CONTINUOUS AND RANDOM VAPNIK-CHERVONENKIS CLASSES

ITAÏ BEN YAACOV

Abstract. Nous démontrons que si T est une théorie dépendante, sa randomisée de Keisler T^R l'est aussi.

Pour faire cela nous généralisons la notion d'une classe de Vapnik-Chervonenkis à des familles de fonctions à valeurs dans $[0,1]$ (une classe de Vapnik-Chervonenkis *continue*), et nous caractérisons les familles de fonctions ayant cette propriété par la vitesse de croissance de la largeur moyenne d'une famille de compacts convexes associés.

In this paper we answer a question lying at the intersection of two currently active research themes in model theory.

The first theme is that of dependent theories, i.e., first order theories which do no possess the independence property, first defined by Shelah [She71]. A formula $\varphi(x, y)$ is said to have the *independence property*, or to be *independent*, in a theory T if for every n one can find a model $\mathcal{M} \vDash T$, $b_i \in M$ for $i < n$, and $a_w \in M$ for $w \subseteq n$, such that $\mathcal{M} \vDash \varphi(a_w, b_i) \iff i \in w$. The theory T has the independence property if at least one formula has it in T; equivalently, a theory T is dependent if every formula is. A stable theory is necessarily dependent. (More generally, a theory T is unstable if and only if it is independent or has the *strict order property*. See [Poi85, Théorème 12.38] for the proof in classical first order logic. It can be adapted easily to continuous logic following standard translation methods.)

The recent use of properties of dependent theories for the solution of the so-called Pillay Conjecture in [HPP] earned them a considerable increase in general interest. It should be pointed out that some refer to dependent theories as NIP (Non Independence Property) theories. Since non-NIP theories are quite "wild", research in this area concentrates on theories which are not non-NIP.

The second theme of research is that of continuous logic and metric structures. Studying metric structures using a model theoretic approach dates back to Henson [Hen76] and Krivine and Maurey [KM81]. Continuous first order logic was much more recently

---

Date: 21 January 2008.

2000 *Mathematics Subject Classification.* 03C95,03C45,52A38.

Key words and phrases. Vapnik-Chervonenkis class, dependent relation, dependent theory, mean width, randomisation.

Research initiated during the workshop "Model theory of metric structures", American Institute of Mathematics Research Conference Centre, 18 to 22 September 2006.

Research supported by ANR chaire d'excellence junior (projet THEMODMET) and by the European Commission Marie Curie Research Network ModNet.

introduced in [BU] as a formalism for this study, shifting the point of view much closer to classical first order logic. The independence property has a natural analogue for metric structures, and one may speak of dependent continuous theories. In contrast with the body of work concerning classical dependent theories, to the best of our knowledge dependent continuous theories have hardly been studied to date.

The intersection of these two themes in which we are interested stems from H. Jerome Keisler's randomisation construction. This first appeared in [Kei99], where to every (complete) first order theory T he associated the first order theory of spaces of random variables in models of T. Since classical first order logic is not entirely adequate for the treatment of spaces of random variables, which are metric by nature, this construction was subsequently improved to produce for every theory T the *continuous* theory T^R of spaces of random variables taking values in models of T (see [BK]). While we shall not go through the details of the construction, we shall point the main properties of T^R in Section 5. The author has shown that

(i) The randomisation of a stable theory is stable ([BK], in preparation).
(ii) On the other hand, the randomisation of a simple unstable theory is not simple. More generally, the randomisation of an independent theory cannot be simple, and is generally wild.

In other words, independent theories are somehow wild with respect to randomisation (even if they do satisfy other tameness properties such as simplicity), while stable theories are tame. It is natural to ask whether the dividing line for tame randomisation lies precisely between dependent and independent theories. A more precise instance of this question would be, is the randomisation of a dependent theory dependent? While Keisler's construction was only stated for a complete classical theory T it can be carried out just as well for an arbitrary continuous theory T (with some minor technical changes) and the question can also be posed when T is continuous.

The article [HPP] mentioned above relates dependent theories and probability measures on types, i.e., with types in the randomised theory. This suggested that dependent theories should be tame with respect to randomisation, so the answer to the question above should be positive. (To our best recollection this was first conjectured by Anand Pillay in the AIM workshop on Model Theory of Metric Structures, September 2006).

In order to give a positive answer we shall consider some purely combinatorial aspects of the independence property, observed by Shelah [She71] and independently by Vapnik and Chervonenkis [VC71] and seek to prove they extend to the continuous setting. (The connection between dependent theories and the work of Vapnik and Chervonenkis was pointed out by Laskowski [Las92].) Doing so we will need a new means for measuring the size of a set, as merely counting points will no longer do. The Gaussian mean width turns out to serve our purposes quite well (the Lebesgue measure of the set once inflated a little will also be useful, but to a much lesser extent). The Gaussian mean width commutes, in a sense, with the randomisation construction, and it follows painlessly that the randomisation of a dependent relation is again a dependent relation. Thus the technical core of this paper has nothing to do with model theory and deals rather with combinatorics and the geometry of convex compacts.

Section 1 consists of a few basic facts regarding convex compacts in $\mathbb{R}^n$ and their mean width.

The combinatorial core of the article is in Section 2 where continuous Vapnik-Chervonenkis classes and dependent relations are characterised via the growth rate of the mean width of an associated family of sets.

Section 3 consists of a technical interlude where we prove that continuous combinations of dependent relations are dependent.

In Section 4 we consider random dependent relations. Using the mean width criterion from Section 2 we show that if a random family of functions is uniformly dependent then its expectation is dependent as well.

The proper model theoretic contents of this paper is restricted to Section 5. We define dependent theories and the randomisation of a theory. The main theorem, asserting that the randomisation of a dependent theory is dependent, follows easily from earlier results. We also extend to continuous logic a classical result saying that in order to verify that a theory is dependent it suffices to verify that every formula $\varphi(x, \bar{y})$ is dependent where x is a single variable.

1. Facts regarding convex compacts and mean width

This section contains few properties of the mean width function. The author is much indebted to Guillaume Aubrun for having introduced him to this notion and its properties. All the results presented here are easy to verify and are either folklore (see for example [AS06]) or (in the case of integrals of convex compacts, as far as we know) minor generalisations thereof.

Let $A \subseteq \mathbb{R}^n$ a bounded set. For $y \in \mathbb{R}^n$ define $h_A(y) = \sup_{x \in A} \langle x, y \rangle$. As a function of y, h_A is positively homogeneous and sub-additive, and thus in particular convex.

Let $u \in S^{n-1}$. The real numbers $t_1 = h_A(u)$ and $t_2 = -h_A(-u)$ are then minimal and maximal, respectively, so that $t_2 \leq \langle x, u \rangle \leq t_1$ for all $x \in A$, i.e., such that A lies between the two hyperplanes $t_2 u + u^\perp$ and $t_1 u + u^\perp$. The *width* of A in the direction $u \in S^{n-1}$ is therefore defined to be $w(A, u) = h_A(u) + h_A(-u)$.

Let $K = \overline{\operatorname{Conv}}(A)$ be the closed convex envelope of A, i.e., the intersection of all closed half-spaces containing A. Then $h_K = h_A$ and:

$$K = \bigcap_{u \in S^{n-1}} \{x \colon \langle x, u \rangle \leq h_K(u)\}.$$

We may thus identify a convex compact $K \subseteq \mathbb{R}^n$ with $h_K \colon S^{n-1} \to \mathbb{R}$. In this case the supremum in the definition of h_K is attained at an extremal point of K.

It is easy to observe that the function $h_K(u)$ is monotone, positively homogeneous and additive in K, i.e., $K \subseteq K' \Longrightarrow h_K(u) \leq h_{K'}(u)$, $h_{\alpha K}(u) = \alpha h_K(u)$ for $\alpha \geq 0$ and

$$\begin{aligned} h_{K+K'}(u) &= \max_{x \in K+K'} \langle x, u \rangle = \max_{y \in K, z \in K'} \langle y + z, u \rangle \\ &= \max_{y \in K} \langle y, u \rangle + \max_{z \in K'} \langle z, u \rangle = h_K(u) + h_{K'}(u). \end{aligned}$$

Let $(X, \mathfrak{B}, \mu)$ be a measure space, $\mathbf{K}$ a mapping from X to the space of convex compacts in $\mathbb{R}^n$. Say that $\mathbf{K}$ is *measurable* (respectively, integrable) if $\omega \mapsto h_{\mathbf{K}(\omega)}(u)$ is for all $u \in S^1$. Notice that $u \mapsto h_{\mathbf{K}(\omega)}(u)$ is $b(\mathbf{K}(\omega))$-Lipschitz where $b(K) = \max\{h_K(u) \colon u \in S^{n-1}\}$. Thus, if $h_{\mathbf{K}}(u)$ is measurable for all u in some dense (and possibly countable) subset of S^{n-1} then $b(\mathbf{K})$ is measurable and thus $\mathbf{K}$ is. If $\mathbf{K}$ is integrable define:

$$h(u) = \int h_{\mathbf{K}}(u)\, d\mu,$$

$$K = \int \mathbf{K}\, d\mu = \bigcap_{u \in S^{n-1}} \{x \colon \langle x, u \rangle \leq h(u)\}.$$

Clearly K is a convex compact, and if $\mathbf{x} \colon X \to \mathbb{R}^n$ satisfies $\mathbf{x}(\omega) \in \mathbf{K}(\omega)$ a.e. then $\int \mathbf{x}\, d\mu \in \int \mathbf{K}\, d\mu$. We claim furthermore that $h_K = h$. Indeed, it is clear by definition of K that $h_K \leq h$. Conversely, given $u \in S^{n-1}$ we may complete it to an orthonormal basis $u_0 = u, u_1, \ldots, u_n$. For each $\omega \in X$ there is a unique $x_\omega \in \mathbf{K}(\omega)$ such that the tuple $(\langle x_\omega, u_0 \rangle, \ldots, \langle x_\omega, u_{n-1} \rangle)$ is maximal in lexicographical order (among all $x \in \mathbf{K}(\omega)$). In particular $\langle x_\omega, u \rangle = h_{\mathbf{K}(\omega)}(u)$. Moreover, the mapping $\mathbf{x} \colon \omega \mapsto x_\omega$ is measurable, $x = \int \mathbf{x}\, d\mu \in K$ and $h_K(u) \geq \langle x, u \rangle = h(u)$. Thus $h_K = h$ as required.

