On the connectedness of random sets of $\mathbb{R}$

Juan Jesús Salamanca
Escuela Politécnica de Ingeniería
Departamento de Estadística e I.O. y D.M.
Universidad de Oviedo
E-33071 Gijón (Spain)
salamancajuan@uniovi.es

Abstract

Random sets arises as distinguished models to study probability under imprecision. This work aims to determine the connectedness of a random set in terms of its capacity functional (the probability of hitting a set). This problem is linked to the celebrated Choquet-Kendall-Matheron theorem, which states that a random closed set is characterized by its capacity functional. Hence, such functional must also characterize any topological property of a random set, as its connectedness. The main disadvantage of the capacity functional is that it is evaluated over a large collection of sets (the family of compact sets). Naturally, we require that the determination of the connectedness must be made using a simple family of sets. We achieve a characterization adapted to these purposes.

Keywords: closed random set; connectedness; capacity functional.
2010 Mathematics subject classification: 52A22; 60B11; 57M15.

1 Introduction

There exist common situations where uncertainty about the probability distribution arises. The theory of Imprecise Probabilities [23] provides a natural scenario to study this kind of problems. In this context, the random sets appears as a model to deal with this imprecision. On a theoretical viewpoint, the random sets have been linked to belief functions [12, 18, 19], p-boxes [3] and fuzzy intervals [5, 6] (see also the classical references [1, 2]). For other approaches to imprecise probabilities, we can refer to [13].

The random intervals (i.e., random sets of $\mathbb{R}$) deserve their own place within random sets [2, 4, 10, 20]. These objects admit a nice interpretation: they can model the imprecision arising in the observation of a random variable [16].

In practice, the random intervals have been used in different clinical trials (for instance, in nausea [22]), in reforcement schedules [15], engineering [7], etc.

This work aims to study the connectedness of a closed random set of $\mathbb{R}$ in terms of simple evaluations of its capacity functional. Recall the seminal Choquet-Kendall-Matheron theorem, that states that the distribution of a random closed set is determined by its capacity functional (see [17] for instance). This fact encourages us to extend that philosophy to the topological properties of a random set; we focus here on the connectedness. It is worth to point out that this problem appears explicitly mentioned in [17, Open Problem 4.37], where applications to percolation theory can arise (see also [14] and [21]).

In broad terms, given a random set X, its capacity functional T_X is a map which associates a compact set with the probability that X intersects with that compact set [17]. Note that the capacity functional is valued over a very large family, the family of compact sets. This is a significant disadvantage in practice and also the main problem to solve in our framework. Then, we claim for a simple family of compact sets which allows us to determine whether X is connected. That claim is also motivated by practical reasons. In fact, the capacity functional is usually restricted to distinguished one-parametric families of compact sets (e.g. circles or segments in $\mathbb{R}^2$, [11]).

The closeness assumption on the random set is necessary. In fact, the capacity functional of a (non-closed) random set cannot determine whether the random set is connected. To prove it, we show two random sets of $[0,1]$, one of them is connected but the other one is not, and sharing the same capacity functional: Let U be a random variable with uniform distribution on $[0,1]$. Take X_1 to be the non-connected random set $X_1 = [0,1]\setminus \{U\}$. Take X_2 to be the deterministic set $[0,1]$. It is easy to see that X_1 and X_2 have the same capacity functional, which takes always the value 1 when it is valued on any compact set of $[0,1]$.

The next section is aimed to show and prove our main result, which answers positively to our requirements.

2 Main results

We begin recalling several notions of a random set that appear throughout this work. Fix a complete probability space $(\Omega, \mathfrak{A}, P)$, and denote by $\mathcal{F}$ the collection of the closed sets of $\mathbb{R}$. A *random closed set* X of $\mathbb{R}$ is a map defined over the probability space and with values on $\mathcal{F}$ such that for any compact set $K \subset \mathbb{R}$ it holds

$$\{\omega : X(\omega) \cap K \neq \emptyset\} \in \mathfrak{A}$$

(see [17]). This condition is interpreted as a measurability hypothesis of X as a map; moreover, this axiom is also linked to the capacity functional (see [9, 17] for a detailed discussion of these topics). We also assume that any random set X satisfies $X(\omega) \neq \emptyset$ for any $\omega \in \Omega$.

Now, we focus on the connectedness of a random closed set. We say that a random closed set X is *connected* if $P(\{\omega : X(\omega) \text{ is connected}\}) = 1$. We start with a technical result.

