random field and in particular one-point covering probabilities

$$p = \mathbf{P}\{x \in X\}$$

do not depend on $x \in \mathbb{R}^d$ and p is called the volume (or area if $d = 2$) fraction of X.

The stationarity assumption makes it possible to estimate probabilities associated with X by averaging over the space. For this, the set X should satisfy the ergodicity assumption and, in order that the central limit theorem holds, also the mixing assumption. These assumptions are essentially the ergodicity and mixing properties of the indicator process if X is regularly closed and so the indicator process identifies its distribution.

Let W be an observation window, for instance a ball of a growing radius.[9] The volume fraction of X can be estimated as

$$\hat{p}_W = \frac{|X \cap W|}{|W|},$$

where $|W|$ is the volume (Lebesgue measure) of W. The properties of estimator $\hat{p}_W$ are well understood, since it can be also written as the spatial average of the indicator process

$$\hat{p}_W = \frac{1}{|W|} \int_W \mathbf{1}_X(x) dx \,.$$

The second-order properties of this estimator depend on the covariance of X defined as

$$C(x_1, x_2) = \mathbf{P}\{x_1, x_2 \in X\}$$

and on the geometry of the window W through its covariogram $\gamma_W(x) = |W \cap (W + x)|$.

The capacity functional $T(K)$ can be estimated by noticing that the probability $\mathbf{P}\{K+a \cap X \neq \emptyset\}$ does not depend on a and so is the volume fraction of the random set

$$X + \check{K} = \{x - y : \ x \in X, \ y \in K\} \,.$$

Therefore, we arrive at the estimator

$$\hat{T}_W(K) = \frac{|(X + \check{K}) \cap W|}{|W|} \,.$$

While this estimator is consistent, it should be used with a care if K is "large" because of the so-called edge effects. Indeed, $(X + \check{K}) \cap W \neq \emptyset$ might result from points of X lying outside W. The easiest way to handle edge effects is to use the plus-sampling, namely observe X inside the enlarged window in order to make inference in the original (smaller) window.

The obtained estimator $\hat{T}_W(K)$ is consistent and its uniform consistence over some family of compact sets can be shown in the same way as in Section 3.1.

[9] It is possible to formulate a general condition on the sequence of growing windows that yield consistent estimators. In general, it suffices to assume that W is a scaled convex set containing the origin in its interior.

Example 3.6 (Boolean model)**.** Let $\Pi = \{x_i,\ i \geq 1\}$ be the stationary Poisson process in $\mathbb{R}^d$ of intensity λ and let $X_1, X_2, \dots$ be a sequence of i.i.d. random compact sets such that X_1 lies within a ball of integrable volume. Then

$$X = \bigcup_{i \geq 1} (x_i + X_i)$$

is a random closed set called the *Boolean model* with the typical grain X_1. This construction makes it possible to construct "complicated" patterns from rather simple components, e.g. in case X_1 is a ball centred at the origin. The statistical issues for the Boolean model consist in estimation of λ and the distribution of X_1 by observing the union set X. Note that in the individual grains $x_i + X_i$ may be not visible in X because of occlusions and so this estimation task is not trivial, see [22].

3.4 Solutions of inequalities and quantiles

Sets appear naturally as solutions of inequalities. Let h be a real-valued function on $\mathbb{R}^d$.[10] Define

$$H(t) = \{x \in \mathbb{R}^d :\ h(x) \leq t\}\,. \tag{3.4}$$

A function h is said to be lower semicontinuous if $h(a) \leq \liminf_{x \to a} h(x)$ for all a, equivalently, that the level sets (3.4) are closed for all t. If the estimator h_n of h is also lower semicontinuous, then the plug-in estimator of $H(t)$

$$H_n(t) = \{x :\ h_n(x) \leq t\} \tag{3.5}$$

is a random closed set. Note that (possibly unknown and estimated) t can be included in h by considering new function $h(x) - t$.

The above estimation problem appears in applications if h is a density function or a cumulative distribution function, if h represent some moment inequalities used to estimate the unknown parameter x or if h are grey values of an image with the aim to threshold it. If h satisfies extra smoothness conditions, then its level sets are smooth and so can be estimated with better rates, see [19, 26].

If $h(x) = \mathbf{P}\{\xi < t\}$ is the left continuous variant of the cumulative distribution function, then $H(t) = (-\infty, x_t]$, where x_t is the t-quantile of the random variable ξ. If ξ is a random vector, we obtain a set-valued quantile, also considered as multivariate generalisation for the value-at-risk.

The properties of the plug-in estimator (3.5) have been studied in [23].

Theorem 3.7. *Assume that, for each compact set K_0,*

$$\eta_n = \sup_{x \in K_0} |h_n(x) - h(x)| \to 0 \quad a.s.\ as\ \ n \to \infty\,. \tag{3.6}$$

The estimator $H_n(t)$ is strongly consistent in the Hausdorff metric, i.e.

$$\rho_H(H_n(p) \cap K_0, H(p) \cap K_0) \to 0 \quad a.s.\ as\, n \to \infty \tag{3.7}$$

for each compact set K_0 if

$$H(t) \subseteq \mathrm{cl}(H(t-))\,, \tag{3.8}$$

where $\mathrm{cl}(H(t-))$ *is the closure of $H(t-) = \{x :\ h(x) < t\}$. If for each x there exists a sequence $\{n(k)\}$ such that $h_{n(k)}(x) > h(x)$ a.s., then* (3.8) *is also a necessary condition.*

[10] The setting can be generalised for collections of functions and so for systems of inequalities.

Assume that the random field

$$\zeta_n(x) = a_n(h(x) - h_n(x)), \quad x \in X, \tag{3.9}$$

has a weak limit $\zeta(x)$ as $n \to \infty$, i.e. $g(\zeta_n) \to g(\zeta)$ in distribution for each continuous functional g on the space $C(X)$ of continuous functions on X with the uniform metric. It is also possible to come up with the limit theorem for the Hausdorff distance between $H(t) \cap K_0$ and its estimator.

3.5 Set estimation

Assume that we would like to estimate an unknown (deterministic) set K by observing a sample of points $\{x_1, \dots, x_n\}$ from it. One of most general estimators is to take the union of balls centred at these sampled points with radius r_n that converges to zero as $n \to \infty$, so that the estimator becomes

$$\hat{K}_n = \cup_{i=1}^n B_{r_n}(x_n).$$

The corresponding estimator is consistent if r_n converges to zero sufficiently slow, see [19]. This construction is similar to the non-parametric curve estimation, which introduces bias in order to obtain a smoother version of the estimator.

While the above estimator is, perhaps, the only possible for generic compact sets K, it can be substantially improved if more information about K is available. For instance, if K is convex, then a much better estimator is obtained as the convex hull P_n of $\{x_1, \dots, x_n\}$, see [27] for one of the first applications of this estimator. There is a vast literature on asymptotic properties of convex hulls, which makes it possible to assess the Hausdorff distance between P_n and K and so to come up with confidence bands for K.

If K has a sufficiently smooth boundary, then further estimators are possible by adjusting the rate at which r_n converges to zero and exploiting the local features of the boundary, see [19].

3.6 Summary: statistical settings

Above we have seen several basic inference settings for random sets.