The *mean width* of K is classically defined as:

$$w(K) = \int_{S^{n-1}} w(K, u) d\mu = 2 \int_{S^{n-1}} h_K(u) d\sigma,$$

where σ is the normalised Lebesgue measure on the sphere.

Lemma 1.1. *The mean width is a monotone, additive, positively homogeneous function of compact convex subsets of $\mathbb{R}^n$. Moreover, if $\mathbf{K}$ is an integrable family of convex compacts then*

$$w\left(\int \mathbf{K}\, d\mu\right) = \int w(\mathbf{K})\, d\mu.$$

Proof. Monotonicity of w follows from monotonicity (in K) of h_K. Additivity and positive homogeneity are special cases of the summability which follows from earlier observations via Fubini's Theorem. $\blacksquare_{1.1}$

As it happens it will be easier to calculate the following variant of the mean width:

Definition 1.2. The *Gaussian mean width* of a convex compact K is defined as

$$w_G(K) = \mathbb{E}[w(K, G_n)] = 2\mathbb{E}[h_K(G_n)],$$

where $G_n \sim N(0, I_n)$ (i.e., $G_n = (g_0, \ldots, g_{n-1})$ where $g_0, \ldots, g_{n-1}$ are independent random variables, $g_i \sim N(0, 1)$).

Since the distribution of $N(0, I_n)$ is rotation-invariant, the random variables $\|G_n\|_2$ and $\frac{G_n}{\|G_n\|_2}$ are independent. Let $\gamma_n = \mathbb{E}[\|G_n\|_2]$. We obtain:

$$w_G(K) = \mathbb{E}[\|G_n\|_2 w(K, G_n/\|G_n\|_2)] = \mathbb{E}[\|G_n\|_2]\mathbb{E}[w(K, G_n/\|G_n\|_2)] = \gamma_n w(K).$$

One can further calculate that

$$\gamma_n = \sqrt{2}\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}.$$

As Γ is log-convex we obtain:

$$\gamma_n \leq \sqrt{2\frac{\Gamma\left(\frac{n}{2}+1\right)}{\Gamma\left(\frac{n}{2}\right)}} = \sqrt{2\frac{n}{2}} = \sqrt{n},$$
$$\gamma_n \geq \sqrt{2\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}} = \sqrt{2\frac{n-1}{2}} = \sqrt{n-1}.$$

Whence:

$$\sqrt{n-1} \leq \gamma_n \leq \sqrt{n}$$

Thus for example, if B^n is the unit ball in $\mathbb{R}^n$ then

$$w_G(B^n) = \gamma_n w(B^n) = 2\gamma_n \approx 2\sqrt{n}.$$

Lemma 1.3. *The Gaussian mean width is a monotone, additive, positively homogeneous, function of compact convex subsets of* $\mathbb{R}^n$*, and for an integrable family* $\mathbf{K}$*:* $w_G\left(\int \mathbf{K}\, d\mu\right) = \int w_G(\mathbf{K})\, d\mu$*.*

Proof. Follows from Lemma 1.1 (or is proved identically). $\blacksquare_{1.3}$

Let us calculate the mean width of the cube $[-1, 1]^n$. The maximum $h_K(y) = \max_{x\in[-1,1]^n}\langle x, y\rangle$ is always attained at an extremal point, i.e., $h_K(y) = \max_{x\in\{-1,1\}^n}\langle x, y\rangle = \|y\|_1$. Thus:

$$w_G([-1, 1]^n) = 2\mathbb{E}[\|G_n\|_1] = 2n\mathbb{E}[|G_1|] = 2n\sqrt{\frac{2}{\pi}}.$$

Thus, for $\varepsilon > 0$ we have:

$$w_G([0, \varepsilon]^n) = \varepsilon n\sqrt{\frac{2}{\pi}}. \tag{1}$$

2. Fuzzy and continuous Vapnik-Chervonenkis classes

Let us start with a few reminders regarding the Vapnik-Chervonenkis classes. We shall follow Chapter 5 of van den Dries [vdD98].

Let us fix a set X and a family of subsets $\mathcal{C} \subseteq \mathcal{P}(X)$. Recall that $[X]^n$ denotes the collection of all subsets of X of size n, and let $\mathcal{P}^f(X) = \bigcup_{n<\omega}[X]^n$ denote the collection of finite subsets of X. For $F \in \mathcal{P}^f(X)$ and $n < \omega$ let:

$$\mathcal{C} \cap F = \{C \cap F \colon C \in \mathcal{C}\},$$
$$f_{\mathcal{C}}(n) = \max\{|\mathcal{C} \cap F| \colon F \in [X]^n\}.$$

Clearly, $f_{\mathcal{C}}(n) \leq 2^n$. Define the *Vapnik-Chervonenkis index* of $\mathcal{C}$, denoted $VC(\mathcal{C})$, to be the minimal d such that $f_{\mathcal{C}}(d) < 2^d$, or infinity if no such d exists. If $VC(\mathcal{C}) < \infty$ then $\mathcal{C}$ is a *Vapnik-Chervonenkis class*.

Let $p_d(x) = \sum_{k<d} \binom{x}{k} \in \mathbb{Q}[x]$, observing this is a polynomial of degree $d - 1$.

Fact 2.1. *If $d = VC(\mathcal{C}) < \infty$ then $f_{\mathcal{C}}(n) \leq p_d(n)$ for all n.*

This can be viewed as a dichotomy result: either $|\mathcal{C} \cap F|$ is maximal (given $|F|$) for arbitrarily large finite $F \subseteq X$, or it is always quite small (polynomial rather than exponential). It follows immediately from the following.

Fact 2.2. *Let F be a finite set, $n = |F|$, and say $\mathcal{D} \subseteq \mathcal{P}(F)$ is such that $|\mathcal{D}| > p_d(n)$. Then F admits a subset $E \subseteq F$, $|E| = d$ such that $|\mathcal{D} \cap E| = 2^d$.*

See [vdD98, Chapter 5] for the proof, which is attributed independently to Shelah [She71] and to Vapnik and Chervonenkis [VC71]. This will also follow as a special case of a result we prove below.

Let us now add a minor twist to the setting, whose motivation will become clear later on. We allow the class $\mathcal{C}$ to contain *fuzzy subsets* of X, i.e., objects C such that for each $x \in X$ at most one of $x \in C$ or $x \notin C$ holds, but possibly neither (in which case it is not known whether x belongs to C or not). This can be formalised by pair $C = (C_1, C_2)$ where $C_1, C_2 \subseteq X$ are disjoint, $C_1 = \{x \in X \colon x \in C\}$, $C_2 = \{x \in X \colon x \notin C\}$.

Let $C \sqsubseteq X$ denote that C is a fuzzy subset of X and let $\mathsf{P}(X)$ denote the collection of fuzzy subsets. If $F \subseteq X$, we say that C *determines a subset of* F if for all $x \in F$ one of $x \in C$ or $x \notin C$ does hold, in which case we define $C \cap F$ as usual, and otherwise we define $F \cap C = *$. We may then define

$$\mathcal{C} \cap F = \{C \cap F \colon C \in \mathcal{C}\} \setminus \{*\},$$
$$f_{\mathcal{C}}(n) = \max\{|\mathcal{C} \cap F| \colon F \in [X]^n\}.$$

Thus $\mathcal{C} \cap F$ is the collection of all subsets of F which members of $\mathcal{C}$ determine. Vapnik-Chervonenkis classes of fuzzy subsets of X and the corresponding index are defined as above, and the standard proofs of Fact 2.1 and of Fact 2.2 hold verbatim.

Our source for classes of fuzzy subsets of X will be the following. Let $Q \subseteq [0,1]^X$ be a collection of functions from X to $[0,1]$. For $0 \leq r < s \leq 1$ and $q \in Q$ we define a fuzzy set $q_{r,s} \sqsubseteq X$ as follows: $x \in q_{r,s}$ if $q(x) \geq s$, $x \notin q_{r,s}$ if $q(x) \leq r$, and it is unknown whether x belongs to $q_{r,s}$ or not if $r < q(x) < s$. We define $Q_{r,s} = \{q_{r,s} : q \in Q\} \subseteq \mathsf{P}(X)$. We say that Q is a Vapnik-Chervonenkis class if $Q_{r,s}$ is for every $0 \leq r < s \leq 1$. Of course the index may vary with r, s. However an easy argument shows that if Q is a Vapnik-Chervonenkis class then for every $\varepsilon > 0$ there exists $d(\varepsilon) < \omega$ which is an upper bound for the Vapnik-Chervonenkis indexes of the classes $Q_{r,r+\varepsilon}$ as r varies in $[0, 1-\varepsilon]$. Notice that in the original case where $\mathcal{C} \subseteq \mathcal{P}(X)$, if $Q = \{\chi_C : C \in \mathcal{C}\}$ is the collection of characteristic functions of members of $\mathcal{C}$ then $Q_{r,s} = \mathcal{C}$ for every $0 \leq r < s \leq 1$, so the subset case is a special case of the function case.

