

## THE GAP BETWEEN PROBABILITY AND PREVALENCE: LONELINESS IN VECTOR SPACES

MAXWELL B. STINCHCOMBE

(Communicated by Claudia Neuhauser)

**ABSTRACT.** The best available definition of a subset of an infinite dimensional, complete, metric vector space  $V$  being “small” is Christensen’s Haar zero sets, equivalently, Hunt, Sauer, and Yorke’s shy sets. The complement of a shy set is a prevalent set. There is a gap between prevalence and likelihood. For any probability  $\mu$  on  $V$ , there is a shy set  $C$  with  $\mu(C) = 1$ . Further, when  $V$  is locally convex, any i.i.d. sequence with law  $\mu$  repeatedly visits neighborhoods of only a shy set of points if the neighborhoods shrink to 0 at any rate.

### 1. SHY SETS AND PREVALENT SETS

Specialized to a real topological vector space  $V$  with a complete, separable, metrisable topology, Christensen’s [2] Haar zero sets are a subclass of the universally measurable sets. A universally measurable  $S$  is a **Haar zero** set if there is a measure  $\mu$ , such that  $\mu(S + v) = 0$  for every  $v \in V$ .<sup>1</sup>

Hunt, Sauer, and Yorke (HSY) [4] more thoroughly investigate properties and examples of shy sets. A set  $S$  is **shy** if it is a subset of a Borel set  $S'$  for which  $\mu(S' + v) = 0$  for every  $v \in V$  and some  $\mu$  such that  $\mu(K) = \mu(V)$  for some compact  $K$ . Expanding HSY’s definition of shyness to subsets of universally measurable sets causes no change to any of their results, and expanding Christensen’s definition to include subsets of Haar zero sets causes no change to his results.

Because a compactly supported measure is a measure,  $S$  being shy implies that  $S$  is a Haar zero set. The tightness of measures on  $V$  implies that  $S$  being a Haar zero set implies that  $S$  is shy. A measure  $\mu$  is **tight** if for every  $\epsilon > 0$  there is a compact set  $C_\epsilon$  such that  $\mu(C_\epsilon) > \mu(V) - \epsilon$ . Because  $V$  is a complete, separable metric space, every measure on  $V$  is tight (e.g. Dudley [3, Theorem 11.5.4]). Suppose that  $\mu$  is a measure such that  $\mu(S + v) = 0$  for all  $v$ . Define  $\mu_\epsilon$  by  $\mu_\epsilon(A) = \mu(A \cap C_\epsilon)$ . For  $0 < \epsilon < \mu(V)$ ,  $\mu_\epsilon$  is a compactly supported measure satisfying  $\mu_\epsilon(S + v) = 0$  for all  $v \in V$ , so that every Haar zero set is shy.

---

Received by the editors March 1, 1999 and, in revised form, April 19, 1999.

1991 *Mathematics Subject Classification.* Primary 28C20, 60B11; Secondary 90B40.

<sup>1</sup>A measure on  $V$  is a countably additive, finite, positive, and non-zero set function on  $\mathcal{V}$ , the Borel  $\sigma$ -field on  $V$ .  $\mathcal{V}^\mu$  denotes the completion of  $\mathcal{V}$  with respect to the measure  $\mu$ , and  $\mathcal{V}^u = \bigcap_\mu \mathcal{V}^\mu$  is the  $\sigma$ -field of universally measurable sets. Every measure  $\mu$  on  $\mathcal{V}$  has a unique extension to  $\mathcal{V}^u$ . Every measure  $\mu_u$  on  $\mathcal{V}^u$  has a unique restriction,  $\mu$ , to  $\mathcal{V}$ , and  $\mu_u$  is the extension of  $\mu$ .

The complement of a shy set is called a **prevalent** set. Important properties of prevalent and shy sets include:

1. shy sets have no interior so that prevalent sets are dense,
2. the countable union of shy sets is shy, equivalently, the countable intersection of prevalent sets is prevalent,
3. if  $V$  is locally compact, that is, finite dimensional, then a set is shy if and only if it has Lebesgue measure 0, and
4. if  $V$  is not locally compact, i.e. is infinite dimensional, then compact sets are shy.