Proposition 2.1. *A random closed set X is connected if and only if for any $a, b \in \mathbb{R}$, $a < b$, the following equation holds*

$$P(\{\omega : X(\omega) \cap (-\infty, a] \neq \emptyset, X(\omega) \cap [a, b] = \emptyset, X(\omega) \cap [b, \infty) \neq \emptyset\}) = 0\,. \tag{1}$$

Proof. Fix $a, b \in \mathbb{R}$ with $a < b$. A $\omega \in \Omega$ satisfying

$$X(\omega) \cap (-\infty, a] \neq \emptyset, X(\omega) \cap [a, b] = \emptyset, X(\omega) \cap [b, \infty) \neq \emptyset$$

means that $X(\omega)$ is not connected. Hence,

$$\begin{aligned} &P(\{\omega : X(\omega) \cap (-\infty, a] \neq \emptyset, X(\omega) \cap [a, b] = \emptyset, X(\omega) \cap [b, \infty) \neq \emptyset\}) \\ &\qquad\qquad \leq P(\{\omega : X(\omega) \text{ is not connected}\})\,. \end{aligned}$$

Therefore, if X is connected, then the equation (1) must hold. To prove the converse, we will assume that X is not connected and the equation (1) holds to find a contradiction.

Given a Lebesgue-measurable set A of $\mathbb{R}$, denote by $\psi(A)$ the collection of all the connected components of A. Thus, we can write: $A = \bigcup_{B_i \in \psi(A)} B_i$, with each B_i connected and $B_i \cap B_j = \emptyset$ if $i \neq j$. Now, define a map $\|\cdot\|$ that associates A with the following non-negative number:

$$\|A\| := \sup_{B_i \in \psi(A)} \mu(B_i)\,,$$

where μ is the Lebesgue measure on $\mathbb{R}$. Moreover, $\|\emptyset\| = 0$ by definition.

Let X be a non-connected random closed set. There exists $\Omega' \subseteq \Omega$ with $P(\Omega') > 0$ such that $X(\omega)$ is not connected for any $\omega \in \Omega'$. Since X is measurable, $[\min X(\omega), \max X(\omega)] \setminus X(\omega)$ is so for any $\omega \in \Omega$. Note an abuse of notation since $\min X(\omega)$ can be $-\infty$ (analogously with $\max X(\omega)$). We have:

$$\|[\min X(\omega), \max X(\omega)] \setminus X(\omega)\| > 0$$

for any $\omega \in \Omega'$. Moreover, $\|[\min X(\omega), \max X(\omega)] \setminus X(\omega)\|$ is measurable as a function on Ω. With these considerations, we deduce:

$$\mathbb{E}\left[\|[\min X(\omega), \max X(\omega)] \setminus X(\omega)\|\right] > 0\,. \tag{2}$$

Now, assume that the equation (1) holds. Hence, the following equation holds for any $h > 0$:

$$\int_{x=-\infty}^{x=\infty} \int_\Omega \mathbb{I}\{\omega : X(\omega) \cap (-\infty, x] \neq \emptyset,$$
$$X(\omega) \cap [x, x+h] = \emptyset,$$
$$X(\omega) \cap [x+h, \infty) \neq \emptyset \} \qquad dP\, dx = 0\,.$$

Here, $\mathbb{I}\{\omega : \ldots\}$ represents the indicator function. Since the integrand is non-negative and measurable against $dP \otimes dx$, we apply the Fubini's theorem to obtain:

$$\begin{aligned}
\int_\Omega \int_{x=-\infty}^{x=\infty} \{\ldots\}\, dx dP &= \int_\Omega \max\{\|[\min X(\omega), \max X(\omega)] \backslash X(\omega)\| - h, 0\}\, dP \\
&= \mathbb{E}\left[\max\{\|[\min X, \max X] \backslash X\| - h, 0\}\right] \\
&\geq \mathbb{E}\left[\|[\min X, \max X] \backslash X\|\right] - h\,.
\end{aligned}$$

This means that $h \geq \mathbb{E}\left[\|[\inf X, \max X] \backslash X\|\right]$ for any $h > 0$. A contradiction is found with the equation (2). □

The main characteristic related to a random set is its capacity functional. Denote by $\mathcal{K}$ the collection of all compact sets of $\mathbb{R}$. The functional $T_X : \mathcal{K} \to [0,1]$ defined by

$$T_X(K) = P\left(\{X \cap K \neq \emptyset\}\right)$$

is said to be the *capacity functional* of X. Note that the capacity functional is defined over a very large collection of sets (the collection of compact sets). As aforementioned, this is a hard disadvantage in practice. Eventually, we are able to obtain a characterization result for connectedness in terms of simple families of sets (in fact, a finite collection of intervals).