Definition 2.3. Let $0 \leq r_i < s_i \leq 1$ be given for $i < n$ and let $A \subseteq [0,1]^n$. We say that A *determines* a subset $w \subseteq n$ between $\bar{r}$ and $\bar{s}$ if there is a point $\bar{a} \in A$ such that $i \in w \Longrightarrow a_i \geq s_i$ and $i \notin w \Longrightarrow a_i \leq r_i$ for all $i < n$. In case $r_i = r$ and $s_i = s$ for all $i < n$ we say that A determines w between r and s.

We say that A *determines a d-dimensional ε-box* If $\varepsilon > 0$, $d \leq n$, and there are $i_0 < \ldots < i_{d-1} < n$ and $\bar{r} \in [0, 1-\varepsilon]^d$ such that $\pi_{\bar{i}}(A) \subseteq [0,1]^d$ determines every subset of d between $\bar{r}$ and $\bar{r} + \varepsilon$.

Finally for $\varepsilon \geq 0$ we say that A *determines a strict d-dimensional ε-box* if it determines a d-dimensional ε'-box for some $\varepsilon' > \varepsilon$.

Thus if $F = \{x_1, \ldots, x_n\} \subseteq X$ then $Q_{r,s} \cap F$ is in bijection with the subsets of n determined by $Q(\bar{x})$ between r and s.

Let us now relate this to the previous section. Let again $Q \subseteq [0,1]^X$ be a collection of functions. For a tuple $\bar{x} \in X^n$ and $q \in Q$ define:

$$q(\bar{x}) = (q(x_0), \ldots, q(x_n)) \in [0,1]^n,$$
$$Q(\bar{x}) = (q(\bar{x}) : q \in Q),$$
$$g_Q(n) = \sup\{w_G(Q(\bar{x})) : \bar{x} \in X^n\}.$$

Lemma 2.4. *If $A \subseteq [0,1]^n$ determines an n-dimensional ε-box then $w_G(A) \geq \varepsilon n\sqrt{\frac{2}{\pi}}$. If A determines a strict n-dimensional ε-box then $w_G(A) > \varepsilon n\sqrt{\frac{2}{\pi}}$.*

Proof. It suffices to prove the first assertion. In this case there are $\bar{r} \in [0, 1-\varepsilon]^n$ and for every $w \subseteq n$ there is $a_w \in A$ such that $a_w(i) \geq r + \varepsilon$ if $i \in w$ and $a_w(i) \leq r$ otherwise. Thus $A \supseteq (a_w : w \subseteq n) \supseteq \prod[r_i, r_i + \varepsilon] = \bar{r}_i + [0, \varepsilon]^n$. It follows that

$$w_G(A) \geq w_G([0,\varepsilon]^n) = \varepsilon n\sqrt{\frac{2}{\pi}}. \qquad \blacksquare_{2.4}$$

In other words, if A determines an n-dimensional ε-box then $\mathrm{Conv}(A)$ contains a set of the form $\bar{r} + [0, \varepsilon]^n$. The converse does not hold in general.

Proposition 2.5. *If $Q \subseteq [0,1]^X$ and $\lim \frac{g_Q(n)}{n} = 0$ then Q is a Vapnik-Chervonenkis class.*

Moreover, for any function $g(n)$ such that $\lim \frac{g(n)}{n} = 0$ and any $\varepsilon > 0$ there is $d(g, \varepsilon) < \omega$ such that for any $Q \subseteq [0,1]^X$, if $g_Q \leq g$ then $d(g, \varepsilon) \geq VC(Q_{r,r+\varepsilon})$ for all $0 \leq r \leq 1-\varepsilon$.

Proof. Let $g = g_Q$ and $\varepsilon > 0$ be given, and find n such that $\frac{g(n)}{n} < \varepsilon\sqrt{\frac{2}{\pi}}$. We claim that $d(g, \varepsilon) = n$ will do.

Indeed, assume not. Then there are $r \in [0, 1-\varepsilon]$ and $F = \{x_0, \dots, x_{n-1}\} \subseteq X$ such that $|Q_{r,r+\varepsilon} \cap F| = 2^n$, i.e., such that $Q(\bar{x})$ determines an n-dimensional ε-box. By Lemma 2.4 we have $g(n) \geq g_Q(n) \geq w_G(Q(\bar{x})) \geq \varepsilon n\sqrt{\frac{2}{\pi}}$, a contradiction. $\blacksquare_{2.5}$

For the converse a little more work is required. Let $\pi\colon \mathbb{R}^n \to \mathbb{R}^{n-1}$ be the projection on the first $n-1$ coordinates. For $A \subseteq \mathbb{R}^n$ and $a \in \mathbb{R}$ let $A_{\leq a} = A \cap (\mathbb{R}^{n-1} \times \,]-\infty, a])$, $A_{>a} = A \cap (\mathbb{R}^{n-1} \times \,]a, +\infty[)$.

Let λ denote the Lebesgue measure.

Lemma 2.6. *Let $A \subseteq [0, \ell+1]^n$ be a Borel set, $\lambda(A) > \ell^d p_d(n)$. Then at least one of the following holds:*

(i) $\lambda(\pi A) > \ell^d p_d(n-1)$.
(ii) *There is $a \in [0, \ell+1]$ such that $\lambda(\pi A_{\leq a} \cap \pi A_{>a+1}) > \ell^{d-1} p_{d-1}(n-1)$.*

Proof. For $x \in [0, \ell+1]^{n-1}$ let

$$A' = \{(x, a) \in [0, \ell+1]^{n-1} \times [0, \ell]\colon x \in \pi A_{\leq a} \cap \pi A_{>a+1}\}$$

$$f_A(x) = \int \chi_A(x, y)\, dy, \qquad f_{A'}(x) = \int \chi_{A'}(x, y)\, dy.$$

Notice that $f_{A'}(x) + \chi_{\pi A}(x) \geq f_A(x)$, integrating which yields:

$$\lambda(A') + \lambda(\pi(A)) \geq \lambda(A) > \ell^d p_d(n).$$

Recall that $p_d(n) = p_{d-1}(n-1) + p_d(n-1)$ and assume that the first case fails, i.e., that $\lambda(\pi A) \leq \ell^d p_d(n-1)$. Then:

$$\lambda(A') + \ell^d p_d(n-1) > \ell^d p_d(n) = \ell^d p_{d-1}(n-1) + \ell^d p_d(n-1)$$

whereby:

$$\lambda(A') > \ell p_{d-1}(n-1).$$

Then there is a such that $\lambda(\{x\colon (x,a) \in A'\}) > p_{d-1}(n-1)$, which is precisely the second case. $\blacksquare_{2.6}$

Lemma 2.7. *Let $A \subseteq [0, \ell+1]^n$ be a Borel set, $\lambda(A) > \ell^d p_d(n)$. Then A determines a strict d-dimensional 1-box.*

Proof. Follows immediately by induction on n using the previous Lemma for the induction step. $\blacksquare_{2.7}$

Lemma 2.8. *Let $A \subseteq [0,1]^n$ and $c > 0$, $\varepsilon \geq 0$ be such that*

$$\lambda(A + [0,c]^n) > (c+\varepsilon)^{n-d}(1-\varepsilon)^d p_d(n).$$

Then A determines a strict d-dimensional ε-box.

Proof. Let $\ell = \frac{1-\varepsilon}{c+\varepsilon}$, so $\ell + 1 = \frac{1+c}{c+\varepsilon}$. Let $B = (c+\varepsilon)^{-1}(A + [0,c]^n)$. Then $B \subseteq [0, \ell+1]^n$ and $\lambda(B) > \ell^d p_d(n)$. By Lemma 2.7 B determines a strict d-dimensional 1-box. Thus $A + [0,c]^n$ determines a strict d-dimensional $(c+\varepsilon)$-box, and A determines a strict d-dimensional ε-box. $\blacksquare_{2.8}$

In case $A \subseteq \{0,1\}^n$ then $\lambda(A + [0,1]^n) = |A|$. If in addition $|A| > p_d(n) = (1+0)^{n-d}(1-0)^d p_d(n)$ then A determines a (strict) d-dimensional 0-box, i.e., an ε-box for some arbitrarily small $\varepsilon > 0$. But given that $A \subseteq \{0,1\}^d$ this is only possible if A determines a d-dimensional 1-box. Thus Fact 2.2 follows as a special case of Lemma 2.8.

Now let us show that if $\lambda(A + [0,c]^n)$ is small then A is small in a different way, namely has small Gaussian mean width.