Suppose that  $S$  is shy and that  $\mu(S + v) \equiv 0$ . Setting  $Q(\cdot) = \mu(\cdot)/\mu(V)$  gives  $Q(S + v) \equiv 0$  and shows that there is no loss in assuming that  $\mu$  is a probability in the definition of shy sets. If  $\mu$  has any atoms, then  $\mu(S + v) \equiv 0$  only if  $S = \emptyset$ , so that the  $\mu$  that matter for shyness are non-atomic. HSY show that the uniform distribution on subsets of finite dimensional subspaces of  $V$  are especially useful non-atomic measures, and this observation will be crucial here.

Unless explicitly noted,  $V$  is infinite dimensional with a complete, translation invariant metric  $d(\cdot, \cdot)$ , the balls around 0 are both convex and balanced,<sup>2</sup> and all measures are probabilities.

## 2. THE GAP

The gap between between likelihood statements and prevalence statements for properties in  $V$  arises from the combination of the tightness of probabilities and the failure of local compactness.

**2.1. The gap for Bayesians.** As Christensen [2, p. 119] notes, there is no probability  $\mu$  such that  $S$  being shy is equivalent to  $\mu(S) = 0$ . Slightly more can be said — any probability  $\mu$  must assign mass 0 to a prevalent set of points.

The **support** of  $\mu$ ,  $\text{supp}(\mu)$ , is defined as the intersection of the closed sets  $F$  such that  $\mu(F) = 1$ . The support of  $\mu$  being large, say all of  $V$  or all of the unit ball in  $V$ , is an indication that  $\mu$  covers many points. However, because  $\mu$  is tight, for every  $\epsilon > 0$  there is a compact  $C_\epsilon$  such that  $\mu(C_\epsilon) > 1 - \epsilon$ . Setting  $C = \bigcup_\epsilon C_\epsilon$  (the union being taken over rational, positive  $\epsilon$ ) gives  $C \subset \text{supp}(\mu) = \text{cl} C$  with  $\mu(C) = 1$ . Being the countable union of compact (hence shy) sets means that  $C$  is shy. Thus,  $V \setminus C$  is prevalent and  $\mu(V \setminus C) = 0$ . If beliefs are modeled as probabilities, even full support beliefs about  $V$  do not cover very much of  $V$ .

**2.2. Intermediate sets.** There are many  $C'$  that are shy and satisfy  $C \subset C' \subset \text{cl} C$  and  $\mu(C) = 1$  where  $C$  is a countable union of compact subsets of  $V$ . A set  $F$  is **approximately flat** if for every  $\epsilon > 0$ , there is a finite dimensional subspace  $W$  of  $V$  such that  $F \subset W + B(0, \epsilon)$  where  $B(v, r)$  is the ball around  $v$  with radius  $r$ . Any finite union of approximately flat sets is approximately flat, and every compact set is approximately flat — let  $W$  be the span of a finite  $\epsilon$ -net.

**Lemma 1.** *For any sequence  $F_n$  of approximately flat sets and any  $r_n \rightarrow 0$ , the set  $C' = \bigcap_m \bigcup_{n \geq m} C_n$  is shy where  $C_n = F_n + B(0, r_n)$ .*

---

<sup>2</sup>A set  $A$  in  $V$  is balanced if for all  $x \in A$ ,  $\lambda x \in A$  whenever  $|\lambda| \leq 1$ . Local convexity of  $V$  implies the existence of a neighborhood basis of balanced, convex sets.

The set  $C'$  in the previous lemma consists of points that are within  $r_n$  of  $F_n$  for infinitely many  $n$ . If  $\mu(F_n) \uparrow 1$ , as must happen for some sequence  $F_n$  of compact (hence approximately flat) sets, then  $\mu(\bigcup_n F_n) = 1$  and  $\bigcup_n F_n \subset C' \subset \text{cl } \bigcup_n F_n$ , and the inclusions may be strict. Taking  $F_n \equiv F$  shows that the closure of any approximately flat set is shy.

*Proof of Lemma 1.* There is no loss in assuming that the sequence  $F_n$  is nested. To show that  $C'$  is shy, it is sufficient to prove the existence of a probability  $Q$  such that  $Q(C' + v) \equiv 0$ . By the easy Borel-Cantelli Lemma,<sup>3</sup> it is sufficient that for all  $v \in V$ ,  $\sum_n Q(C_n + v) < \infty$ .