(i) Estimate parameter of the distribution of a random set X by observing a sample $X_1, \dots, X_n$ (or a variant for stationary sets).

(ii) Estimation of a deterministic set by a random set constructed as solution of inequalities or by observing a sample of points from this set.

The following settings are also important in view of application to partially identified models.

(iii) Estimate a parameter of a random set X by observing a sample of selections of independent copies of X. In contrast to **(i)** we observe x_i from X_i for $i = 1, \dots, n$. This setting arises, e.g. in analysis of games with multiple equilibria. In this case it is usually not possible to come up with a single estimated value for the parameter specifying the distribution of X, since many distributions of X may be compatible with the observed sample of selections.

(iv) Make inference about the distribution of possible selections by observing a sample of realisations of X. Again, here many distributions of selections are possible.

Example 3.8 (Domination)**.** Let $X = (-\infty, \xi]$, where ξ is normally distributed with unknown mean m and the known variance (say $\sigma^2 = 1$). Assume that we observe a sample of values $x_1, \ldots, x_n$ being selections of independent copies of X. By Theorem 2.6 an in view of the suitable reduction of the family of test compact sets, we need to find the parameters of ξ (and so of X) by checking that ξ dominates the observed family of selections. A possible identification region is given by the family of all $m \in \mathbb{R}$ such that $\Phi(t-m) \leq \hat{F}_n(t)$, where Φ is the cumulative distribution function of the standard normal distribution and $\hat{F}_n$ is the empirical distribution function of the observed sample $x_1, \ldots, x_n$.

4 Minkowski sums

4.1 Expectation of a random set

The space of closed sets is not linear, which causes substantial difficulties in defining the expectation for a random set. One way described below relies on the representation a random set using the family of its selections.

Let X be a random closed set in $\mathbb{R}^d$. If X possesses at least one integrable selection, then X is called integrable. For instance, if X is almost surely non-empty compact and its norm $\|X\|$ is an integrable random variable (then X is said to be *integrably bounded*), then all selections of X are integrable and so X is integrable too. Later on we usually assume that X is integrably bounded.

Definition 4.1. The *(selection or Aumann) expectation* $\mathbf{E}X$ of an integrable random closed set X is closure of the family of all expectations for its integrable selections.

If X is an integrably bounded subset of $\mathbb{R}^d$, then the expectations of all its selections form a closed set and there is no need to take an additional closure. The so defined expectation depends on the probability space where X is defined. For instance, the deterministic set $X = \{0, 1\}$ defined on the trivial probability space has expectation $\mathbf{E}X = \{0, 1\}$, since it has only two trivial (deterministic) selections, see Remark 2.2. However, if X is defined on a non-atomic probability space, then its selections are $\xi = \mathbf{1}_A$ for all events $A \subset \Omega$, so that $\mathbf{E}\xi = \mathbf{P}(A)$ and the range of possible values for $\mathbf{E}\xi$ constitutes the whole interval $[0, 1]$.

The following result shows that the selection expectation is a convex set. A useful mathematical tool suitable to describe a convex set K in the Euclidean space $\mathbb{R}^d$ is its *support function* defined as

$$h_K(u) = \sup\{\langle x, u\rangle : \ x \in K\},$$

where $\langle x, u\rangle$ denotes the scalar product. Note that the support function is finite for all u if K is bounded. Sometimes the support function is considered only for u from the unit sphere, since it is one-homogeneous meaning that $h_K(cu) = c h_K(u)$ for all $c \geq 0$.

Exercise 4.2. Show that the support function $h_K(-u)$ equals $h_{\check{K}}(u)$, where $\check{K}$ is the set centrally symmetric to K with respect to the origin.

Theorem 4.3. *If an integrably bounded X is defined on a non-atomic probability space, then* $\mathbf{E}X$ *is a convex set and*

$$\mathbf{E}h_X(u) = h_{\mathbf{E}X}(u), \quad u \in \mathbb{R}^d. \tag{4.1}$$

Proof. The convexity of the Aumann expectation can be derived from the Lyapunov theorem which says that the range of any vector-valued measure is a convex set (to see this in the one-dimensional case note that real-valued measures do not have gaps in the ranges of their values). Let ξ_1 and ξ_2 be two integrable selections of X. Define the vector-valued measure

$$\lambda(A) = (\mathbf{E}(\mathbf{1}_A\, \xi_1), \mathbf{E}(\mathbf{1}_A\, \xi_2))$$

for all measurable events A. The closure of its range is convex, $\lambda(\emptyset) = (0,0)$ and $\lambda(\Omega) = (\mathbf{E}\xi_1, \mathbf{E}\xi_2)$. Let $\alpha \in (0,1)$. Thus, there exists an event A such that

$$|\alpha \mathbf{E}\xi_i - \mathbf{E}(\mathbf{1}_A\, \xi_i)| < \varepsilon/2\,, \quad i = 1,2\,.$$

Define the selection

$$\eta = \mathbf{1}_A\, \xi_1 + \mathbf{1}_{A^c}\, \xi_2\,.$$

Then

$$\|\alpha \mathbf{E}\xi_1 + (1-\alpha)\mathbf{E}\xi_2 - \mathbf{E}\eta\| < \varepsilon$$

for arbitrary $\varepsilon > 0$, whence $\mathbf{E}X$ is convex.

Now establish the relationship to support functions. Let $x \in \mathbf{E}X$. Then there exists a sequence ξ_n of selections such that $\mathbf{E}\xi_n \to x$ as $n \to \infty$. Furthermore

$$h_{\{x\}}(u) = \lim_{n\to\infty} \langle \mathbf{E}\xi_n, u\rangle = \lim_{n\to\infty} \mathbf{E}\langle \xi_n, u\rangle \le \mathbf{E}h_X(u)\,.$$

Finally, for each unit vector u and $\varepsilon > 0$ define a half-space as

$$Y_\varepsilon = \{x :\ \langle x, u\rangle \ge h_X(u) - \varepsilon\}\,.$$

Then $Y_\varepsilon \cap X$ is a non-empty random closed set, which has a selection ξ_ε, such that

$$h_{\{\xi_\varepsilon\}}(u) \ge h_X(u) - \varepsilon\,.$$

Taking the expectation confirms that $h_{\mathbf{E}X}(u) \ge \mathbf{E}h_X(u)$. $\square$

In other words, the expectation of a random set is a convex set whose support function equals the expected support function of X. The convexifying effect of the selection expectation limits its applications in such areas like image analysis, where it is sometimes essential to come up with averaging scheme for images, see [20, Sec. 2.2] for a collection of further definitions of expectations. However, it appears very naturally in the law of large numbers for random closed sets as described in the following section.

Example 4.4. Let $X = B_\xi(\eta)$ be the closed ball of radius $\xi > 0$ centred at $\eta \in \mathbb{R}^d$, where both ξ and η are integrable. Then its expectation is the ball of radius $\mathbf{E}\xi$ centred at $\mathbf{E}\eta$.

Exercise 4.5. Show that $\mathbf{E}X = \{a\}$ is a singleton if and only if X is a random singleton itself, i.e. $X = \{\xi\}$.