Let us work with the extended real line $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty, -\infty\}$. Let

$$Q := \{p_0 = -\infty, p_1, p_2, \ldots, p_n, p_{n+1} = +\infty\}$$

be a *finite set of ordered points of* $\overline{\mathbb{R}}$; that is, by definition, $p_i < p_j$ if $i < j$, and $p_1, \ldots, p_n \in \mathbb{R}$). From that points, we are able to build a collection of sets $V = \{v_i\}_{i=1}^n$ as follows: $v_i = [p_{i-1}, p_{i+1}]$ for $1 \leq i \leq n$. We call *vertices* to the elements of V. The terminology comes from the fact that we will build a finite linear graph G_Q. Two vertices, v_i and v_j, of G_Q are connected by an edge if $|i - j| = 1$; in such case, the edge models the intersection of both sets. More precisely, denote by $E = \{e_i\}_{i=1}^{n-1}$ the set of edges such that $e_i = [p_i, p_{i+1}]$ is the edge whose vertices are v_i and v_{i+1}, for $i = 1, \ldots, n-1$.

Now, we define a map ψ which associates a set A with a subgraph of G_Q. Given a set A, $\psi(A)$ is the subgraph of G_Q whose vertices are $\{v_i : v_i \cap A \neq \emptyset\}$ and whose edges are

$\{e_i : e_i \cap A \neq \emptyset\}$. To show that ψ is well defined, we need to prove that any edge of $\psi(A)$ has its associated vertices also belonging to $\psi(A)$. In fact, if $e_i \in \psi(A)$, we have $[p_i, p_{i+1}] \cap A \neq \emptyset$, and consequently $[p_{i-1}, p_{i+1}] \cap A \neq \emptyset$ and $[p_i, p_{i+2}] \cap A \neq \emptyset$; that means that v_i and v_{i+1} (the vertices associated to the edge) also belong to $\psi(A)$.

Observe that if A is connected, then $\psi(A)$ is also connected as a graph. In this case, the following formula holds (see [8]):

$$\#\text{vertices}(\psi(A)) - \#\text{edges}(\psi(A)) = 1\,, \tag{3}$$

where here and so on, # denotes the cardinality.

Remark 2.2. More generally, for a set A and a graph G, we have

$$\#\text{vertices}(\psi(A)) - \#\text{edges}(\psi(A)) = \#\text{connected components}(\psi(A))\,.$$

If A is not connected but closed, then we can find suitable points $Q = \{p_0 = -\infty, p_1, p_2, p_3 = +\infty\}$ such that $\psi(A)$ is not a connected subgraph of G_Q. In fact, since A is closed, we can find real numbers p_1 and p_2 satisfying

$$A \cap (-\infty, p_1] \neq \emptyset\ ,\ A \cap [p_1, p_2] = \emptyset\ ,\ A \cap [p_2, \infty) \neq \emptyset\,.$$

Then $\psi(A)$ is a graph with two vertices and no edge.

Now, let X be a random closed set. From the remark above,

$$\mathbb{E}\left[\#\text{vertices}(\psi(X))\right] - \mathbb{E}\left[\#\text{edges}(\psi(X))\right] \geq 1\,.$$

If X is connected, $\psi(X)$ is also connected (as a graph). Therefore, equation (3) implies:

$$\mathbb{E}\left[\#\text{vertices}(\psi(X))\right] - \mathbb{E}\left[\#\text{edges}(\psi(X))\right] = 1\,. \tag{4}$$

Next, we have to compute the previous expectations. To do that, we need the following technical result,

Lemma 2.3 (Robbin's theorem)**.** *Let Y be a random set of a finite space $C = \{c_i\}_{i \in N \subset \mathbb{N}}$. It holds*

$$\mathbb{E}\left[\#Y\right] = \sum_{i \in N} P\left(c_i \in Y\right)\,.$$

From the previous result, we compute:

$$\begin{aligned}\mathbb{E}\left[\#\text{vertices}(\psi(A))\right] &= \sum_{i=1}^{n} P\left(\{[p_{i-1}, p_{i+1}] \cap X \neq \emptyset\}\right)\\ &= \sum_{i=1}^{n} T_X\left([p_{i-1}, p_{i+1}]\right)\,.\end{aligned}$$