The construction divides  $\mathbb{N}$  into sets  $M_k$ , where  $M_k = \{n_{k-1} < n \leq n_k\}$ ,  $k = 1, 2, \dots$ . For each  $k$ , there is a probability  $Q_k$  supported on a finite dimensional, compact set of radius  $2\bar{r}_k$  such that  $\sum_k \bar{r}_k < \infty$ . Further, for all  $v \in V$ ,  $p_k = \sum_{n \in M_k} Q_k(C_n + v) < 1/2^k$ . The probability  $Q$  will be defined as the convolution  $Q_1 * Q_2 * \dots$ , and  $\sum_n Q(C_n + v)$  will be shown to be less than or equal to  $\sum_k p_k$ , which is finite.

Set  $n_0 = 0$ . This paragraph gives the inductive construction of  $Q_k$  such that  $\sum_{n \in M_k} Q_k(C_n + v) < 1/2^k$ . Given  $n_{k-1}$ , let  $\bar{r}_k = \max\{r_n : n > n_{k-1}\}$  (which exists because  $r_n \rightarrow 0$ ). Pick  $n_k$  such that for all  $n \geq n_k$ ,  $r_n < \bar{r}_k/2$ . Because the  $F_n$  are nested and  $r_n \leq \bar{r}_k$  for each  $n \in M_k$ ,  $C_n = F_n + B(0, r_n) \subset F_{n_k} + B(0, \bar{r}_k)$  for each  $n \in M_k$ . Pick  $0 < \epsilon < \bar{r}_k/2$  so that  $(\bar{r}_k + \epsilon)/2\bar{r}_k < 3/4$ . Because  $F_{n_k}$  is approximately flat, there is a finite dimensional subspace,  $W_k$ , of  $V$  such that  $F_{n_k} \subset W_k + B(0, \epsilon)$ . Pick an integer  $N_k$  such that  $(3/4)^{N_k} < 1/2^k(n_k - n_{k-1})$ . Pick  $N_k$  linearly independent points  $e_i, i = 1, \dots, N_k$ , such that  $d(e_i, 0) = 1$  (where  $d(\cdot, \cdot)$  is the complete, translation invariant metric on  $V$ ) and  $d(e_i, W_k) = 1$ . Parametrize the subspace  $V_k$  spanned by the  $e_i$  by  $a \in \mathbb{R}^{N_k}$ ,  $a \leftrightarrow \sum_i a_i e_i$ . Let  $A_k$  be the set of  $a$  in  $\mathbb{R}^{N_k}$  such that  $d(\sum_i a_i e_i, 0) \leq 2\bar{r}_k$ .  $A_k$  is a convex, balanced set with non-empty interior. Take  $Q'_k$  as the uniform distribution on  $A_k$  so that  $Q'_k(\{a : d(\sum_i a_i e_i, 0) \leq \bar{r}_k + \epsilon\}) < (3/4)^{N_k}$ . Let  $Q_k$  be the corresponding image law on  $V_k$ . For all  $v \in V$ ,  $Q_k(W_k + B(0, \bar{r}_k + \epsilon) + v) \leq (3/4)^{N_k} < 1/2^k(n_k - n_{k-1})$ . Further,  $C_n + v \subset W_k + B(0, \bar{r}_k + \epsilon) + v$ . Combining,  $Q_k(C_n + v) < 1/2^k(n_k - n_{k-1})$  for each  $n \in M_k$ .

Because  $\bar{r}_{k+1} < \bar{r}_k/2$ ,  $\sum_k 2\bar{r}_k < \infty$  so that the convolution  $Q = Q_1 * Q_2 * \dots$  is well-defined. For each  $k$ , let  $P_k$  be the convolution of the  $Q_j, j \neq k$ , so that  $Q = Q_k * P_k$ . Pick any  $n \in M_k$  and any  $v \in V$ .  $Q(C_n + v) = \int_V Q_k(C_n + v - s) dP_k(s)$  so that  $Q(C_n + v) < 1/2^k(n_k - n_{k-1})$ . This implies that  $\sum_n Q(C_n + v) < \sum_k 1/2^k = 1$  so that  $Q(C' + v) = 0$ , as was to be shown.  $\square$

**2.3. The gap for approximate frequentists.** Frequentist interpretations of  $\mu$  also do not cover much of  $V$ , at least not at any rate of approximation.