Exercise 4.6. Assume that X is isotropic, i.e. X coincides in distribution with its arbitrary rotation around the origin. Identify $\mathbf{E}X$.

Sometimes it is necessary to check if a given point x belongs to a convex set, e.g. to the Aumann expectation $\mathbf{E}X$. This can be done by comparing the support functions of x and of $\mathbf{E}X$. Namely $x \in \mathbf{E}X$ if and only if

$$\sup_{u\in\mathbb{R}^d} \Big[\langle x, u\rangle - h_{\mathbf{E}X}(u)\Big] = 0\,.$$

Alternatively, the supremum can be taken over all u with norm one, in which case the supremum is less than or equal to zero.

4.2 Law of large numbers and the central limit theorem for Minkowski sums

The most common distance on the family $\mathcal{K}$ of compact sets is the Hausdorff distance defined as

$$\rho_H(K,L) = \inf\{r > 0:\ K \subset L^r,\ L \subset K^r\}\,,$$

where K^r denotes the closed r-neighbourhood of K, i.e. the set of all points within distance r from K. Recall that the *Minkowski sum* of two compact sets K and L is defined as

$$K + L = \{x + y:\ x \in K,\ y \in L\}\,.$$

In particular K^r is the Minkowski sum of K and the closed ball of radius r, centred at the origin. The same definition applies if one of the summands is compact and the other is closed. However if the both summands are closed (and not necessarily compact), then the sum is not always closed and one typically inserts the closure in the definition.

Support functions linearise the Minkowski sum, i.e.

$$h_{K+L}(u) = h_K(u) + h_L(u)\,, \quad u \in \mathbb{R}^d\,.$$

The homogeneity property of support functions makes it possible to define them only on the unit sphere $\mathbb{S}^{d-1}$ in $\mathbb{R}^d$. Then the uniform metric for support functions on the sphere turns into the Hausdorff distance between compact sets. Namely

$$\rho_H(K,L) = \sup_{u\in\mathbb{S}^{d-1}} |h_K(u) - h_L(u)|$$

and also

$$\|K\| = \rho_H(K,\{0\}) = \sup_{u\in\mathbb{S}^{d-1}} |h_K(u)|\,.$$

Consider a sequence of i.i.d. random compact sets $X_1, X_2, \dots$ all distributed as a random compact set X. It should be noted that the mere existence of such sequence implies that the probability space is non-atomic.

Theorem 4.7 (Law of large numbers for random sets, see [1])**.** *If $X, X_1, X_2, \dots$ are i.i.d. integrably bounded random compact sets and $S_n = X_1 + \cdots + X_n$, $n \geq 1$, are their Minkowski sums, then*

$$\rho_H(n^{-1}S_n, \mathbf{E}X) \to 0 \quad \textit{a.s. as} \quad n \to \infty\,.$$

Proof. Let us prove the result assuming that X is almost surely convex. Then

$$h_{n^{-1}S_n}(u) = \frac{1}{n}\sum_{i=1}^{n} h_{X_i}(u) \to \mathbf{E}h_X(u) = h_{\mathbf{E}X}(u)\text{a.s. as } n \to \infty$$

by a strong law of large numbers in a Banach space specialised for the space of continuous functions on the unit ball with the uniform metric. The uniform metric on this space corresponds to the Hausdorff metric on convex compact sets, whence the strong law of large numbers holds.

In order to rid of the convexity assumption we rely on the following result known as Shapley–Folkman–Starr theorem. If $K_1, \dots, K_n$ are compact subsets of $\mathbb{R}^d$ and $n \geq 1$, then

$$\rho_H(K_1 + \cdots + K_n, \operatorname{conv}(K_1 + \cdots + K_n)) \leq \sqrt{d} \max_{1\leq i\leq n} \|K_i\|\,.$$

Note that the number of summands does not appear in the factor on the right-hand side. For instance, if $K_1 = \cdots = K_n = K$, then one obtains that the distance between the sum of n copies of K and the sum of n copies of the convex hull of K is at most $\sqrt{d}\|K\|$.

A not necessarily convex X can be replaced by its convex hull $\mathrm{conv}(X)$, so that it remains to show that

$$n^{-1}\rho_H(K_1+\cdots+K_n, \mathrm{conv}(K_1+\cdots+K_n)) \le \frac{\sqrt{d}}{n}\max_{1\le i\le n}\|X_i\| \to 0 \quad \text{a.s. as } n\to\infty\,.$$

The latter follows from the integrable boundedness of X. Indeed, then we have $n\mathbf{P}(\|X\|>n)\to 0$ as $n\to\infty$, and

$$\mathbf{P}(\max_{i=1,\dots,n}\|X_i\|\ge nx) = 1-(1-\mathbf{P}(\|X\|\ge nx))^n \to 0\,.$$

□

Numerous generalisations of the above strong law of large numbers deal with random subsets of Banach spaces and possibly unbounded random closed sets in the Euclidean space, see [20].

The formulation of the central limit theorem is complicated by the fact that random sets with expectation zero are necessarily singletons. Furthermore, it is not possible to define Minkowski subtraction as the opposite operation to the addition. For instance, it is not possible to find a set that being added to a ball produces a triangle. Therefore, its not possible in general to normalise successive sums of random compact sets.

Note that the classical limit theorem may be (a bit weaker) formulated as the convergence of the normalised distance between the empirical mean and the expectation to the absolute value of a normally distributed random variable. In other words, it is possible to replace subtraction of two numbers with a distance between them. This idea is used to formulate a central limit theorem for Minkowski sums of random sets.

In order to formulate a limit theorem for random closed sets we need to define a centred Gaussian random field $\zeta(u)$ on the unit sphere $\mathbb{S}^{d-1}$ in $\mathbb{R}^d$ which shares the covariance structure with the random closed set X, i.e.

$$\mathbf{E}[\zeta(u)\zeta(v)] = \mathrm{cov}(h_X(u), h_X(v))\,, \quad u,v\in\mathbb{S}^{d-1}\,.$$

Since the support function of a compact set is Lipschitz, it is possible to show that the random field ζ has a continuous modification.

Theorem 4.8 (Central Limit Theorem, see [31])**.** *Let $X_1, X_2, \dots$ be i.i.d. copies of a random closed set X in $\mathbb{R}^d$ such that $\mathbf{E}\|X\|^2<\infty$. Then $\sqrt{n}\rho_H(n^{-1}S_n, \mathbf{E}X)$ converges in distribution as $n\to\infty$ to $\sup\{|\zeta(u)|:\ u\in\mathbb{S}^{d-1}\}$.*

Proof. For convex random sets the result follows from the central limit theorem for continuous random functions on the unit sphere. The general non-convex case is proved by an application of the Shapley–Folkman–Starr theorem. □

It is not clear how to interpret geometrically the limit $\zeta(u)$, $u\in\mathbb{S}^{d-1}$, in the central limit theorem for random sets.

