Vladimir G. Pestov

On a result of K. P. Hart about non-existence of measurable solutions to the discrete expectation maximization problem

Commentationes Mathematicae Universitatis Carolinae, Vol. 64 (2023), No. 3, 353–358

Persistent URL: `http://dml.cz/dmlcz/152303`

On a result of K. P. Hart about non-existence of measurable solutions to the discrete expectation maximization problem

Vladimir G. Pestov

Abstract. It was shown that there is a statistical learning problem – a version of the expectation maximization (EMX) problem – whose consistency in a domain of cardinality continuum under the family of purely atomic probability measures and with finite hypotheses is equivalent to a version of the continuum hypothesis, and thus independent of ZFC. K. P. Hart had subsequently proved that no solution to the EMX problem can be Borel measurable with regard to an uncountable standard Borel structure on X, and so the independence result could just be an artefact of a model allowing non-measurable learning rules. In this note we reinforce the point somewhat by observing that such a solution cannot even be Lebesgue measurable.

Keywords: expectation maximization problem; EMX; continuum hypothesis; independence of ZFC; measurability

Classification: 68T05, 03E35

1. Introduction

The default model of statistical learning assumes that datapoints belong to a standard Borel space, whose measurable sigma-algebra is generated by a complete separable metric, and the learning rule – the mapping associating a hypothesis to every sample – is Borel (or at least universally) measurable. However, it certainly makes sense to push the limits of the model by dropping some of the restrictions and studying the consequences. An interesting recent work by S. Ben-David, P. Hrubeš, S. Moran, A. Shpilka, and A. Yehudayoff [2], see [1] for a more detailed exposition, illustrates what happens if the requirement of universal measurability of the learning rule is dropped. In this case, there is a learning problem – the expectation maximization (EMX) problem – whose consistency in the domain X of cardinality continuum under the family of all purely atomic probability measures and having all finite sets as posible hypotheses is equivalent to a version of the continuum hypothesis, and thus, in the case $X = \mathbb{R}$, independent of

DOI 10.14712/1213-7243.2023.031

Holder of the DCR-A fellowship 300050/2022-4 of the Program of Scientific and Technological Development of the State of Paraíba by CNPq and FAPESQ.

the standard axioms of the Zermelo–Fraenkel set theory with the axiom of choice (ZFC). That a solution to the EMX problem cannot be Borel measurable if X is identified with the interval $\mathbb{I} = [0,1]$ (in other words, if X is an uncountable standard Borel space), was immediately proved by K. P. Hart in [3]. Thus, the independence of the EMX learning problem of ZFC could just be an artefact of a model allowing non-measurable learning rules. In this note, we show that such a solution cannot even be Lebesgue measurable.

The EMX problem calls to guess, probably approximately correctly, a set $S(\sigma)$ having a nearly full measure on the basis of a random finite unlabelled sample, σ. The sample follows an unknown probability distribution, μ, with regard to which the measure of the hypothesis is calculated as well. Thus, the error and the confidence of the guess are supposed to be uniformly bounded over a given family of probability measures on the domain (a measurable space). Finally, the hypothesis $S(\sigma)$ should belong to a specified family of sets, measurable with regard to each measure from our specified family.

In more exact terms, given a family M of probability measures on a measurable space $(X, \mathcal{A})$, and a family $\mathcal{H} \subseteq \mathcal{A}$ (the hypothesis class), does there exist a map

$$S\colon \bigcup_{n=1}^{\infty} X^n \to \mathcal{H},$$

so that, given $\varepsilon, \delta > 0$, there is N so that for all $n \geq N$ and each $\mu \in M$,

$$P[\mu(S(\sigma_n)) > 1-\varepsilon] > 1-\delta?$$

Here σ_n denotes the independent and identically distributed (i.i.d.) n-sample following the law μ. The probability P refers, in a measure-theoretic reformulation, to the product measure $\mu^{\otimes n}$ on the set X^n of all unlabelled n-samples σ. When talking of the measurability of S, one can understand it as the measurability of the associated map

$$\bigcup_{n=1}^{\infty} X^n \times X \to \{0,1\}.$$

A basic example of an EMX problem admitting a Borel measurable solution is when X is the Euclidean space $\mathbb{R}^d$ with the standard Borel structure, the family of measures consists of all Borel probability measures, and the hypotheses are all compact subsets. The mapping S can associate to every sample σ the smallest closed ball around the origin containing all points of σ. Indeed, given $\delta, \varepsilon > 0$ and an unknown distribution μ, let B be the largest open ball around the origin having μ-measure less than or equal to $1-\varepsilon$. Then the corresponding closed ball $\bar{B}$ has the measure greater than or equal to $1-\varepsilon$. The probability for all n

