3rd *BYMAT Conference* (2020)

Distribution functions and probability measures on linearly ordered topological spaces

José F. Gálvez-Rodríguez
Universidad de Almería
jgr409@ual.es

Miguel A. Sánchez-Granero
Universidad de Almería
misanche@ual.es

Abstract: In this work we describe a theory of a cumulative distribution function (in short, cdf) on a separable linearly ordered topological space (LOTS) from a probability measure defined in this space. This function can be extended to the Dedekind-MacNeille completion of the space where it does make sense to define the pseudo-inverse. Moreover, we study the properties of both functions (the cdf and the pseudo-inverse) and get results that are similar to those which are well-known in the classical case.

Resumen: En este trabajo describimos una teoría sobre la función de distribución acumulada en un espacio topológico linealmente ordenado separable a partir de una medida de probabilidad definida en este espacio. Esta función se puede extender a la completación Dedekind-MacNeille del espacio donde tiene sentido definir la pseudo-inversa. Además, estudiamos las propiedades de ambas funciones (la función de distribución y la pseudo-inversa) y obtenemos resultados similares a los conocidos en el caso clásico.

Keywords: probability, measure, Dedekind-MacNeille completion, cumulative distribution function, linearly ordered topological space.

MSC2010: 60E05, 60B05.

Reference: Gálvez-Rodríguez, José F., and Sánchez-Granero, Miguel A. "Distribution functions and probability measures on linearly ordered topological spaces". In: *TEMat monográficos*, 2 (2021): *Proceedings of the 3rd BYMAT Conference*, pp. 203-206. ISSN: 2660-6003. URL: https://temat.es/monograficos/article/view/vol2-p203.

1. Introduction

This work collects some results on a theory of a cumulative distribution function (cdf) on a separable linearly ordered topological space (LOTS).

In [2], we described a theory of a cumulative distribution function on a separable linearly ordered topological space. Moreover, we showed that this function plays a similar role to that played in the classical case and studied its pseudo-inverse, which allowed us to generate samples of the probability measure that we used to define the distribution function. In [3], we extended a cdf defined on a separable linearly ordered topological space, X, to its Dedekind-MacNeille completion, $DM(X)$. That completion is, indeed, a compactification. Moreover, we proved that each function satisfying the properties of a cdf on $DM(X)$ is the cdf of a probability measure defined on $DM(X)$. Indeed, if X is compact, a similar result can be obtained in this context. Finally, the compactification $DM(X)$ lets us generate samples of a distribution in X. By following this research line, the next step is exploring some conditions on X such that, given a function F with the properties of a cdf, we can ensure that there exists a unique probability measure on X such that its cdf is F. This is completely developed in [4].

For further reference about the classical measure theory see, for example, [5].

2. Preliminaries: linearly ordered topological spaces

First, we recall the definition of a linear order:

Definition 1. A partially ordered set $(X, \leq)$ (that is, a set X with the binary relation $\leq$ that is reflexive, antisymmetric and transitive) is totally ordered if every $x, y \in X$ are comparable, that is, $x \leq y$ or $y \leq x$. In this case, the order is said to be total or linear. ◂

For further reference about partially ordered spaces see, for example, [1].

From the previous definition, we can talk about linearly ordered topological spaces:

Definition 2. A linearly ordered topological space (in short, LOTS) is a triple $(X, \tau, \leq)$ where $(X, \leq)$ is a linearly ordered set and where τ is the topology of the order. ◂

The topology of the order is defined as follows:

Definition 3. Let X be a set which is linearly ordered by $<$. We define the order topology τ on X by taking the sub-basis $\{\{x \in X : x < a\} : a \in X\} \cup \{\{x \in X : x > a\} : a \in X\}$. ◂

In the rest of this work, we will assume that X is a separable LOTS, μ is a probability measure on the Borel σ-algebra of X, $\sigma(X)$, and τ is the order topology in X.