Lemma 2.9. *Let $A \subseteq [0,1]^n$, $c > 0$. Then*

$$w_G(A) \leq (1+c)\sqrt{2n \log(\lambda(A + [0,c]^n)/c^n)}.$$

Proof. Let us first observe that if $\varphi\colon \mathbb{R}^n \to \mathbb{R}$ is a linear functional then the convexity of the exponential function implies:

$$\begin{aligned}\exp(\varphi(x)) &= \exp\left(\int_{y \in x+[-c,c]^n} \varphi(y)\,(2c)^{-n} d\lambda(y)\right) \\ &\leq (2c)^{-n} \int_{x+[-c,c]^n} \exp(\varphi(y))\, d\lambda(y).\end{aligned}$$

Let $A' = 2A - 1 \subseteq [-1,1]^n$ and $B' = A' + [-c,c]^n = 2(A + [0,c]^n) - (1+c)$. Then by the previous observation we have:

$$\sup\{\exp(\varphi(x))\colon x \in A'\} \leq (2c)^{-n} \int_{B'} \exp(\varphi(x))\, d\lambda(x).$$

Let $\beta > 0$ be an arbitrary parameter for the time being. For a fixed $x \in \mathbb{R}^n$ we have $\beta\langle x, G_n\rangle \sim N(0, \beta^2\|x\|^2)$, and a straightforward calculation yields $\mathbb{E}[\exp(\beta\langle x, G_n\rangle)] =$

$\exp(\beta^2\|x\|^2/2)$. Using concavity of the logarithm we obtain:

$$\begin{aligned} w_G(A') &= 2\mathbb{E}\left[\sup\left\{\langle x, G_n\rangle \colon x \in A'\right\}\right] \\ &= \frac{2}{\beta}\mathbb{E}\left[\log\left(\sup\left\{\exp(\beta\langle x, G_n\rangle) \colon x \in A'\right\}\right)\right] \\ &\leq \frac{2}{\beta}\log\left(\mathbb{E}\left[(2c)^{-n}\int_{B'}\exp(\beta\langle x, G_n\rangle)\, d\lambda(x)\right]\right) \\ &= \frac{2}{\beta}\log\left((2c)^{-n}\int_{B'}\mathbb{E}\left[\exp(\beta\langle x, G_n\rangle)\right]\, d\lambda(x)\right) \\ &= \frac{2}{\beta}\log\left((2c)^{-n}\int_{B'}\exp\left(\frac{\beta^2\|x\|^2}{2}\right) d\lambda(x)\right) \\ &\leq \frac{2}{\beta}\log\left((2c)^{-n}\lambda(B')\exp\left(\frac{\beta^2(1+c)^2 n}{2}\right)\right) \\ &= \frac{2\log(\lambda(B')/(2c)^n)}{\beta} + \beta(1+c)^2 n. \end{aligned}$$

Minimum is attained when $\beta = \frac{\sqrt{2\log(\lambda(B')/(2c)^n)}}{(1+c)\sqrt{n}}$, and substituting we obtain:

$$w_G(A') \leq 2(1+c)\sqrt{2n\log(\lambda(B')/(2c)^n)}.$$

Finally, $w_G(A) = w_G(A')/2$ and $\lambda(A+[0,c]^n) = \lambda(B')/2^n$, whence the desired inequality. $\blacksquare_{2.9}$

Lemma 2.10. *Let $A \subseteq [0,1]^n$ be finite, $|A| = N$. Then $w_G(A) \leq \sqrt{2n\log N}$.*

Proof. For c small enough we have $\lambda(A+[0,c]^n) = Nc^n$, so $w_G(A) \leq (1+c)\sqrt{2n\log N}$ and thus $w_G(A) \leq \sqrt{2n\log N}$. $\blacksquare_{2.10}$

Our proof of Lemma 2.9 is based on a direct argument due to M. Talagrand for Lemma 2.10.

Theorem 2.11. *Let $Q \subseteq [0,1]^X$ be a collection of functions. Then the following are equivalent:*

(i) *Q is a Vapnik-Chervonenkis class.*
(ii) *For every $\varepsilon > 0$ there is d such that for every $\bar{x} \in X^d$, $Q(\bar{x})$ does* not *determine a d-dimensional ε-box.*
(iii) $\lim \frac{g_Q(n)}{n} = 0$.

Proof. For (i) $\Longrightarrow$ (ii) we shall prove the contra-positive. So assume that for some $\varepsilon > 0$ this fails, i.e., for all d there are $\bar{x} \in X^d$, $r_0, \ldots, r_{d-1}$ and $\left\{q_w \colon w \subseteq d\right\} \subseteq Q$ satisfying $q_w(x_i) \leq r_i$ if $i \in w$ and $q_w(x_i) \geq r_i + \varepsilon$ if $i \notin w$. Thus there must be a subset of r_i of size at least $d' = \lceil d\varepsilon/2\rceil$ which are at distance at most $\varepsilon/2$ from one another, and we might as well assume these are $r_0 \leq r_1 \leq \ldots \leq r_{d'-1} \leq r_0 + \varepsilon/2 = r$. For $i < d'$ we $q_w(x_i) \leq r$

if $i \in w$ and $q_w(x_i) \geq r + \varepsilon/2$ if $i \notin w$. This works for arbitrarily large d, and thus for arbitrarily large d'. Thus Q is not a Vapnik-Chervonenkis class. (And considering $d' = \lfloor d\varepsilon/m \rfloor$ we can get $q_w(x_i) \geq r + \varepsilon(1 - \frac{1}{m})$.)

Let us now show (ii) $\Longrightarrow$ (iii). Let us fix $\varepsilon > 0$, and let d be as in the hypothesis. By Lemma 2.8 and Lemma 2.9 we have for all $c > 0$ and $\bar{x} \in X^n$:

$$\lambda(Q(\bar{x}) + [0,c]^n) \leq (c+\varepsilon)^{n-d}(1-\varepsilon)^d p_d(n),$$
$$w_G(Q(\bar{x})) \leq (1+c)\sqrt{2n \log(\lambda(Q(\bar{x}) + [0,c]^n)/c^n)}$$

Whereby:

$$g_Q(n) \leq (1+c)\sqrt{2n \log\left(\left(1 + \frac{\varepsilon}{c}\right)^n \left(\frac{1-\varepsilon}{c+\varepsilon}\right)^d p_d(n)\right)},$$
$$\frac{g_Q(n)}{n} \leq (1+c)\sqrt{2\log\left(1 + \frac{\varepsilon}{c}\right) + \frac{1}{n}\log\left(\left(\frac{1-\varepsilon}{c+\varepsilon}\right)^d p_d(n)\right)}.$$

As n goes to infinity the second term under the root disappears. In addition we have $\log(1 + \varepsilon/c) \leq \varepsilon/c$, and we obtain: $\overline{\lim} \frac{g_Q(n)}{n} \leq (1+c)\sqrt{\frac{2\varepsilon}{c}}$. Minimum is reached when $c = 1$ in which case $\overline{\lim} \frac{g_Q(n)}{n} \leq \sqrt{8\varepsilon}$. This holds for every $\varepsilon > 0$, whereby $\lim \frac{g_Q(n)}{n} = 0$ as desired.

(iii) $\Longrightarrow$ (i) was proved in Proposition 2.5. $\blacksquare_{2.11}$

Notice that the proof also tells us in fact something more precise:

Corollary 2.12. *Assume that $\overline{\lim} \frac{g_Q(n)}{n} = C > 0$. Then for some r the class $Q_{r,r+C^2/8}$ is* not *a Vapnik-Chervonenkis class.*

In case $Q \subseteq \{0,1\}^X$, i.e., for collection of characteristic functions, a box if exists has size one, so we can get better bounds .

Proposition 2.13. *Let $\mathcal{C} \subseteq \mathcal{P}(X)$, $Q = \{\chi_C : C \in \mathcal{C}\} \subseteq [0,1]^X$. Then $g_Q(n) \leq \sqrt{2n \log p_d(n)}$ and for n big enough $g_Q(n) \leq \sqrt{2dn \log n}$, where $d = VC(\mathcal{C})$.*

Proof. Let $d = VC(\mathcal{C}) < \infty$. For every $\bar{x} \in X^n$ we have $|Q(\bar{x})| \leq f_{\mathcal{C}}(n) \leq p_d(n)$ by Fact 2.1. For n large enough we have $p_d(n) \leq n^d$ and by Lemma 2.10:

$$g_Q(n) \leq \sqrt{2n \log p_d(n)} \leq \sqrt{2dn \log n}. \qquad \blacksquare_{2.13}$$

We can now switch to a more symmetric situation. Let X and Y be two sets, $S \subseteq X \times Y$. For $x \in X$ let $S_x = \{y \in Y : (x,y) \in S\}$ and for $y \in Y$ let $S^y = \{x \in X : (x,y) \in S\}$. Thus S gives rise to two families of subsets $S^Y = \{S^y : y \in Y\} \subseteq \mathcal{P}(X)$ and $S_X = \{S_x : x \in X\} \subseteq \mathcal{P}(Y)$.

Similarly, if $S \sqsubseteq X \times Y$ we may define $S^y \sqsubseteq X$ by $x \in S^y \iff (x, y) \in S$ and $x \notin S^y \iff (x, y) \notin S$. Continuing as above we obtain two families of fuzzy subsets $S^Y \subseteq \mathsf{P}(X)$ and $S_X \subseteq \mathsf{P}(Y)$.

Fact 2.14. *Let $S \sqsubseteq X \times Y$. Then S_X is a Vapnik-Chervonenkis class if and only if S^Y is, in which case $VC(S_X) \leq 2^{VC(S^Y)}$ and vice versa.*

We say in this case that S is a dependent *relation.*

Proof. In case $S \subseteq X \times Y$ this is proved in [vdD98, Chapter 5]. The case of a fuzzy relation, while not considered there, is identical. $\blacksquare_{2.14}$

Finally, a function $\varphi \colon X \times Y \to [0, 1]$ gives rise to two families of functions $\varphi^Y = \{\varphi^y \colon y \in Y\} = \{\varphi(\cdot, y) \colon y \in Y\} \subseteq [0, 1]^X$ and similarly $\varphi_X = \{\varphi_x \colon x \in X\} \subseteq [0, 1]^Y$.