points of a random i.i.d. sample σ following the distribution μ to belong to B is less than or equal to $(1-\varepsilon)^n$. Thus, with confidence greater than or equal to $1-(1-\varepsilon)^n$, the ball $S(\sigma)$ contains $\bar{B}$ and thus has measure greater than or equal to $1-\varepsilon$. If $n \geq \log\delta/\log(1-\varepsilon)$, then $1-(1-\varepsilon)^n \geq 1-\delta$, so our S indeed solves the EMX problem.

In the specific version of the problem considered by the authors, the domain is just any set, X, equipped with a sigma-algebra containing the singletons, and the family of probability measures in question, $P_a(X)$, consists of all purely atomic measures on X. The hypothesis class consists of all finite subsets of the domain. Denote $[X]^{<\infty}$ the family of all finite subsets of X. Now the question becomes: does there exist a map,

$$S\colon \bigcup_{n=1}^{\infty} X^n \to [X]^{<\infty},$$

with the property that for every $\varepsilon, \delta > 0$ there is $n = n(\varepsilon, \delta)$ so that

$$\forall\, m \geq n,\ \forall\, \mu \in P_a(X), \quad P[\mu(S(\sigma_m)) > 1-\varepsilon] > 1-\delta?$$

The central theorem of [2], [1] states that such an S exists if and only if the cardinality of X is less than $\aleph_\omega$. In particular, the domain of real numbers, $\mathbb{R}$, admits a solution to the EMX problem over purely atomic measures if and only if the continuum equals $\aleph_n$ for some natural n, and thus the assertion is independent of ZFC.

The main criticism of the result belongs to K. P. Hart in [3], who has in particular shown that, if $X = [0,1]$, no Borel measurable map S with the above properties can exist. (Of course the conclusion now holds for any uncountable standard Borel space X.)

Below we notice that a map S having the required properties cannot even be Lebesgue measurable. The argument is a variation on the original argument from [3].

2. The argument

Let $[\mathbb{I}]^m$ denote, for $m \in \mathbb{N}$, the family of all m-subsets of the interval given the Vietoris topology. Thus, two finite sets A and B with m elements are ε-close if A is included in the ε-neighbourhood of B and vice versa. We will identify $[\mathbb{I}]^m$ with a subset of all elements of $\mathbb{I}^m$ of the form

$$x = (x_1, x_2, \ldots, x_m), \qquad x_1 < x_2 < \ldots < x_m.$$

The image of $[\mathbb{I}]^m$ in $\mathbb{I}^m$ is an open m-simplex, having the Lebesgue measure $1/m!$. We will denote it by the same symbol, $[\mathbb{I}]^m$, and equip with the $l^\infty(m)$ distance (which corresponds to the Vietoris distance) until the end of the argument.

Here is the main technical tool showing that a Lebesgue measurable finite-to-one compression function $[\mathbb{I}]^{m+1} \to [\mathbb{I}]^m$ cannot exist.

Lemma 1. *There exists no finite-to-one Lebesgue measurable map*

$$\kappa\colon [\mathbb{I}]^{m+1} \to [\mathbb{I}]^m$$

having the property $\kappa(\sigma) \subseteq \sigma$.

Proof: Let $\kappa\colon [\mathbb{I}]^{m+1} \to [\mathbb{I}]^m$ be a Lebesgue measurable map having the property $\kappa(\sigma) \subseteq \sigma$.

Fix any point $x = (x_1, \dots, x_{m+1}) \in [\mathbb{I}]^{m+1}$, and define

$$\varepsilon = \frac{1}{3} \min_{1 \le i < j \le m+1} d(x_i, x_j) > 0.$$

Let $\gamma > 0$ be the Lebesgue measure of the open ball $B_\varepsilon(x)$ taken in $[\mathbb{I}]^{m+1}$ (seen as a simplex with l^∞ metric). According to Luzin's theorem, there is a compact set $K \subseteq [\mathbb{I}]^{m+1}$ having measure greater than $1-\gamma$ and such that $\kappa|_K$ is continuous, thus uniformly continuous. Choose $\delta \le \varepsilon$ so small that if $\sigma, \tau \in K$ and $d(\sigma, \tau) < \delta$, then $d(\kappa(\sigma), \kappa(\tau)) < \varepsilon$.