One can attempt to define Gaussian random sets as those whose support function becomes a Gaussian process on $\mathbb{S}^{d-1}$. All such sets have however degenerate distributions. Namely, X is a

Gaussian random set if and only if $X = \xi + M$, where ξ is a Gaussian random vector in $\mathbb{R}^d$ and M is a deterministic convex compact set. This is seen by noticing that the so-called *Steiner point*

$$s(X) = \frac{1}{\kappa_d} \int_{\mathbb{S}^{d-1}} h_K(u)\, u\, du \tag{4.2}$$

is a Gaussian random vector that a.s. belongs to X, where κ_d is the volume of the d-dimensional unit ball. Thus, $M = X - \xi$ with $\xi = s(X)$ has Gaussian non-negative support function, which is then necessarily degenerated, so that M is deterministic.

Note that the Steiner point provides a natural selection of X. Further selections can be obtained by changing the measure used in the right-hand side (4.2) to a more general measure on the unit sphere.

4.3 Zonoids

One particularly important example of expectations of random sets appears if one takes expectation of random segments. Let $X = [0, \xi]$ be a segment in $\mathbb{R}^d$ with end-points being the origin and ξ. Note that possible translations of X only result in translations of the expectation and so do not influence its shape.

Assume that ξ is integrable. Then the selection expectation of X is a convex compact set Z_ξ with the support function

$$h_{Z_\xi}(u) = \mathbf{E} \max(0, \langle u, \xi \rangle) = \mathbf{E}\langle u, \xi \rangle_+ \,,$$

where a_+ denotes $\max(0, a)$, i.e. the positive part of $a \in \mathbb{R}$. The set Z_ξ is called a *zonoid* of ξ.[11]

The support function of zonoid equals the expectation of $(u_1\xi_1 + \cdots + u_d\xi_d)_+$. If $\xi_1, \ldots, \xi_d$ are prices and $u_1, \ldots, u_d$ are weights, then this expectation (if taken with respect to a martingale measure) becomes the price of an exchange option on d assets. By considering $(1, \xi)$, i.e. the extended (lifted) variant of ξ, it is possible to interpret the prices of all basket options. Indeed, the expectation of $[0, (1, \xi)]$ being a random set in $\mathbb{R}^{d+1}$ is a convex set $\hat{Z}_\xi$ in $\mathbb{R}^{d+1}$ called the lift zonoid of ξ and having the support function

$$h_{\hat{Z}_\xi}(u) = \mathbf{E}(u_0 + \langle u, \xi \rangle)_+ \,, \qquad u_0 \in \mathbb{R},\ u \in \mathbb{R}^d \,.$$

In finance, the value of u_0 is called a strike.

Exercise 4.9. Prove that the values of $\mathbf{E}(\xi - k)_+$ for an integrable random variable ξ and all $k \in \mathbb{R}$ determine uniquely the distribution of ξ.

While the lift zonoid of ξ uniquely determines the distribution of ξ (first it determines the distribution of $\langle u, \xi \rangle$ and then of ξ itself by the Cramér–Wold device), the zonoid Z_ξ does not uniquely determines the distribution of ξ. For example, if ξ is a random variable, then its zonoid is the segment with end-points being the expectations of the positive and negative parts of ξ.

[11] In general, a zonotope is the Minkowski sum of a finite number of segments, i.e. the expectation of a random segment that takes only a finite number of possible values. A zonoid is the limit of zonotopes in the Hausdorff metric and so is the expectation of a random segment, possibly with none of the end-points being zero. However, it is always possible to translate zonoid, so that it becomes the expectation of $[0, \xi]$ for some ξ.

Example 4.10 (Interval regression, see [6]). In the setting of Example 1.1 the estimators of the slope θ_1 and intercept θ_2 are obtained as

$$(\theta_1, \theta_2) = \Sigma(x)^{-1} \begin{pmatrix} \mathbf{E}(y) \\ \mathbf{E}(xy) \end{pmatrix}, \qquad \Sigma(x) = \begin{bmatrix} 1 & \mathbf{E}x \\ \mathbf{E}x & \mathbf{E}x^2 \end{bmatrix}.$$

Since y is a selection of $Y = [y_L, y_U]$, xy is a selection of xY and so the pair (y, xy) is a selection of random segment

$$G = \left\{ \begin{pmatrix} y \\ xy \end{pmatrix} : \ y_L \le y \le y_U \right\} \subset \mathbb{R}^2 .$$

Thus, the least squares estimator of $\theta = (\theta_1, \theta_2)$ is given by $\Sigma(x)^1 \bar{G}_n$ where $\bar{G}_n$ is the average of n segments, so that $\bar{G}_n$ estimates the selection expectation $\mathbf{E}G$.

Exercise 4.11. Let $(x, y) \in X$, where X is a random set in the plane. For instance, if y is interval-valued, then X is a vertical segment, if both x and y are interval-valued, then X is a rectangle, etc. The aim is to characterise the ordinary least square regression of y onto x if a sample of random set $X_1, \dots, X_n$ is observed.

5 Heavy tails, stability and unions

5.1 Stable random sets

As shown in Section 4.2, Gaussian random sets have degenerate shape. A general α-stable random set X satisfies

$$a^{1/\alpha} X_1 + b^{1/\alpha} X_2 \sim (a+b)^{1/\alpha} X \tag{5.1}$$

for all $a, b > 0$, where X_1, X_2, X are i.i.d. random sets. In particular, Gaussian random set X appears if $\alpha = 2$. By centring with a random vector, it is possible to assume that X contains the origin. Then writing (5.1) for the support function of X we arrive at the conclusion that $h_X(u)$ is α-stable non-negative random variable. If $\alpha \in [1, 2]$, this is possible only if $h_X(u)$ is constant and so X is a random translation of a deterministic set as in the Gaussian case.

However, nontrivial α-stable random sets exist if $\alpha \in (0, 1)$. Such sets can be constructed as the sum of series

$$X = c \sum \Gamma_k^{-1/\alpha} Z_k ,$$

where $\Gamma_k = \zeta_1 + \cdots + \zeta_k$ with i.i.d. standard exponential $\zeta_1, \zeta_2, \dots$ and $Z_1, Z_2, \dots$ are i.i.d. random compact sets.

Stable random sets necessarily have heavy-tailed distribution, e.g. if X is α-stable with $\alpha \in (0, 1)$, then $h_X(u)$ is not integrable for each u. In other words, some points of X with high probability have large norm. Heavy tailed random sets can be constructed by taking convex hull of a Poisson point process whose intensity decays sufficiently slow at the infinity. In contrast, if the intensity decays sufficiently slow, e.g. like in the Gaussian case, then the convex hull of points is not heavy-tailed.

5.2 Unions of random sets

While the arithmetic summation scheme for random variables gives rise to the Gaussian distribution in the limit, the maximum of random variables gives rise to extreme value distributions. Along the same line, the Minkowski summation scheme for random sets being singletons reduces to the classical limit theorem for sums of random vectors, while taking unions of random sets generalises the maximum (or minimum) scheme for random variables. Notice that if $X_i = (-\infty, \xi_i]$, $i = 1, 2 \ldots$, then

$$X_1 \cup \cdots \cup X_n = (-\infty, \max(\xi_1, \ldots, \xi_n)] \,.$$

Let $X, X_1, X_2, \ldots$ be a sequence of i.i.d. random closed sets in $\mathbb{R}^d$ and let $a_n > 0$, $n \geq 1$, be a sequence of non-negative normalising constants. The weak convergence of the random set

$$Z_n = a_n(X_1 \cup \cdots \cup X_n)$$

to a random closed set Z is defined by specialising the general concept of weak convergence of probability measures to the space $\mathcal{F}$ of closed sets. In particular, a necessary and sufficient condition for this is the convergence of capacity functionals on sets K such that Z touches the boundary of K with probability zero. The capacity functional of the set Z_n is easy to find as

$$T_{Z_n}(K) = 1 - (1 - T_X(a_n^{-1}K))^n \,.$$

Various convergence results for unions of random sets can be found in [20, Ch. 4].