Proposition 2.15. *Let X and Y be sets, $\varphi \colon X \times Y \to [0, 1]$ any function. Then φ^Y is a Vapnik-Chervonenkis class if and only if φ_X is.*

In that case we say that φ is dependent.

Proof. For $0 \leq r < s \leq 1$ define $\varphi_{r,s} \sqsubseteq X \times Y$ as usual. Then $(\varphi_{r,s})_X = (\varphi_X)_{r,s}$ is a Vapnik-Chervonenkis class if and only if $(\varphi_{r,s})^Y = (\varphi^Y)_{r,s}$ is. $\blacksquare_{2.15}$

Lemma 2.16. *A uniform limit of dependent functions is dependent.*

Proof. Let $\varphi_n \colon X \times Y \to [0, 1]$ be dependent converging uniformly to φ. Assume φ is independent, so say $\varphi_{r,r+3\varepsilon}$ is independent for some $\varepsilon > 0$ and $r \in [0, 1 - 3\varepsilon]$. Let n be large enough such that $|\varphi - \varphi_n| \leq \varepsilon$. Then $(\varphi_n)_{r+\varepsilon,r+2\varepsilon}$ is independent, contrary to hypothesis. $\blacksquare_{2.16}$

3. Crushing convex compacts

Let $K \subseteq \mathbb{R}^n$ be a convex compact, $u \in S^{n-1}$ a fixed direction vector. We would like to construct a new convex compact K_u by crushing all points below the hyperplane $u^\perp$ to the hyperplane. Define the two half spaces and a mapping $S \colon \mathbb{R}^n \to \mathbb{R}^n$ as follows:

$$H^+ = \{x \in \mathbb{R}^n \colon \langle x, u \rangle \geq 0\}, \qquad H^- = \{x \in \mathbb{R}^n \colon \langle x, u \rangle \leq 0\},$$

$$S(x) = \begin{cases} x & x \in H^+ \\ P_{u^\perp}(x) & x \in H^-. \end{cases}$$

We then let

$$K_u = \mathrm{Conv}(S(K)) = \mathrm{Conv}((K \cap H^+) \cup P_{u^\perp}(K \cap H^-)).$$

We would like to show that $w_G(K_u) \leq w_G(K)$.

Lemma 3.1. *Let K, u and K_u be as above. If $h_K(-u) \leq 0$ then $K = K_u$. If $h_K(-u) \geq 0$ then we have for $y \in \mathbb{R}^n$, $y' = P_{u^\perp}(y)$:*

$$h_{K_u}(y) = \max(h_K(y), h_K(P_{u^\perp}(y))) \qquad y \in H^+,$$

$$h_{K_u}(y) \leq \min\{h_K(z) \colon z \in [y, P_{u^\perp}(y)]\} \qquad y \in H^-.$$

Proof. If $h_K(-u) \leq 0$ then $K \subseteq H^+$ and $S(K) = K$. Consider the case $h_K(-u) \geq 0$. In that case clearly $h_{K_u}(u) = 0$ and h_{K_u} agrees with h_K on $u^\perp$. Let us also observe that if $y \in \mathbb{R}^n$ then the $h_{K_u}(y) = \langle x, y\rangle$ for some extremal point $x \in K_u$, in which case we have in fact $x \in S(K)$. Thus we always have $h_{K_u}(y) = \langle S(x), y\rangle$, $x \in K$.

Let us consider the case where $y \in H^+$, i.e., $y = y' + \lambda u$ where $y' \perp u$ and $\lambda \geq 0$. Say $h_K(y) = \langle x, y\rangle$, $x \in K$. Then $\langle S(x), y\rangle \geq \langle x, y\rangle$ and thus $h_{K_u} \geq h_K(y)$. Since h_{K_u} is sub-additive we also have $h_{K_u}(y) = h_{K_u}(y' + \lambda u) + h_{K_u}(-\lambda u) \geq h_{K_u}(y')$. Thus $h_{K_u}(y) \geq \max(h_K(y), h_K(y'))$. On the other hand, we know that $h_{K_u}(y) = \langle S(x), y\rangle$ for some $x \in K$. If $x \in H^+$ then $h_{K_u}(y) \leq h_K(y)$. If $x \in H^-$ then $\langle S(x), y\rangle = \langle S(x), y'\rangle = \langle x, y'\rangle$ so $h_{K_u}(y) \leq h_K(y')$. Either way $h_{K_u}(y) \leq \max(h_K(y), h_K(y'))$ and the first case is proved.

Now assume $y \in H^-$. Let us make first some general observations. First, if $h_{K_u}(y) = \langle S(x), y\rangle$, $x \in K$, then $\langle S(x), y\rangle \leq \langle x, y\rangle$ whereby $h_{K_u}(y) \leq h_K(y)$. Now write $y = y' - \lambda u$ where $y' \perp u$ and $\lambda \geq 0$. Let $z \in [y, y'] \subseteq H^-$, i.e., $z = y' - \mu u$ for $\mu \in [0, \lambda]$. Then $h_{K_u}(y) \leq h_{K_u}(z) + h_{K_u}(-(\lambda - \mu)u) = h_{K_u}(z) \leq h_K(z)$. We have thus shown that $h_{K_u}(y) \leq \min\{h_K(z) \colon z \in [y', y]\}$. $\blacksquare_{3.1}$

Can the second inequality be improved to an equality? Either way, the inequalities we have suffice to prove:

Lemma 3.2. *Let K and K_u be as above, $y' \in u^\perp$ and $\lambda \geq 0$. Then*

$$h_{K_u}(y' + \lambda u) + h_{K_u}(y' - \lambda u) \leq h_K(y' + \lambda u) + h_K(y' - \lambda u).$$

Proof. Consider the mapping $s(t) = h_{K_u}(y' + tu)$, which we know to be convex. If $s(0) \leq s(\lambda)$ then $h_{K_u}(y' + \lambda u) = h_K(y' + \lambda u)$, and we already know that $h_{K_u}(y' - \lambda u) \leq h_K(y' - \lambda u)$.

If $s(0) \geq s(\lambda)$ then $h_{K_u}(y' + \lambda u) = h_K(y')$. By convexity of s it must be decreasing for all $t \leq 0$, so in particular

$$h_{K_u}(y - \lambda u) \leq \min\{s(t) \colon t \in [-\lambda, 0]\} = s(0) = h_K(y').$$

Thus:

$$h_{K_u}(y' + \lambda u) + h_{K_u}(y' - \lambda u) \leq h_K(2y') \leq h_K(y' + \lambda u) + h_K(y' - \lambda u). \qquad \blacksquare_{3.2}$$

Proposition 3.3. *Let K and K_u be as above. Then $w(K_u) \leq w(K)$ and $w_G(K_u) \leq w_G(K)$.*

Proof. It will be enough to prove the first inequality. For $y \in S^{n-1}$ let y' always denote $P_{u^\perp}(y)$ and $\lambda = |\langle y, u\rangle|$. We have:

$$\begin{aligned}
2w(K_u) &= 2\int_{S^{n-1}} w(K_u, y)\, d\sigma(y) \\
&= \int_{S^{n-1}} (w(K_u, y' + \lambda u) + w(K_u, y' - \lambda u))\, d\sigma(y) \\
&= \int_{S^{n-1}} \big(h_{K_u}(y' + \lambda u) + h_{K_u}(y' - \lambda u) \\
&\qquad\qquad + h_{K_u}(-y' + \lambda u) + h_{K_u}(-y' - \lambda u)\big)\, d\sigma(y) \\
&\le \int_{S^{n-1}} \big(h_{K}(y' + \lambda u) + h_{K}(y' - \lambda u) \\
&\qquad\qquad + h_{K}(-y' + \lambda u) + h_{K}(-y' - \lambda u)\big)\, d\sigma(y) \\
&= \ldots = 2w(K).
\end{aligned}$$

■3.3

Now let $K \subseteq \mathbb{R}^n$ be a convex compact, $(e_i \colon i < n)$ the canonical base, and define

$$\begin{aligned}
K^+ &= (\ldots(K_{e_0})_{e_1}\ldots)_{e_{n-1}} \\
&= \mathrm{Conv}\big((x_0 \vee 0, \ldots, x_{n-1} \vee 0) \colon \bar{x} \in K\big).
\end{aligned}$$

Corollary 3.4. *Let $K \subseteq \mathbb{R}^n$ be a convex compact. Then $w_G(K^+) \le w_G(K)$.*

We remind the reader that for $x, y \in [0,1]$ we define $\neg x = 1 - x \in [0,1]$ and $x \dot{-} y = \max(x - y, 0) \in [0,1]$. Moreover, for every $n \ge 1$, the family of functions $[0,1]^n \to [0,1]$ one can construct with the three operations $\{\frac{x}{2}, \neg x, x \dot{-} y\}$ is dense in the space all continuous functions from $[0,1]^n$ to $[0,1]$ (see [BU]).

Corollary 3.5. *Let X and Y be sets, $\varphi, \psi \colon X \times Y \to [0,1]$. Then $g_{(\neg\varphi)^Y} = g_{\varphi^Y}$, $g_{(\varphi/2)^Y} = \frac{1}{2} g_{\varphi^Y}$ and $g_{(\varphi \dot{-} \psi)^Y} \le g_{\varphi^Y} + g_{\psi^Y}$.*

Thus, if φ and ψ are Vapnik-Chervonenkis classes then so are $\neg\varphi$, $\frac{1}{2}\varphi$ and $\varphi \dot{-} \psi$.