Following a similar reasoning,

$$\mathbb{E}\left[\#\text{edges}(\psi(X))\right] = \sum_{i=1}^{n-1} T_X\left([p_i, p_{i+1}]\right)\,.$$

At this point, we have proved one implication of our main result:

Theorem 2.4. *A random closed set X of $\mathbb{R}$ is connected if and only if any finite set of ordered points $\{p_0 = -\infty, p_1, p_2, \ldots, p_n, p_{n+1} = +\infty\}$ satisfy*

$$\sum_{i=1}^{n} T_X\left([p_{i-1}, p_{i+1}]\right) - \sum_{i=1}^{n-1} T_X\left([p_i, p_{i+1}]\right) = 1\,. \tag{5}$$

To prove the converse of that result, we prove that there exists a finite set of ordered points that does not satisfy the equation (5) when X is not connected.

If X is a non-connected random closed set, then the Proposition 2.1 allows us to find p_1 and p_2 ($p_1 < p_2$) satisfying

$$\Omega' := \{\omega : X(\omega) \cap (-\infty, p_1] \neq \emptyset, X(\omega) \cap [p_1, p_2] = \emptyset, X(\omega) \cap [p_2, \infty) \neq \emptyset\}$$

$$P\left(\Omega'\right) > 0\,.$$

For these points, the left part of the equation (5) reads as follows:

$$\int_\Omega \Big[\quad \mathbb{I}\left\{\omega : X(\omega) \cap (-\infty, p_2] \neq \emptyset\right\} + \mathbb{I}\left\{\omega : X(\omega) \cap [p_1, \infty) \neq \emptyset\right\} - \mathbb{I}\left\{\omega : X(\omega) \cap [p_1, p_2] \neq \emptyset\right\} \quad \Big]\, dP\,.$$

To compute this integral, we split Ω into the following pairwise disjoint sets:

$$\begin{aligned}
\Omega_1 &:= \{\omega : X(\omega) \cap (-\infty, p_2] \neq \emptyset\,, X(\omega) \cap [p_1, \infty) = \emptyset\,, X(\omega) \cap [p_1, p_2] = \emptyset\} \\
\Omega_2 &:= \{\omega : X(\omega) \cap (-\infty, p_2] = \emptyset\,, X(\omega) \cap [p_1, \infty) \neq \emptyset\,, X(\omega) \cap [p_1, p_2] = \emptyset\} \\
\Omega_3 &:= \Omega' \\
\Omega_4 &:= \{\omega : X(\omega) \cap (-\infty, p_2] \neq \emptyset\,, X(\omega) \cap [p_1, \infty) \neq \emptyset\,, X(\omega) \cap [p_1, p_2] \neq \emptyset\}
\end{aligned}$$

It can be checked that

$$\int_{\Omega_i} \Big[\quad \mathbb{I}\left\{\omega : X(\omega) \cap (-\infty, p_2] \neq \emptyset\right\} + \mathbb{I}\left\{\omega : X(\omega) \cap [p_1, \infty) \neq \emptyset\right\} - \mathbb{I}\left\{\omega : X(\omega) \cap [p_1, p_2] \neq \emptyset\right\} \quad \Big]\, dP = P\left(\Omega_i\right)$$

for $i = 1, 2, 4$. However, we have

$$\int_{\Omega_3=\Omega'} [\ldots]\, dP \;\; = \;\; 2\, P\,(\Omega')$$

This means that the left side of the equation (5) equals $1 + P\,(\Omega')$. Since we have assumed that $P\,(\Omega') > 0$, the equation of our main result does not hold.

3 Conclusions

We have obtained a characterization of the connectedness of a random closed of $\mathbb{R}$ set in terms of its capacity functional. The main advantage of this formula is that we only need evaluations of the capacity functional over a finite collection of intervals. This fact implies that it can be easily implemented in practice. More precisely, a finite family of evaluations of the capacity over suitable intervals can yield to state that the random set is not connected. In fact, observe that the distances between the finite set of ordered points are key to detect the non-connectedness. To illustrate this idea, assume that a random set X is contained in an interval, let us say $[a, b]$. Consider finite sets of ordered points of the form: $\mathcal{I}_n = \left\{p_0 = -\infty, p_1 = a, p_2 = a + \frac{b-a}{n}, \ldots, p_j = a + j\,\frac{b-a}{n}, \ldots, p_n = b, p_{n+1} = \infty\right\}$. It is straightforward to see that X is connected if and only if the equation (5) holds over $\mathcal{I}_n$ for any $n \in \mathbb{N}$. Clearly, bigger n, easier to detect the non-connectedness. However, at the same time, more complexity in computability and in recollecting the data. It depends on each case where to fix an appropriate n in view of this reasoning.