Here we will discuss properties of random sets that can appear in the limit. In a more general triangular array schemes the limits are union-infinitely-divisible, while in the above described scheme the limit Z is necessarily union-stable.

Definition 5.1. A random closed set is said to be *union-infinitely divisible* if Z coincides in distribution with the union of i.i.d. random closed sets $Z_{11}, \ldots, Z_{nn}$ for each $n \geq 2$.
A random closed set Z is said to be *union-stable* if Z coincides in distribution with $a_n^{-1}(Z_1 \cup \cdots \cup Z_n)$ for each $n \geq 2$ with normalising constants $a_n > 0$, where $Z_1, \ldots, Z_n$ are i.i.d. copies of Z.

Exercise 5.2. Assume that Z coincides in distribution with the union of its n i.i.d. copies for some $n \geq 2$. Show that such Z is necessarily deterministic.

For the following it is essential to single out the deterministic part of a random set. A point x is said to be a fixed point of X if $x \in X$ with probability one. The set of fixed points is denoted by F_X. For instance, if $X = (-\infty, \xi]$ with exponentially distributed ξ, then $F_X = (-\infty, 0]$, while F_X is empty if ξ is normally distributed.

Theorem 5.3. *A random closed set X is union-infinitely-divisible if and only if there exists a completely alternating upper semicontinuous functional $\psi : \mathcal{K} \to [0, \infty]$ such that*

$$T(K) = 1 - e^{-\psi(K)} \,, \quad K \in \mathcal{K} \,,$$

and $\psi(K) < \infty$ whenever $K \cap F_X = \emptyset$.

Example 5.4. Let X be the Poisson point process with intensity measure Λ. Then $T(K) = 1 - e^{-\Lambda(K)}$ and X is union-infinitely-divisible. Indeed, X equals in distribution the union of n i.i.d. Poisson processes, each with intensity measure $n^{-1}\Lambda$.

The functional ψ shares nearly the same properties as T apart from the fact that the values of ψ are no longer required to lie in $[0, 1]$. The functional ψ defines a locally measure μ on $\mathcal{F}$, such that $\mu(\{F : F \cap K \neq \emptyset\}) = \psi(K)$. The measure μ defines a Poisson processes on $\mathcal{F}$ such that the "points" of this process are actually closed sets. Then X is the union of Poisson process on $\mathcal{F}$ with intensity measure μ.

Theorem 5.5. *A random closed set is union-stable if and only if its capacity functional admits representation $T(K) = 1 - e^{-\psi(K)}$ with ψ being homogeneous, i.e.*

$$\psi(sK) = s^\alpha \psi(K), \quad K \in \mathcal{K},\ K \cap F_X = \emptyset,$$

for some $\alpha \neq 0$ and all $s > 0$, and also $F_X = sF_X$ for all $s > 0$.

The proof of the above theorem relies on solving some functional equations for capacity functionals of random sets, quite similar to the corresponding characterisation of max-stable random variables. The major complication stems from the fact that for any random variable ξ the equivalence of the distributions of ξ and $c\xi$ immediately implies that $c = 1$. This is however not the case for random sets, e.g. $X = \{t \geq 0 : w_t = 0\}$, the set of zeros for the standard Brownian motion, coincides in distribution with cX for each $c > 0$. Another example of such X is a randomly rotated cone in $\mathbb{R}^d$. The key step in the proof of Theorem 5.5 aims to show that the union-stability property rules out all such self-similar random sets.

Exercise 5.6. Let $X = (-\infty, \xi]$ on $\mathbb{R}$. Characterise all distributions of ξ that correspond to union-stable X and confirm that these distributions are exactly extreme value distributions of Fréchet and Weibull type.

If X is a Poisson process, then its union-stability property implies that the intensity of the process is homogeneous, i.e. $\Lambda(sx) = s^{\alpha-d}\lambda(x)$ for all $x \in \mathbb{R}^d$ and $s > 0$.

5.3 The Boolean model

The Poisson process gives rise to an important model of random sets that produces random sets which are infinitely divisible for unions. In order to define the so-called Boolean model of random sets, consider a Poisson process Π in $\mathbb{R}^d$ with intensity measure Λ and a sequence of i.i.d. random compact sets $X_0, X_1, X_2, \ldots$. One calls the points x_i of Π *germs* and the random sets X_i *grains*. Then consider the union of grains placed at locations specified by the germs, i.e. define

$$X = \bigcup_{x_i \in \Pi} (x_i + X_i).$$

In order to ensure that the union remains a closed set, one requires that the typical grain X_0 grains is not too big. For instance, if X_0 almost surely contains the origin, the condition $\mathbf{E}\|X_0\|^d < \infty$ guarantees that X is a random closed set, which is then called the *Boolean model*.[12] If Λ is the Lebesgue measure, then X is said to be a *stationary* Boolean model, which is by now the best understood model of a stationary random closed set.

Recall that $\check{K} = \{-x : x \in K\}$ denotes the centrally symmetric set to $K \subset \mathbb{R}^d$.

Theorem 5.7. *If X is a Boolean model, then*

$$T_X(K) = 1 - \exp\{-\mathbf{E}\Lambda(\check{X}_0 + K)\}, \quad K \in \mathcal{K}.$$

[12]The necessary and sufficient condition for the closedness of X is the finiteness of $\mathbf{E}\Lambda(X_0^r)$ for some $r > 0$, where X_0^r is the closed r-neighbourhood of X_0.

Proof. Let Π'_Λ be the set of all points $x_n \in \Pi_\Lambda$ with $(X_n + x_n) \cap K \neq \emptyset$. In other words, Π'_Λ is a result of a thinning procedure applied to the Poisson point process Π_Λ. Then each point $x \in \Pi_\Lambda$ belongs also to Π'_Λ with probability

$$g(x) = \mathbf{P}\{(X_0 + x) \cap K \neq \emptyset\} = \mathbf{P}\{x \in \check{X}_0 + K\}.$$

Then, by the Poisson property and Fubini's theorem,

$$\begin{aligned} T_X(K) = \mathbf{P}\{X \cap K \neq \emptyset\} &= 1 - \mathbf{P}\{\Pi'_\lambda = \emptyset\} \\ &= 1 - \exp\left\{-\int_{\mathbb{R}^d} g(x)\Lambda(dx)\right\} \\ &= 1 - \exp\{-\mathbf{E}\Lambda(\check{X}_0 + K)\}. \end{aligned}$$

□

Example 5.8. Let X_0 be a ball of radius η such that $\mathbf{E}\eta^d < \infty$. Then the stationary Boolean model exists and has the capacity functional

$$T_X(K) = 1 - \exp\{-\lambda \mathbf{E}|K^\eta|\},$$

where $|\cdot|$ denotes the Lebesgue measure and λ is the intensity of the stationary Poisson process of germs. If K is convex, then the Lebesgue measure of K^η is the polynomial of order d in η whose coefficients depend on the geometry of K.