3. Defining the cumulative distribution function

First we give the definition of the cumulative distribution function of a probability measure defined on X.

Definition 4. Given a probability measure μ on X, its cumulative distribution function (in short, cdf) is the function $F: X \to [0, 1]$ defined by $F(x) = \mu(\{a \in X : a \leq x\})$, for each $x \in X$. ◂

This function satisfies some properties that we collect next, and that are similar to those which are well-known in the classical theory of distribution functions:

Proposition 5. *Let μ be a probability measure on X, and F its cdf. Then,*

(i) *F is monotonically non-decreasing.*

(ii) *F is right τ-continuous.*

(iii) $\sup F(X) = 1$.

(iv) *If there does not exist* $\min X$, *then* $\inf F(X) = 0$.

Once we have defined a cdf and studied its properties, it does make sense to ask ourselves if, given a function, F, satisfying the properties collected above, there exists a probability measure μ on X such that its cdf, F_μ, is F. This is a question we will answer with the help of a structure we analyse in the next section.

Indeed, if we work with F we can get the measure of each interval in X.

Proposition 6. *If F is the cdf of a probability measure μ on X, then $\mu(\{x \in X : a < x \leq b\}) = F(b) - F(a)$, for each $a, b \in X$ such that $a < b$.*

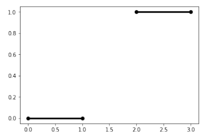
The next result is about the continuity of a cdf:

Proposition 7. *Let $x \in X$, μ be a probability measure on X and F its cdf. If $\mu(\{x\}) = 0$, then F is τ-continuous at x.*

However, the converse is not true. To show that, we include an example where the cdf of a probability measure on a space is a step cdf which is continuous with respect to τ.

Example 8. Let $X = [0,1] \cup [2,3]$ with the usual order. If μ is a probability measure defined on X by $\mu(\{2\}) = 1$, then its cdf is the function $F\colon X \to [0,1]$ defined by

$$F(x) = \begin{cases} 0 & \text{if } x \leq 1, \\ 1 & \text{if } x \geq 2. \end{cases}$$



Note that F is a step cdf and it is τ-continuous. ◂

4. The extension of a cdf to the Dedekind-MacNeille completion

Definition 9. The Dedekind-MacNeille completion of a partially ordered set X is defined to be $DM(X) = \{A \subseteq X : A = (A^u)^l\}$ ordered by inclusion ($A \leq B$ if and only if $A \subseteq B$), where A^u (resp. A^l) is the set of upper (resp. lower) bounds of A. ◂

We also define $\phi\colon X \to DM(X)$ as the embedding given by $\phi(x) = (\{a \in X : a \leq x\})$, for each $x \in X$. For further reference about cuts and the Dedekind-MacNeille completion see, respectively, [6] and [7].

The next result gives us that each cdf on a LOTS can be naturally extended to its Dedekind-MacNeille completion.

Proposition 10. *$DM(X)$ is, indeed, a compactification of X and F can be extended to a cdf, $\widetilde{F}$, on $DM(X)$ by defining $\widetilde{F}\colon DM(X) \to [0,1]$ by $\widetilde{F}(A) = \inf F(A^u)$, for each $A \in DM(X)$.*

5. The pseudo-inverse of a cdf

Definition 11. Let F be a cdf. We define the pseudo-inverse of F by $G\colon [0,1] \to DM(X)$ given by $G(r) = \{x \in X : F(x) \geq r\}^l$, for each $r \in [0,1]$. ◂

This function satisfies some properties which are similar to those which are well-known in the classical case for the pseudo-inverse of a cdf, as we can see next:

Proposition 12. *The following hold:*

(i) *G is monotonically non-decreasing.*

(ii) *G is left τ-continuous.*

(iii) *$G(r) \leq \phi(x)$ if and only if $r \leq F(x)$, for each $x \in X$ and each $r \in [0,1]$.*