Denote

$$K' = K \cap B_\varepsilon(x).$$

The set K' has a strictly positive Lebesgue measure. Therefore, there exists a point $y \in K'$ whose δ-neighbourhood has a strictly positive Lebesgue measure (because K' is precompact, so can be covered with finitely many balls of radius δ). Denote $K'' = K' \cap B_\delta(y)$.

Assume without loss of generality that

$$\kappa(y) = (y_1, y_2, \dots, y_m),$$

that is, the coordinate $(m+1)$ is removed. (If it is another coordinate, we will just apply a permutation to the simplex and to K'. This mapping will of course send the simplex image of $[\mathbb{I}]^{m+1}$ to another subsimplex of $\mathbb{I}^{m+1}$, but it preserves both the Lebesgue measure and the l^∞-metric.)

For any $z \in B_\delta(y)$, we have $d(y, z) < \delta$, thus, if $z \in K''$, then $d(\kappa(y), \kappa(z)) < \varepsilon$. Consequently, for all i, $\kappa(z)_i \in B_\varepsilon(\kappa(y)_i)$, and in particular, $\kappa(z)$ is also obtained by removing the last coordinate of z. We conclude that, for all $z \in K''$,

$$\kappa(z) = \pi_{[1,m]}(z),$$

the coordinate projection on the first m coordinates.

By the Fubini theorem,

$$\mu(K'') = \int_0^1 \mu^{(m)}(\pi^{-1}_{[1,m]}(z) \cap K'')\, \mathrm{d}\lambda(z),$$

and since $\mu(K'') > 0$ for a set of points z of positive measure the set $\kappa^{-1}(z)$ is infinite. □

Recall that $[\mathbb{I}]^{<\infty}$ is the family of all finite subsets of the closed interval $\mathbb{I} = [0,1]$, and $P_a(\mathbb{I})$ is the set of all purely atomic probability measures on $\mathbb{I}$.

Theorem 2. *There is no Lebesgue measurable map*

$$S\colon \bigcup_{n=1}^{\infty} \mathbb{I}^n \to [\mathbb{I}]^{<\infty}$$

with the property that for every $\varepsilon, \delta > 0$ there is $n = n(\varepsilon, \delta)$ so that

$$\forall m \geq n,\ \forall \mu \in P_a(\mathbb{I}), \quad P[\mu(S(\sigma_m)) > 1 - \varepsilon] > 1 - \delta.$$

The result is deduced from Lemma 1 word for word as in [3]: if there existed a Lebesgue measurable solution S to the EMX problem for the class of finite sets under all purely atomic measures, then there would exist a Lebesgue measurable finite-to-one compression function $\kappa\colon [\mathbb{I}]^{m+1} \to [\mathbb{I}]^m$, because the choice of a point to remove can be done in a Borel measurable fashion (e.g. by always removing the smallest possible point).

Acknowledgement. The author is grateful to Samuel Gomes da Silva and Ciro Russo for inviting him to read and explain the results of [1], [2], [3] in a minicourse at the Logic, Set Theory, and Topology week of the Federal University of Bahía, Salvador, BA, Brazil in March 2020.

References

[1] Ben-David S., Hrubeš P., Moran S., Shpilka A., Yehudayoff A., *A learning problem that is independent of the set theory ZFC axioms*, available at arXiv 1711.05195v1 [cs.LG] (2017), 17 pages.

[2] Ben-David S., Hrubeš P., Moran S., Shpilka A., Yehudayoff A., *Learnability can be undecidable*, Nat. Mach. Intell. **1** (2019), no. 1, 44–48.

[3] Hart K. P., *Machine learning and the continuum hypothesis*, Nieuw Arch. Wiskd. (5) **20** (2019), no. 3, 214–217.

V. G. Pestov:
Departamento de Matemática, Universidade Federal da Paraíba,
Cidade Universitaria, 58051-900 João Pessoa, PB, Brasil
and:
Departement of Mathematics and Statistics, University of Ottawa,
STEM Complex, 150 Louis-Pasteur Private, Ottawa, Ontario, ON K1N,
Canada

E-mail: vladimir.pestov@uottawa.ca

(*Received* May 25, 2023, *revised* August 13, 2023)