While it is rather easy to find the hitting probability for the Boolean model, it is considerably more complicated task to find the probability that X covers a non-finite set K.

Exercise 5.9. Find the probability $p(x) = \mathbf{P}\{x \in X\}$ for the Boolean model X, which is called the coverage function of X. By Robbins' formula find the expected Lebesgue measure of $X \cap W$ for an observation window W.

5.4 Other expectations of sets

It should be noted that the selection expectation is well suited to deal with convex random sets, in view of its convexifying effect. There exist numerous other definitions of expectation for random sets which do not have this convexifying effect.

For instance, if a random set X is star-shaped, i.e. $x \in X$ implies that the segment $[0, x]$ is contained in X, then X can be described by its radial function $r_X(u)$ that depends on direction u and so the expected radial function $\mathbf{E}r_X(u)$ becomes the radial function of a deterministic star-shaped set called the star-shaped expectation of X. This definition however heavily depends on the relative position of X with respect to the origin.

Numerous further definitions of expectations are discussed in [20].

6 Set-valued random functions

6.1 Examples of set-valued processes

A family of random sets X_t indexed by time (discrete or continuous) is naturally called a *set-valued process.* The time argument may also represent or include other covariates. A single-valued stochastic process ξ_t such that $\xi_t \in X_t$ a.s. for all t is called a selection of the set-valued process. Set-valued processes (or processes of random sets) are discussed by [13] and [20, Ch. 5].

Example 6.1 (Growth model)**.** Let $X_1, X_2, \dots$ be random closed sets, for instance, independent identically distributed. Then it is possible to define an increasing set-valued process (sequence) by taking unions

$$Z_n = X_1 \cup \dots \cup X_n$$

or convex hulls like $\mathrm{conv}(X_1, \dots, X_n)$ or Minkowski sums

$$Y_n = X_1 + \dots + X_n \,.$$

The limit theorems for Minkowski sums and unions clarify the limiting behaviour of so defined processes.

Note that it is considerably more difficult to define tractable models of set-valued processes that also may decrease with positive probability.

Example 6.2. Consider a random game that depends on some parameters. The Nash equilibrium has a closed graph as function of parameters, see [8, Sec. 1.3.2]. Thus, in case of random game, the set of equilibria becomes a set-valued process.

Example 6.3. Let ξ_t, $t \in \mathbb{R}^d$, be real-valued upper semicontinuous stochastic process. Then

$$X_a = \{t : \ \xi_t \le a\}\,, \quad a \in \mathbb{R}\,,$$

is an increasing set-valued process indexed by the level a. For instance, if ξ_t is an estimate for the cumulative distribution function of a random vector, then X_a represent a possible related quantile process. It is possible to define addition of functions by adding the corresponding level sets, so that the so-called level sum of functions is a function whose level sets equal the Minkowski sums of level sets of the two summands. The law of large numbers for Minkowski sums then yield the law of large numbers for so defined sums of functions.

A rich source of set-valued functions is provided by solutions of differential inclusions, see [2]. Let x_t, $t \in \mathbb{R}$, be a differentiable function. Then

$$\frac{dx}{dt} \in F(x, t)$$

is called a *differential inclusion*, where F is a set-valued function of x and t. If F is random, then the family of solutions $X_t = \{x_t\}$ for each given t builds a set-valued process. It should be noted that not all functions y_t such that $y_t \in X_t$ for all t are solutions of the differential inclusions, since such functions may be very irregular.

It is also possible to consider stochastic differential equations with set-valued coefficients. However, since there is no natural analogue of the set-valued Brownian motion, there is no analogue of the set-valued stochastic differential.

6.2 Set-valued martingales

An integrable set-valued process Z_t with discrete time $t = 0, \dots, T$ or continuous time $t \in [0, T]$ is said to be a set-valued $(\mathfrak{F}_t)$-martingale if $\mathbf{E}(Z_t|\mathfrak{F}_s) = Z_s$ a.s. for all $0 \le s \le t \le T$. The conditional expectation of Z_t is defined as the convex set whose support function is the conditional expectation of the support function of Z_t, see [20, Sec. 2.1.6] and [11] for the case of unbounded

Z_t. It is well known (see [12] and [20, Sec. 5.1.1]) that a *set-valued martingale* admits at least one martingale selection and, moreover, it has a countable dense family of martingale selections, the so-called *Castaing representation*, see [20, Th. 5.1.12]. In the following we consider only the case of the discrete time.

Note that a supermartingale is defined by the inclusion relation $\mathbf{E}(Z_t|\mathfrak{F}_s) \subset Z_s$ a.s. for all $0 \le s \le t \le T$ and a submartingale by $\mathbf{E}(Z_t|\mathfrak{F}_s) \supset Z_s$ a.s. This situation is quite typical for set-valued process obtained as convex hulls of martingales, since the supremum of martingales naturally yields a submartingale. In difference to the classical case of real-valued martingales, it is not possible to pass from a super- to submartingale by considering $-Z_t$.

For instance, consider geometric Brownian motions $\zeta_t = s_0 \exp\{\sigma W_t - \sigma^2 t/2\}$ with volatility σ that belongs to an interval $[\sigma', \sigma'']$, see [3] and [24] for a discussion of such partially specified volatility models. The union Z_t of all their paths is an interval-valued submartingale, see [18, Ex. 9.2]. Its upper bound $\zeta_t^a = \sup Z_t$ is a numerical submartingale, while the lower bound $\zeta_t^b = \inf Z_t$ is a numerical supermartingale. Applying the Doob decomposition theorem to the both of them it is possible to come up with an interval-valued martingale that is the largest set-valued martingale inscribed in Z_t for all t. In spaces of dimension 2 and more the problem of finding the largest set-valued martingale inscribed in a set-valued process is not yet studied.

The specific nature of the financial setting calls for the study of set-valued processes that are not necessarily martingales themselves or are martingales with respect to another (risk-neutral) probability measure $\mathbf{Q}$ equivalent to $\mathbf{P}$.

Consider now the mere question of the existence of a $\mathbf{Q}$-martingale selection ζ_t of a set-valued process Z_t, where $\mathbf{Q}$ is equivalent to the underlying probability measure $\mathbf{P}$. We then call ζ an *equivalent martingale selection*.

Rokhlin [29] studied the existence of a martingale selection for a sequence of relatively open convex set-valued process X_t with discrete time $t = 0, \dots, T$. The main results of [29] establishes that the existence of martingale selection is equivalent to the a.s. non-emptiness of the recursively defined set-valued process W_t, $t = 0, \dots, T$. One sets W_T to be the closure of X_T and defines W_{t-1} to be the closure of X_{t-1} intersected with the relative interior of Y_{t-1}, the latter being the convex hull of the support of the conditional expectation of W_t given $\mathfrak{F}_{t-1}$. The *support* for a random set is defined by [28] as the union of the supports for all selections of the random set.