As a future work, we would consider topological manifolds as the spaces over which the random set is defined. In this case, the topological complexity increases significantly (even in $\mathbb{R}^2$). This is a challenge that would be worth addressing.

Acknowledgements

The author is partially supported by the Spanish MICINN Grant with FEDER funds PGC2018-098623-B-I00.

References

[1] A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Ann. Math. Statist.*, 38:325–339, 1967.

[2] A. P. Dempster. Upper and lower probabilities generated by a random closed interval. *Ann. Math. Statist.*, 39:957–966, 1968.

[3] Sébastien Destercke, Didier Dubois, and Eric Chojnacki. Unifying practical uncertainty representations: I. generalized p-boxes. *International Journal of Approximate Reasoning*, 49(3):649–663, 2008.

[4] Didier Dubois and Henri Prade. The mean value of a fuzzy number. *Fuzzy sets and systems*, 24(3):279–300, 1987.

[5] Didier Dubois and Henri Prade. Random sets and fuzzy interval analysis. *Fuzzy Sets and Systems*, 42(1):87–101, 1991.

[6] Irwin R Goodman. *Fuzzy sets as equivalence classes of random sets*, volume 327. Pergamon Press, Oxford, 1982.

[7] Jia Guo and Xiaoping Du. Reliability sensitivity analysis with random and interval variables. *International journal for numerical methods in engineering*, 78(13):1585–1617, 2009.

[8] Allen Hatcher. *Algebraic topology.* Cambridge University Press, 2002.

[9] Charles J. Himmelberg and T. Parthasarathy. Measurable relations. *Fund. math*, 87 (1):53–72, 1975.

[10] Cliff Joslyn. Distributional representations of random interval measurements. In *Uncertainty Analysis in Engineering and Sciences: Fuzzy Logic, Statistics, and Neural Network Approach*, pages 37–51. Springer, 1998.

[11] Marie Kratz and Werner Nagel. On the capacity functional of excursion sets of gaussian random fields on $\mathbb{R}^2$. *Advances in Applied Probability*, 48(3):712–725, 2016.

[12] Elmar Kriegler and Hermann Held. Utilizing belief functions for the estimation of future climate change. *International journal of approximate reasoning*, 39(2-3):185–209, 2005.

[13] Georges Matheron. *Random sets and integral geometry.* Wiley New York, 1975.

[14] Ronald Meester and Rahul Roy. *Continuum percolation*, volume 119. Cambridge University Press, 1996.

[15] JR Millenson. Random interval schedules of reinforcement. *Journal of the Experimental Analysis of Behavior*, 6(3):437–443, 1963.

[16] Enrique Miranda, Inés Couso, and Pedro Gil. Random intervals as a model for imprecise information. *Fuzzy Sets and Systems*, 154(3):386–412, 2005.

[17] Ilya S. Molchanov. *Theory of random sets*, volume 19. Springer, 2005.

[18] Hung T Nguyen. On random sets and belief functions. *Journal of Mathematical Analysis and Applications*, 65(3):531–542, 1978.

[19] Hung T Nguyen. *An introduction to random sets.* Chapman and Hall/CRC, 2006.

[20] Wolfgang Stadje. Coverage problems for random intervals. *SIAM Journal on Applied Mathematics*, 49(5):1538–1551, 1989.

[21] Pedro Terán. Choquet theorem for random sets in polish spaces and beyond. In Sébastien Destercke, Thierry Denoeux, María Ángeles Gil, Przemyslaw Grzegorzewski, and Olgierd Hryniewicz, editors, *Uncertainty Modelling in Data Science*, pages 208–215, Cham, 2019. Springer International Publishing. ISBN 978-3-319-97547-4.

[22] Peter F Thall and John M Lachin. Analysis of recurrent events: Nonparametric methods for random-interval count data. *Journal of the American Statistical Association*, 83 (402):339–347, 1988.

[23] Peter Walley. *Statistical reasoning with imprecise probabilities.* Chapman and Hall, London, 1991.