Proof. Clearly $g_{(\neg\varphi)^Y} = g_{\varphi^Y}$, $g_{(\varphi/2)^Y} = \frac{1}{2} g_{\varphi^Y}$. We are left with $g_{(\varphi \dot{-} \varphi)^Y} \le g_{\varphi^Y} + g_{\psi^Y}$.

Consider the function $\varphi - \psi \colon X \times Y \to [-1,1]$, and observe that for $\bar{x} \in X^n$ we have $(\varphi \dot{-} \psi)^Y(\bar{x}) = \big((\varphi - \psi)^Y(\bar{x})\big)^+ \subseteq [0,1]^n$. We thus have:

$$w_G\left((\varphi \dot{-} \psi)^Y(\bar{x})\right) = w_G\left(\left((\varphi - \psi)^Y(\bar{x})\right)^+\right) \le w_G\left((\varphi - \psi)^Y(\bar{x})\right).$$

On the other hand we also have $(\varphi - \psi)^Y \subseteq \varphi^Y - \psi^Y \subseteq [0,1]^X$, and for $\bar{x} \in X^n$:

$$w_G\left((\varphi - \psi)^Y(\bar{x})\right) \le w_G\left(\varphi^Y(\bar{x}) - \psi^Y(\bar{x})\right) = w_G\left(\varphi^Y(\bar{x})\right) + w_G\left(\psi^Y(\bar{x})\right).$$

Thus $w_G\left((\varphi \dot{-} \psi)^Y(\bar{x})\right) \le w_G\left(\varphi^Y(\bar{x})\right) + w_G\left(\psi^Y(\bar{x})\right)$, whereby $g_{(\varphi \dot{-} \psi)^Y} \le g_{\varphi^Y} + g_{\psi^Y}$. ■3.5

Lemma 3.6. *Let X and Y be sets, $\varphi, \psi \colon X \times Y \to [0,1]$ dependent. Then $\neg\varphi$, $\frac{1}{2}\varphi$ and $\varphi \dot{-} \psi$ are dependent as well.*

Proof. By Corollary 3.5. $\blacksquare_{3.6}$

Proposition 3.7. *Let X and Y be sets, $\varphi_n \colon X \times Y \to [0,1]$ dependent functions for $n < \omega$, and let $\psi \colon [0,1]^\omega \to [0,1]$ be an arbitrary continuous function. Then $\psi \circ (\varphi_n) \colon X \times Y \to [0,1]$ is dependent.*

Proof. By results in [BU] one can approximate ψ uniformly with expressions written with $\neg$, $\frac{1}{2}$ and $\dot{-}$. Such expressions in the φ_n are dependent by Lemma 3.6. Thus $\psi \circ (\varphi_n)$ is a uniform limit of dependent functions, and is therefore dependent by Lemma 2.16. $\blacksquare_{3.7}$

4. Random dependent relations and functions

In this section X and Y will be sets as before. However, we will be interested here in dependent relations and functions on $X \times Y$ which may vary (randomly).

Let Ω be an arbitrary set for the time being. A family of relations on $X \times Y$, indexed by Ω, can be viewed as a relation $S \subseteq \Omega \times X \times Y$. For every $\omega \in \Omega$ we obtain a relation $S_\omega \subseteq X \times Y$ and we may view S equivalently as a function $S \colon \Omega \to \mathcal{P}(X \times Y)$. Similarly, a family of $[0,1]$-valued functions on $X \times Y$ will be given as $\varphi \colon \Omega \times X \times Y \to [0,1]$ or equivalently as $\varphi \colon \Omega \to [0,1]^{X \times Y}$ sending $\omega \mapsto \varphi_\omega = \varphi(\omega, \cdot, \cdot)$. The usual passage from S to its characteristic function χ_S commutes with these equivalent presentations.

We say that such a family $S = \{S_\omega \colon \omega \in \Omega\}$ is *uniformly dependent* if there is $d = d(S)$ such that $VC((S_\omega)^Y) \leq d$ for every $\omega \in \Omega$. Similarly a family $\varphi = \{\varphi_\omega \colon \omega \in \Omega\}$ is uniformly dependent if for every $\varepsilon > 0$ there is $d = d(\varphi, \varepsilon)$ such that $VC\left((\varphi_\omega^Y)_{[r,r+\varepsilon]}\right) \leq d$ for every $r \in [0, 1-\varepsilon]$ and $\omega \in \Omega$. Clearly S is uniformly dependent if and only if χ_S is.

It follows from the proof of Theorem 2.11 that $\varphi = \{\varphi_\omega \colon \omega \in \Omega\}$ is uniformly dependent if and only if there is a function $g \colon \mathbb{N} \to \mathbb{R}$ such that $\lim \frac{g(n)}{n} = 0$ and $g_{\varphi_\omega^Y} \leq g$ for every ω. Indeed, in case φ is uniformly dependent then for every ω, n and ε we obtain:

$$g_{\varphi_\omega^Y}(n) \leq 2n\sqrt{2\varepsilon + \frac{d(\varphi,\varepsilon)}{n} \log\left(\frac{1-\varepsilon}{1+\varepsilon}\right) + \frac{\log p_{d(\varphi,\varepsilon)}(n)}{n}}.$$

Then a function g as desired can be obtained by:

$$g(n) = \inf_{0<\varepsilon<1} 2n\sqrt{2\varepsilon + \frac{d(\varphi,\varepsilon)}{n} \log\left(\frac{1-\varepsilon}{1+\varepsilon}\right) + \frac{\log p_{d(\varphi,\varepsilon)}(n)}{n}}.$$

Let us now consider *random* relations and functions on $X \times Y$. We fix a probability space $(\Omega, \mathfrak{B}, \mu)$. From now on we will only consider families S or φ such that for $(x, y) \in X \times Y$ the event $\{\omega \colon (x, y) \in S_\omega\}$ or the function $\omega \mapsto \varphi_\omega(x, y)$ are measurable. We may then define functions $\mathbb{P}[S], \mathbb{E}[\varphi] \colon X \times Y \to [0,1]$ by

$$\mathbb{P}[S](x,y) = \mathbb{P}[(x,y) \in S], \qquad \mathbb{E}[\varphi](x,y) = \mathbb{E}[\varphi(x,y)].$$

If S is measurable then so is χ_S which is given by $(\chi_S)_\omega = \chi_{(S_\omega)}$ and then $\mathbb{E}[\chi_S] = \mathbb{P}[S]$.

Theorem 4.1. *Let X, Y be countable sets, $\varphi_\omega\colon X \times Y \to [0,1]$ a random family of functions on $X \times Y$. Then $g_{\mathbb{E}[\varphi]^Y} \leq \mathbb{E}[g_{\varphi_\omega^Y}]$ (and the latter is measurable).*

In particular, if φ is uniformly dependent then $\mathbb{E}[\varphi]$ is dependent.

Proof. Let us fix n and let $\bar{x} \in X^n$. Define

$$\mathbf{K}_{\bar{x}}(\omega) = \overline{\mathrm{Conv}}\big(\varphi_\omega^Y(\bar{x})\big) \subseteq [0,1]^n.$$

Each $\mathbf{K}_{\bar{x}}(\omega)$ is a convex compact and

$$g_{\varphi_\omega^Y}(n) = \sup_{\bar{x} \in X^n} w_G(\mathbf{K}_{\bar{x}}(\omega)).$$

Since Y is assumed countable the family $\mathbf{K}_{\bar{x}}$ is measurable for every $\bar{x} \in X$. It is moreover bounded and therefore integrable. Since X is also assumed countable the function $\omega \mapsto g_{\varphi_\omega^Y}(n)$ is measurable as well. For a fixed tuple $\bar{x}$ we have $\mathbb{E}[\varphi]^Y(\bar{x}) \subseteq \mathbb{E}[\mathbf{K}_{\bar{x}}]$. Thus

$$w_G\left(\mathbb{E}[\varphi]^Y(\bar{x})\right) \leq w_G(\mathbb{E}[\mathbf{K}]_{\bar{x}}) = \mathbb{E}[w_G(\mathbf{K}_{\bar{x}})] \leq \mathbb{E}[g_{\varphi_\omega^Y}(n)].$$

It follows that $g_{\mathbb{E}[\varphi]^Y} \leq \mathbb{E}[g_{\varphi_\omega^Y}]$ as desired.

If φ is uniformly dependent then there is $g\colon \mathbb{N} \to \mathbb{R}$ such that $\frac{g(n)}{n} \to 0$ and $g \geq g_{\varphi_\omega^Y}$ for every $\omega \in \Omega$. Then $g_{\mathbb{E}[\varphi]^Y} \leq g$ as well and by Theorem 2.11 $\mathbb{E}[\varphi]$ is dependent. $\blacksquare_{4.1}$

Corollary 4.2. *Let X, Y be sets, $\varphi = \{\varphi_\omega \colon \omega \in \Omega\}$ a measurable family of uniformly dependent functions. Then $\mathbb{E}[\varphi]\colon X \times Y \to [0,1]$ is dependent.*

Proof. If not then this is witnesses on countable subsets $X_0 \subseteq X$ and $Y_0 \subseteq Y$, contradicting Theorem 4.1. $\blacksquare_{4.2}$

Corollary 4.3. *Let X, Y be sets, $S = \{S_\omega \colon \omega \in \Omega\}$ a measurable family of uniformly dependent relations. Then $\mathbb{P}[S]\colon X \times Y \to [0,1]$ is dependent.*

Proof. Apply Corollary 4.2 to χ_S. $\blacksquare_{4.3}$

5. Dependent and randomised theories

In this final section we settle the model theoretic problem which motivated the present study. This consists mostly of translating consequences of previous results to the model theoretic setting. In order to avoid blowing this section up disproportionately with a lot of introductory material we assume the reader is already familiar with the basics of classical model theory (see Poizat [Poi85]) and its generalisation to continuous logic (see [BU]).