6.3 Transaction costs

As conventional martingales are extremely important for mathematical finance, the set-valued martingales also can be used to bring financial models closer to reality by accounting for transaction costs.

Transaction costs in a currency market with d traded currencies can be described by means of a *bid-ask matrix* $\Pi = (\pi^{ij})_{ij=1,d}$ where π^{ij} is the number of units of currency i (or, more generally, an asset number i) needed as a payment for one unit of currency j, see [16] and [30]. It is always assumed that Π has all positive entries, the diagonal entries are all 1 and $\pi^{ij} \le \pi^{ik}\pi^{kj}$, the latter means that a chain of exchanges never beats the result of a direct transaction.

In the time-dependent stochastic setting the bid-ask matrix Π_t is a random adapted matrix-valued process that depends on (discrete) time and is called a *bid-ask process*. By discounting prices, it is always assumed that all interest rates are zero.

A currency portfolio is a vector $x \in \mathbb{R}^d$ that represents the number of physical units of the currencies held by an investor. The set of portfolios available at price zero is the cone $-\hat{K}(\Pi)$ in

$\mathbb{R}^d$ spanned by the negative basis vectors $-e_i$, $1 \le i \le d$, and the vectors $-\pi^{ij}e_i + e_j$ for all i and j. The latter portfolios can be realised by borrowing π^{ij} units of asset i and buying with this one unit of j. This portfolio can be liquidated by buying π^{ij} units of asset i with at most $\pi^{ji}\pi^{ij} \le \pi^{jj} = 1$ unit of asset j. The cone of portfolios available at price zero is centrally symmetric to the *solvency cone* $\hat{K}(\Pi)$, see [9] and [30]. The hats in these notation indicate that the elements of the cones represent physical units of the assets.

Consider any x from the cone $-\hat{K}(\Pi)$. If the assets were priced at $s = (s_1, \dots, s_d)$, then this portfolio would cost $\langle x, s\rangle = \sum x_i s_i$, which should be at most zero. Thus, s might be used as a price system if and only if $\langle x, s\rangle \le 0$ for all $x \in -\hat{K}(\Pi)$. In other words, all *consistent price systems* form a cone $K^*(\Pi)$, which is polar to $-\hat{K}(\Pi)$. The cone $K^*(\Pi)$ is also called the positive dual cone to the solvency cone $\hat{K}(\Pi)$. Note that $K^*(\Pi) \setminus \{0\}$ is a subset of the interior of $\mathbb{R}^d_+$. Later on we abbreviate $\hat{K}(\Pi_t)$ as $\hat{K}_t$ with a similar agreement for other related cones.

Having started with zero initial endowment, by trading in one time unit it is possible to arrive at a portfolio which belongs to $-\hat{K}_0$. If we write $L^0(G, \mathfrak{F})$ for the family of all $\mathfrak{F}$-measurable random vectors with values in a set $G \subset \mathbb{R}^d$, then in T time units one can arrive at any portfolio from

$$\hat{A}_T = \sum_{t=0}^{T} L^0(-\hat{K}_t, \mathfrak{F}_t)\,,$$

which is called the set of *attainable portfolios*. The sum in the right-hand side is the Minkowski sum of subsets of $L^0(\mathbb{R}^d, \mathfrak{F}_T)$. The major issue then is to determine whether $\hat{A}_T$ might contain portfolios with a positive value, which then would lead to a sure profit without investment. The typical way to prove such results is to show the closedness of the cone $\hat{A}_T$ in L^0 and then apply the separation argument. Note that the Minkowski sum of closed but non-compact sets in a linear space is not necessarily closed.

The bid-ask process $(\Pi_t)_{t=0}^T$ is said to satisfy the *no-arbitrage* property (NA) if the intersection of $\hat{A}_T$ and the family $L^0(\mathbb{R}^d_+, \mathfrak{F}_T)$ of non-negative $\mathfrak{F}_T$-measurable random vectors is exactly $\{0\}$. In order to avoid arbitrage for the agents, who might profit from transaction costs by cancelling the effects of opposite operations and pocketing the transaction costs on the both, the sequence of portfolios is sometimes assumed to be increasing, see [14] and [18], i.e. both the short and long positions are assumed to be non-decreasing in t. However, this assumption is generally superfluous.

It is apparent that attainable portfolios in this model of transaction costs are random vectors that take values in certain polyhedral cones. These cones are determined by (possibly random) bid-ask matrices and so can be treated as random polyhedral closed cones in $\mathbb{R}^d$. Grigoriev [9] showed that for two assets (i.e. for $d = 2$) the no-arbitrage property is equivalent to the existence of an $(\mathfrak{F}_t)$-martingale ζ_t that a.s. takes values in $K^*_t \setminus \{0\}$ for all $t = 0, \dots, T$. This martingale $(\zeta_t)_{t=0}^T$ is called a *consistent price process* associated with $(\Pi_t)_{t=0}^T$. Counterexamples confirm that this result does not hold for $d \ge 4$, see [30], and for all $d \ge 3$ in case of general polyhedral cones K^*_t not necessarily associated with bid-ask matrices, see [9, Ex. 5.1]. The case of bid-ask matrices and $d = 3$ is still open. In general, an equivalent interpretation of the no-arbitrage condition using consistent price processes in the space of arbitrary dimension $d \ge 3$ is not yet known.

In view of this, [30] has shown the equivalence of two stronger statements: the robust no-arbitrage property and the existence of a strictly consistent price process. The *robust no-arbitrage* means the existence of a bid-ask process $\tilde{\Pi}$ with smaller bid-ask spreads

$$\frac{1}{\pi^{ji}} < \frac{1}{\tilde{\pi}^{ji}} < \tilde{\pi}^{ij} < \pi^{ij}\,, \quad 1 \le i \ne j \le d\,,$$

which also satisfies the no-arbitrage property, meaning that the agent might offer some non-zero discounts on transaction costs while still ensuring no-arbitrage. A consistent price process ζ_t is said to be *strict* if ζ_t a.s. belongs to the relative interior of K_t^* for all t, where the *relative interior* of a set is the interior taken in the smallest affine subspace containing this set. The consistent price system $\tilde{K}_t^*$ generated by $\tilde{\Pi}$ is a subset of the relative interior of K_t^*. The appearance of the relative interior in this context is quite natural, since if K_t^* lies in a hyperplane, certain assets (or some combinations of them) can be freely exchanged. One says that the model exhibits *efficient friction* if K_t^* has non-empty interior for all t. In this case all exchanges involve positive transaction costs, the interior of K_t^* coincides with the relative interior and so each strictly consistent price process automatically belongs to the interior of K_t^*.

Assume now that the first asset represents a money account. Since this asset is traded without transaction costs, the prices of other assets can be represented as bid and ask prices in relation to the money account. Let $S_t^{a,i}$ and $S_t^{b,i}$ be the ask and bid prices (so that $[S_t^b, S_t^a]$ is the *bid-ask spread*) for asset $i = 2, \dots, d$ at time t. If we set $S_t^{a,1} = S_t^{b,1} = 1$ for all t, then it is possible to define a bid-ask matrix by setting $\pi_t^{ij} = S_t^{a,j}/S_t^{b,i}$. In the opposite direction, a bid-ask matrix admits a bid-ask spread representation with a money account if $\pi_t^{ij} = \pi_t^{i1}\pi_t^{1j}$ for all time moments t and every assets i and j. While this is always possible if $d = 2$, this assumption restricts the possible family of solvency cones in dimension $d \geq 3$.