6. Relationship between a probability measure and its cdf

Once we have defined and studied the main properties of a cdf and its pseudo-inverse, we answer the question made in Section 3 about the univocal relationship between a probability measure and its cdf in the context of separable LOTS. For that purpose, the main result is the next one:

Theorem 13. *Let X be a separable LOTS such that $DM(X) \setminus \phi(X)$ is countable and $F\colon X \to [0,1]$ be a monotonically non-decreasing and right τ-continuous function satisfying $\sup F(X) = 1$ and $\sup F(A) = \inf F(A^u)$, for each $A \in DM(X)$. Moreover, $\inf F(X) = 0$ if there does not exist the minimum of X. Then, there exists a unique probability measure on X, μ, such that $F = F_\mu$.*

What is more, the pseudo-inverse of a cdf is also univocally determined by its probability measure:

Theorem 14. *Let X be a separable LOTS such that $DM(X)\setminus\phi(X)$ is countable and let $G\colon [0,1] \to DM(X)$ be a monotonically non-decreasing and left τ-continuous function such that $\sup G^{-1}(< A) = \inf G^{-1}(> A)$, for each $A \in DM(X) \setminus \phi(X)$, $G(0) = \min DM(X)$, $G^{-1}(\max DM(X)) \subseteq \{1\}$ if there does not exist the maximum of X and $G^{-1}(\min DM(X)) = \{0\}$ if there does not exist the minimum of X. Then, there exists a unique probability measure on X, μ, such that G is the pseudo-inverse of F_μ.*

7. Applications

7.1. Generating samples

First, we can get the measure of each subset in the Borel σ-algebra of X from the pseudo-inverse of a cdf.

Proposition 15. *Let μ be a probability measure. Then, $\mu(A) = l(G^{-1}(A))$ for each $A \in \sigma(X)$, where l is the Lebesgue measure.*

That procedure lets us generate samples of a distribution, similarly to the classical procedure for distribution functions in the real line.

Remark 16. We can also calculate integrals with respect to μ, so, for $g\colon X \to \mathbb{R}$, it holds that

$$\int g(x)\,\mathrm{d}\mu(x) = \int g(G(t))\,\mathrm{d}t.$$

◂

7.2. A goodness-of-fit test

In this subsection, we give a goodness-of-fit test whose idea is similar to the one followed by the Kolmogorov-Smirnov test, but in a more general context. Suppose that we are given a random sample on a separable LOTS according to a certain cumulative distribution function. Our purpose is testing if that distribution comes from a certain cdf F. Let us denote by F_n the empirical cumulative distribution function of the sample and define the statistic $D_n = \sup_{x\in X}|F_n(x) - F(x)|$, then the next statement holds.

Theorem 17. *Given a separable LOTS, X, and $n \in \mathbb{N}$, the distribution of D_n is the same for each cdf, F_μ, satisfying that $\mu(\{x\}) = 0$, for each $x \in X$.*

References

[1] Dushnik, D. and Miller, E. W. “Partially ordered sets”. In: *American Journal of Mathematics* 63 (1941), pp. 600–610.

[2] Gálvez-Rodríguez, J. F. and Sánchez-Granero, M. A. “The distribution function of a probability measure on a linearly ordered topological space”. In: *Mathematics* 7.9 (2019), p. 864.

[3] Gálvez-Rodríguez, J. F. and Sánchez-Granero, M. A. “The distribution function of a probability measure on the Dedekind-MacNeille completion”. In: *Topology and its Applications* 275 (2020), No. 107010.

[4] Gálvez-Rodríguez, J. F. and Sánchez-Granero, M. A. “Equivalence between distribution functions and probability measures on a LOTS”. Preprint.

[5] Halmos, P. R. *Measure Theory*. New York: Springer-Verlag, 1974.

[6] MacNeille, H. M. “Partially ordered sets”. In: *Transactions of the American Mathematical Society* 42.3 (1937), pp. 416–460.

[7] Roman, S. *Lattices and ordered sets*. New York: Springer-Verlag, 2008.