Let T be a (classical or continuous) first order theory.

Definition 5.1. We say that a formula $\varphi(\bar{x}, \bar{y})$ is *dependent* in T if for every $\mathcal{M} \vDash T$, $\varphi^{\mathcal{M}}$ is dependent on $M^n \times M^m$.

We say that T is *dependent* if all formulae are.

In the case of a classical theory this is equivalent to the original definition (see Laskowski [Las92]) and it extends naturally to continuous logic. If T is a continuous dependent theory then by Lemma 2.16 every definable predicate in T is dependent. In addition, it is easy to see using compactness that if $\varphi(\bar{x}, \bar{y})$ is dependent in T then it is uniformly so in all models of T.

For a continuous language $\mathcal{L}$ let $\mathcal{L}^R$ consists of a n-ary predicate symbol $\mathbb{E}[\varphi(\bar{x})]$ for every n-ary $\mathcal{L}$-formula $\varphi(\bar{x})$.

Theorem 5.2. *For every $\mathcal{L}$-theory T (dependent or not) there is a (unique) $\mathcal{L}^R$-theory T^R such that:*

(i) *For every $p(\bar{x}) \in \mathrm{S}_n(T^R)$ there is a unique Borel probability measure ν_p on $\mathrm{S}_n(T^R)$ such that for every n-ary predicate symbol $\mathbb{E}[\varphi(\bar{x})] \in \mathcal{L}^R$:*

$$\mathbb{E}[\varphi(\bar{x})]^p = \int \varphi^q \, d\nu_p(q).$$

The mapping $p \mapsto \nu_p$ is a bijection between $\mathrm{S}_n(T^R)$ and the space of regular Borel probability measures. We will consequently identify the two spaces, thus identifying p with ν_p.

(ii) *The topology on $\mathrm{S}_n(T^R)$ is the one of weak convergence. In other words, this is the weakest topology such that for every continuous function $\varphi\colon \mathrm{S}_n(T) \to \mathbb{C}$ the mapping $\mu \mapsto \int \varphi \, d\mu$ is continuous.*

(iii) *For a mapping $f: m \to n$, the corresponding mapping $f^{*,R}\colon \mathrm{S}_n(T^R) \to \mathrm{S}_m(T^R)$ is given by associating to each type in $\mathrm{S}_n(T^R)$, being a measure on $\mathrm{S}_n(T)$, its image measure on $\mathrm{S}_m(T)$ via the application $f^*\colon \mathrm{S}_n(T) \to \mathrm{S}_m(T)$. (Since $f^*\colon \mathrm{S}_n(T) \to \mathrm{S}_m(T)$ is continuous between compact spaces, the image of a regular measure is regular.)*

(iv) *The distance predicate coincides with $\mathbb{E}[d(x, y)]$.*

Moreover, T^R eliminates quantifiers.

Since every classical first order theory can be viewed as a continuous theory, the same applies if T is a classical theory. In this case we may prefer to write $\mathbb{P}[\varphi(\bar{x})]$ instead of $\mathbb{E}[\varphi(\bar{x})]$. (In fact, the precise counterpart of $\mathbb{E}[\varphi(\bar{x})]$ is $\mathbb{P}[\neg\varphi(\bar{x})]$ since 1 is "False", but this is a minor issue.) In particular the distance predicate is then given by $\mathbb{P}[x \neq y]$.

Proof. Uniqueness follows from the fact that the type spaces are entirely described.

In the case T is a classical theory, the explicit construction appears in [BK], where Keisler's original construction [Kei99] is transferred from classical logic to the more adequate setting of continuous logic.

A similar construction can in principle be carried out when T is a continuous theory. Alternatively, let us consider $\mathrm{S}_n(T^R)$ as a mere symbol denoting the space of regular Borel probability measures on $\mathrm{S}_n(T)$. Let $\mathrm{S}(T^R)$ denote the mapping $n \mapsto \mathrm{S}_n(T^R)$ and let us equip it with the topological and functorial structure described in items (ii),(iii). Then $\mathrm{S}(T^R)$ is an open Hausdorff type-space functor in the sense of [Ben03]. The predicates of $\mathcal{L}^R$ can be interpreted in models of $\mathrm{S}(T^R)$ as per item (i), in which case $\mathbb{E}[d(x,y)]$ defines a metric on the models. By results appearing in [BU] a continuous theory T^R exists in *some* language whose type space functor is $\mathrm{S}_n(T^R)$. Since the n-ary $\mathcal{L}$-formulae are dense among all continuous functions $\mathrm{S}_n(T) \to [0,1]$, the atomic $\mathcal{L}^R$-formulae $\mathbb{E}[\varphi(\bar{x})]$ separate types. It follows that T^R can be taken to be an $\mathcal{L}^R$-theory and that it eliminates quantifiers as such. We leave the details to the reader. $\blacksquare_{5.2}$

Members of models of T^R should be thought of as random variables in models of T. If $\mathbf{a}, \mathbf{b}, \ldots \in \mathcal{M} \vDash T^R$ then their type $\mathrm{tp}^R(\mathbf{a}, \mathbf{b}, \ldots)$, viewed as a probability measure, should be thought of as the distribution measure of the $\mathrm{S}_n(T)$-valued random variable $\omega \mapsto \mathrm{tp}(\mathbf{a}(\omega), \mathbf{b}(\omega), \ldots)$. Similarly $\mathbb{E}[\varphi(\mathbf{a}, \mathbf{b}, \ldots)]$ is the expectation of the random variable $\omega \mapsto \varphi(\mathbf{a}(\omega), \mathbf{b}(\omega), \ldots)$, and so on. As we said in the introduction it is natural to ask whether the randomisation of a dependent theory is dependent.

Theorem 5.3. *Let T be a dependent first order theory (classical or continuous). Then T^R is dependent as well.*

Proof. Every classical theory can be identified with a continuous theory via the identification of T with 0, of F with 1 and of $=$ with d. We may therefore assume that T is continuous.

Let us first consider a formula of the form $\varphi(\bar{x}, \bar{y}) = \mathbb{E}[\psi(\bar{x}, \bar{y})]$. Let $\mathcal{M} \vDash T^R$, and we need to show that $\varphi^{\mathcal{M}}$ is dependent on $M^n \times M^m$. Let us enumerate $M^n = \{\bar{a}_i : i \in I\}$, $M^m = \{\bar{b}_j : j \in J\}$. Let $p = \mathrm{tp}(M^n, M^m/\varnothing)$. We may write it as $p(\bar{x}_i, \bar{y}_j)_{i \in I, j \in J} \in \mathrm{S}_{I \cup J}(T^R)$, and identify it with a probability measure μ on $\Omega = \mathrm{S}_{(I \times n) \cup (J \times m)}(T)$ such that for every formula $\rho(\bar{z})$ of the theory T, $\bar{z} \subseteq \{\bar{x}_i, \bar{y}_j\}_{i \in I, j \in J}$:

$$\mathbb{E}[\rho(\bar{z})]^p = \int_\Omega \rho(\bar{z})^q \, d\mu(q).$$

For $i \in I$, $j \in J$ and $q \in \Omega$ define: $\chi_q(i,j) = \psi(\bar{x}_i, \bar{y}_j)^q$. Then $\chi = \{\chi_q : q \in \Omega\}$ is a measurable family of $[0,1]$-valued functions on $I \times J$ and $\varphi(\bar{a}_i, \bar{b}_j) = \varphi(\bar{x}_i, \bar{y}_j)^p = \mathbb{E}[\chi](i,j)$ where expectation is with respect to μ. Since T is dependent the family $\{\chi_q : q \in \Omega\}$ is uniformly dependent. By Corollary 4.2 $\mathbb{E}[\chi] : I \times J \to [0,1]$ is dependent. Equivalently, $\varphi : M^n \times M^m \to [0,1]$ is dependent.

We have thus shown that every atomic formula is dependent. By Lemma 3.6 every quantifier free formula is dependent. By quantifier elimination and Lemma 2.16 every formula is dependent. $\blacksquare_{5.3}$

We conclude this paper with a few extensions of classical results regarding dependent formulae and theories to continuous logic.

Lemma 5.4. *The following are equivalent for a formula $\varphi(\bar{x}, \bar{y})$:*

(i) *The formula φ is independent.*

(ii) *There exist a tuple $\bar{a}$ an, indiscernible sequence $(\bar{b}_n : n < \omega)$ and $0 \leq r < s \leq 1$ such that:*

$$\varphi(\bar{a}, \bar{b}_{2n}) \leq r, \qquad \varphi(\bar{a}, \bar{b}_{2n+1}) \geq s.$$

(iii) *There exist a tuple $\bar{a}$ and indiscernible sequence $(\bar{b}_n : n < \omega)$ such that* $\lim \varphi(\bar{a}, \bar{b}_n)$ *does not exists.*

Proof. (i) $\Longrightarrow$ (ii). Assume φ is independent, and let us work in a sufficiently saturated model. Then there are $0 \leq r < s \leq 1$ such that for all m there are $(\bar{b}_n : n < m)$ and $(\bar{a}_w : w \subseteq m)$ satisfying:

$$\varphi(\bar{a}_w, \bar{b}_n) \leq r \Longleftrightarrow n \in w, \qquad \varphi(\bar{a}_w, \bar{b}_n) \geq s \Longleftrightarrow n \notin w.$$

By compactness there exists an infinite sequence $(b_n : n < \omega)$ such that for every finite $u \subseteq \omega$ and every $w \subseteq u$ there are $\bar{a}_{u,w}$ such that for all $n \in u$:

$$\varphi(\bar{a}_{u,w}, \bar{b}_n) \leq r \Longleftrightarrow n \in w, \qquad \varphi(\bar{a}_{u,w}, \bar{b}_n) \geq s \Longleftrightarrow n \notin w.$$

By standard arguments using Ramsey's Theorem there exists an indiscernible sequence $(\bar{b}_n : n < \omega)$ having the same property. In particular for every m there exists $\bar{a}_m$ such that for all $n < m$:

$$\varphi(\bar{a}, \bar{b}_{2n}) \leq r, \qquad \varphi(\bar{a}, \bar{b}_{2n+1}) \geq s.$$

The existence of $\bar{a}$ as desired now follows by compactness.