If $d = 2$, i.e. for the bid-ask spread $[S_t^b, S_t^a]$ on a single asset, [9, Cor. 2.9] showed that the no-arbitrage condition is equivalent to the existence of a martingale S_t with respect to a probability measure equivalent to the original probability measure $\mathbf{P}$, such that $S_t \in [S_t^b, S_t^a]$ a.s. for all $t = 0, \dots, T$. This means that the interval-valued process $[S_t^b, S_t^a]$ possesses a martingale selection with respect to an equivalent measure. The earlier results in this direction go back to [15]. Although (for $d \geq 3$) this setting is more restrictive than the bid-ask matrix formulation, it corresponds to an intuitive perception of a price that belongs to some interval and is not simply rescalable as in the case of price systems given by cones. Furthermore, the family of all martingale selections can be used to determine the upper and lower prices of claims. These prices are obtained by taking supremum and infimum of the expectations of claim payoffs obtained by substituting in the payoff function all possible martingale selections.

For multiple assets with a money account, [14] considered *set-valued price processes* being rectangles or parallelepipeds. While this setting perfectly fits into the idea of bid-ask matrices, it does not take into account possible discounts for simultaneous transactions on several related assets. This *link-save* effect has been noticed by [18]. There it was assumed that prices of several assets at time t are described by a convex set Z_t in the first quadrant, so that the price of a combination of assets $u = (u_1, \dots, u_d) \in \mathbb{R}^d$ (expressed in physical units) is given by the support function $h_{Z_t}(u)$ of the set Z_t. The sublinearity of support functions corresponds again to the sublinearity property of prices, since the price of the combination $u' + u''$ does not exceed the sum of the individual prices of u' and u''. It is shown in [18] that the existence of a martingale selection of the set-valued process $(Z_t)_{t=0}^T$ in the square integrable case is equivalent to the no-arbitrage property. If we add a money account to this model, then the cone with base $\{1\} \times Z_t$ becomes an analogue of the cone K_t^* in Kabanov's model of transaction costs. It should be noted however that the corresponding cone is not a polyhedral cone any longer.

References

[1] Z. Artstein and R. A. Vitale. A strong law of large numbers for random compact sets. *Ann. Probab.*, 3:879–882, 1975.

[2] J.-P. Aubin and A. Cellina. *Differential Inclusions.* Springer, Berlin, 1984.

[3] M. Avellaneda, A. Levy, and A. Parás. Pricing and hedging derivative securities in markets with uncertain volatilities. *Appl. Math. Finance*, 2:73–88, 1995.

[4] A. Beresteanu, I. Molchanov, and F. Molinari. Sharp identification regions in games. Working Paper. Cornell University, http://cemmap.ifs.org.uk/wps/cwp1508.pdf, 2008.

[5] A. Beresteanu, I. Molchanov, and F. Molinari. Sharp identification regions in models with convex moment predictions. *Econometrica*, 2011.

[6] A. Beresteanu and F. Molinari. Asymptotic properties for a class of partially identified models. *Econometrica*, 76:763–814, 2008.

[7] R. M. Dudley. A course on empirical processes. In *École d'Été de Probabilités de Saint Flour XII*, volume 1097 of *Lect. Notes Math.*, pages 1–142. Springer, Berlin, 1984.

[8] D. Fudenberg and J. Tirole. *Game Theory.* MIT Press, Cambridge, MA, 1991.

[9] P. G. Grigoriev. On low dimensional case in the fundamental asset pricing theorem with transaction costs. *Statist. Decisions*, 23:33–48, 2005.

[10] L. Heinrich. Asymptotic properties of minimum contrast estimators for parameters of Boolean models. *Metrika*, 31:349–360, 1993.

[11] Ch. Hess. On multivalued martingales whose values may be unbounded: martingale selectors and Mosco convergence. *J. Multivariate Anal.*, 39:175–201, 1991.

[12] Ch. Hess. Conditional expectation and martingales of random sets. *Pattern Recognition*, 32:1543–1567, 1999.

[13] S. Hu and N. S. Papageorgiou. *Handbook of Multivalued Analysis*, volume 1. Kluwer, Dordrecht, 1997.

[14] E. Jouini. Price functionals with bid-ask spreads: an axiomatic approach. *J. Math. Econ.*, 34:547–558, 2000.

[15] E. Jouini and H. Kallal. Martingales and arbitrage in securities markets with transaction costs. *J. Econ. Th.*, 66:178–197, 1995.

[16] Yu. M. Kabanov. Hedging and liquidation under transaction costs in currency markets. *Finance and Stochastics*, 3:237–248, 1999.

[17] T. Kamae, U. Krengel, and G. L. O'Brien. Stochastic inequalitites on partially ordered spaces. *Ann. Probab.*, 5:899–912, 1977.

[18] K. Kaval and I. Molchanov. Link-save trading. *J. Math. Econ.*, 42:710–728, 2006.

[19] A. P. Korostelev and A. B. Tsybakov. *Minimax Theory of Image Restoration.* Springer, New York, 1993.

[20] I. Molchanov. *Theory of Random Sets.* Springer, London, 2005.

[21] I. S. Molchanov. A characterization of the universal classes in the Glivenko–Cantelli theorem for random closed sets. *Theory Probab. Math. Statist.*, 41:85–89, 1990. Translation from *Teor. Veroyatn. Mat. Stat.* 41: 74–78, 1989.

[22] I. S. Molchanov. *Statistics of the Boolean Model for Practitioners and Mathematicians.* Wiley, Chichester, 1997.

[23] I. S. Molchanov. A limit theorem for solutions of inequalities. *Scand. J. Statist.*, 25:235–242, 1998.

[24] P. A. Mykland. Financial options and statistical prediction intervals. *Ann. Statist.*, 31:1413–1438, 2003.

[25] R. Pyke. A uniform central limit theorem for partial sum processes indexed by sets. *London Math. Soc., Lect. Notes Ser.*, 79:219–240, 1983.

[26] Ph. Rigolet and R. Vert. Optimal rates in plug-in estimators of density level sets. *Bernoulli*, 15:1154–1178, 2009.

[27] B. D. Ripley and J.-P. Rasson. Finding the edge of a Poisson forest. *J. Appl. Probab.*, 14:483–491, 1977.

[28] D. B. Rokhlin. A martingale selection problem in the finite discrete-time case. *Theory Probab. Appl.*, 50:420–435, 2006.

[29] D. B. Rokhlin. A theorem on martingale selection for relatively open convex set-valued random sequences. *Math. Notes*, 81:543–548, 2007.

[30] W. Schachermayer. The fundamental theorem of asset pricing under proportional transaction costs in finite discrete time. *Math. Finance*, 14:19–48, 2004.

[31] W. Weil. An application of the central limit theorem for Banach-space-valued random variables to the theory of random sets. *Z. Wahrsch. verw. Gebiete*, 60:203–208, 1982.