(ii) $\Longrightarrow$ (iii). Immediate.

(iii) $\Longrightarrow$ (i). Assume that $(\bar{b}_n : n < \omega)$ is indiscernible and $\lim_n \varphi(\bar{a}, \bar{b}_n)$ does not exist. Then there are $0 \leq r < s \leq 1$ such that $\varphi(\bar{a}, \bar{b}_n) < r$ and $\varphi(\bar{a}, \bar{b}_n) > s$ infinitely often. Then for every m and every $w \subseteq m$ we can find $n_0 < \ldots < n_{m-1} < \omega$ such that $\varphi(\bar{a}, \bar{b}_{n_i}) < r$ if $i \in w$ and $\varphi(\bar{a}, \bar{b}_{n_i}) > s$ otherwise. By indiscernibility we can then find $\bar{a}_w$ such that $\varphi(\bar{a}, \bar{b}_i) < r$ if $i \in w$ and $\varphi(\bar{a}, \bar{b}_i) > s$ if $i \in m \smallsetminus w$. Then φ is independent. $\blacksquare_{5.4}$

Lemma 5.5. *Let $\bar{a}$ be a tuple, $(\bar{b}_n : n < \omega)$ an indiscernible sequence of tuples, and let $\varphi_s(\bar{x}, \bar{y}_0 \ldots \bar{y}_{k_s-1})$ be dependent formulae for $s \in S$. Then there exists in an elementary extension of $\mathcal{M}$ an $\bar{a}$-indiscernible sequence $(\bar{c}_n : n < \omega)$ such that for all $s \in S$:*

$$\varphi_s(\bar{a}, \bar{c}_0 \ldots \bar{c}_{k_s-1}) = \lim \varphi_s(\bar{a}, \bar{b}_n \ldots \bar{b}_{n+k_s-1}).$$

Proof. For $k < \omega$ let I_k consist of all increasing tuples $\bar{n} \in \omega^k$. We define a partial ordering on I_k saying that $\bar{n} < \bar{n}'$ if $n_{k-1} < n'_0$. Then standard arguments using Ramsey's Theorem and compactness yield an $\bar{a}$-indiscernible sequence $(\bar{c}_n : n < \omega)$ such that for every k and every formula $\varphi(\bar{x}, \bar{y}_0, \ldots, \bar{y}_{k-1})$:

$$\varliminf_{\bar{n} \in I_k} \varphi(\bar{a}, \bar{b}_{\bar{n}}) \leq \varphi(\bar{a}, \bar{c}_0, \ldots, \bar{c}_{k-1}) \leq \varlimsup_{\bar{n} \in I_k} \varphi(\bar{a}, \bar{b}_{\bar{n}}),$$

where $\bar{b}_{\bar{n}} = \bar{b}_{n_0}, \ldots, \bar{b}_{n_{k-1}}$.

Let us now fix $s \in S$. If $(\bar{n}_m : m < \omega)$ is an increasing sequence in I_{k_s} then $(\bar{b}_{\bar{n}_m} : m < \omega)$ is an indiscernible sequence so $\lim_m \varphi_s(\bar{a}, \bar{b}_{\bar{n}_m})$ exists. Moreover, given two increasing sequences in I_{k_s} we can choose a third increasing sequence alternating between the two, so the limit does not depend on the choice of sequence. It follows that $\lim_{\bar{n}\in I_{k_s}} \varphi_s(\bar{a}, \bar{b}_{\bar{n}})$ exists, and the assertion follows. $\blacksquare_{5.5}$

Theorem 5.6. *Assume T is independent. Then there exists a formula $\varphi(x, \bar{y})$, where x is a singleton, which is independent.*

Proof. Let $\varphi(\bar{x}, \bar{y})$ be an independent formula such that $\bar{x}$ has minimal length. If it is of length one we are done. If not, we may write $\bar{x} = x\bar{z}$ and $\varphi = \varphi(x\bar{z}, \bar{y})$.

By Lemma 5.4 there are $a\bar{b}$ and a sequence $(\bar{c}_n : n < \omega)$ is a model of T as well as $0 \leq r < s \leq 1$ such that

$$\varphi(a\bar{b}, \bar{c}_{2n}) \leq r, \qquad \varphi(a\bar{b}, \bar{c}_{2n+1}) \geq s.$$

Choosing $t \in (r, s)$ dyadic and replace φ with $m(\varphi \dot{-} t)$ for m large enough we may assume that $r = 0$ and $s = 1$. For $m < \omega$ let:

$$\psi_m(\bar{z}, \bar{y}_0, \ldots, \bar{y}_{2m-1}) = \inf_x \bigvee_{i<m} (\varphi(x\bar{z}, \bar{y}_{2i}) \vee \neg\varphi(x\bar{z}, \bar{y}_{2i+1})).$$

By assumption of minimality of $\bar{x}$ the formulae ψ_n must be dependent. By Lemma 5.5 there is a $\bar{b}$-indiscernible sequence $(\bar{c}'_n : n < \omega)$ such that for all m:

$$\psi_m(\bar{b}, \bar{c}'_0 \ldots \bar{c}_{2m-1}) = \lim \psi_m(\bar{b}, \bar{c}_n \ldots \bar{c}_{n+2m-1}). \tag{2}$$

We know that $\psi_m(\bar{b}, \bar{c}_{2n}, \ldots, \bar{c}_{2n+2m-1}) = 0$, as this is witnessed by a. Therefore the limit in (2) must be equal to zero, and thus $\psi_m(\bar{b}, \bar{c}'_0 \ldots \bar{c}_{2m-1}) = 0$ for all m. By a compactness argument there exists a' such that

$$\varphi(a'\bar{b}, \bar{c}'_{2n}) = 0, \qquad \varphi(a'\bar{b}, \bar{c}'_{2n+1}) = 1.$$

Changing our point of view a little we observe that $(\bar{b}\bar{c}'_n : n < \omega)$ is an indiscernible sequence and

$$\varphi(a', \bar{b}\bar{c}'_{2n}) = 0, \qquad \varphi(a', \bar{b}\bar{c}'_{2n+1}) = 1.$$

Thus $\varphi(x, \bar{z}\bar{y})$ is independent and $\bar{x}$ was not minimal after all. $\blacksquare_{5.6}$

References

[AS06] Guillaume Aubrun and Stanisław J. Szarek, *Tensor product of convex sets and the volume of separable states on N qudits*, Physical Review A **73** (2006), 022109.

[Ben03] Itaï Ben Yaacov, *Positive model theory and compact abstract theories*, Journal of Mathematical Logic **3** (2003), no. 1, 85–118.

[BK] Itaï Ben Yaacov and H. Jerome Keisler, *Randomizations of models as metric structures*, in preparation.

[BU] Itaï Ben Yaacov and Alexander Usvyatsov, *Continuous first order logic and local stability*, Transactions of the AMS, to appear.

[Hen76] C. Ward Henson, *Nonstandard hulls of Banach spaces*, Israel Journal of Mathematics **25** (1976), 108–144.

[HPP] Ehud Hrushovski, Kobi Peterzil, and Anand Pillay, *Groups, measure and the nip*, Journal of the AMS, to appear.

[Kei99] H. Jerome Keisler, *Randomizing a model*, Advances in Mathematics **143** (1999), no. 1, 124–158.

[KM81] Jean-Louis Krivine and Bernard Maurey, *Espaces de Banach stables*, Israel Journal of Mathematics **39** (1981), no. 4, 273–295.

[Las92] Michael C. Laskowski, *Vapnik-Chervonenkis classes of definable sets*, Journal of the London Mathematical Society. Second Series **45** (1992), no. 2, 377–384.

[Poi85] Bruno Poizat, *Cours de théorie des modèles*, Nur al-Mantiq wal-Ma'rifah, 1985.

[She71] Saharon Shelah, *Stability, the f.c.p., and superstability; model theoretic properties of formulas in first order theory*, Annals of Mathematical Logic **3** (1971), no. 3, 271–362.

[VC71] V. N. Vapnik and A. Ya. Chervonenkis, *On the uniform convergence of relative frequencies of events to their probabilities*, Theory of Probability and Applications **16** (1971), no. 2, 264–280.

[vdD98] Lou van den Dries, *Tame topology and o-minimal structures*, London Mathematical Society Lecture Note Series, vol. 248, Cambridge University Press, Cambridge, 1998.

Itaï Ben Yaacov, Université de Lyon, Université Lyon 1, Institut Camille Jordan, UMR 5208 CNRS, 43 boulevard du 11 novembre 1918, F-69622 Villeurbanne Cedex, France

URL: `http://math.univ-lyon1.fr/~begnac/`