Let  $Y_i, i = 1, 2, \dots$ , be a sequence of i.i.d. random variables defined on a probability space  $(\Omega, \mathcal{F}, P)$  and having distribution  $\mu$  on  $V$ . It is (well-)known that for  $P$ -almost every  $\omega$ , the set  $Y_1(\omega), Y_2(\omega), \dots$  is dense in the support of  $\mu$ , so that repeated draws from  $\mu$  may seem to cover a large part of  $V$ . However, there is a shy set  $C', C \subset C' \subset \text{supp}(\mu)$ , consisting of all points whose neighborhoods are visited infinitely often when the neighborhoods shrink. Conversely, there is a prevalent

<sup>3</sup>The easy Borel-Cantelli Lemma says that, for a measure  $P$ , if  $\sum_n P(E_n) < \infty$ , then  $P(E) = 0$  where  $E = \bigcap_m \bigcup_{n \geq m} E_n$ . The proof is  $P(E) \leq P(\bigcup_{n \geq m} E_n) \leq \sum_{n \geq m} P(E_n) \downarrow 0$  so that  $P(E) = 0$ .

set of points in  $V$  whose neighborhoods are almost never visited by repeated draws from the distribution  $\mu$  provided the neighborhoods shrink to 0.

For  $v \in V$  and  $r_n \rightarrow 0$ , an  $r_n$  **funnel at**  $v$  is the sequence of balls,  $B(v, r_n)$ , centered at  $v$  with radius  $r_n$ . Let  $\lambda_n \uparrow \infty$ . For  $v \in V$ , let  $A(v)$  be the event that  $Y_1, \dots, Y_{\lambda_n}$  visits the  $r_n$  funnel at  $v$  for infinitely many  $n$ . More explicitly,  $A(v)$  is the set of  $\omega$  such that  $Y_i(\omega) \in B(v, r_n)$  for some  $i \leq \lambda_n$  for infinitely many  $n$ . A point  $v \in V$  is  $(r_n, \lambda_n)$ -**lonely** if  $P(A(v)) = 0$ . Equivalently,  $v$  is  $(r_n, \lambda_n)$ -lonely if, with probability 1, the  $r_n$ -ball around  $v$  eventually receives no more visits from  $Y_1, \dots, Y_{\lambda_n}$ .

Loneliness is prevalent.

**Theorem.** *For any probability  $\mu$ , any  $r_n \rightarrow 0$ , and any  $\lambda_n \uparrow \infty$ , the set of  $(r_n, \lambda_n)$ -lonely points is prevalent.*

*Proof.* Pick a sequence  $e_n \downarrow 0$  such that  $\sum_n e_n < \infty$ . Because  $\mu$  is tight, there is a nested sequence of compact (hence approximately flat) subsets,  $K_n \subset K_{n+1}$ , of  $V$  and a sequence  $d_n \downarrow 0$  such that  $\mu(K_n) > (1 - d_n)$  and  $(1 - (1 - d_n)^{\lambda_n}) \leq e_n$ . Set  $B_n = \{\omega : (\exists i \leq \lambda_n)[Y_i(\omega) \notin K_n]\}$  so that  $P(B_n) \leq e_n$ .

For  $v \in V$ , let  $A_n(v)$  be the set  $\{\omega : (\exists i \leq \lambda_n)[Y_i(\omega) \in B(v, r_n)]\}$ . The set of  $\omega$  that are in  $A_n(v)$  for infinitely many  $n$  is  $A(v) = \bigcap_m \bigcup_{n \geq m} A_n(v)$ . If  $d(v, K_n) \geq r_n$ , then  $P(A_n(v)) \leq P(B_n) \leq e_n$ . By the Borel-Cantelli Lemma (again),  $P(A(v)) = 0$  if  $\sum_n P(A_n(v)) < \infty$ . To show that  $v$  is lonely, it is sufficient to show that  $\sum_n P(A_n(v)) < \infty$ . This means that  $v$  is lonely if  $d(v, K_n) \geq r_n$  for all but finitely many  $n$  — if there exists an  $m$  such that for all  $n \geq m$ ,  $d(v, K_n) \geq r_n$ , then  $\sum_n P(A_n(v)) \leq (m - 1) + \sum_{n \geq m} e_n < \infty$ .

Let  $C_n = K_n + B(0, r_n)$  so that  $C' = \bigcap_m \bigcup_{n \geq m} C_n$  is the complement of the set of  $v$  satisfying  $d(v, K_n) \geq r_n$  for all but finitely many  $n$ . It is sufficient to show that  $C'$  is shy, and this is the conclusion of Lemma 1.  $\square$

### 3. COMMENTS

The comments address several issues: why, in retrospect at least, the Theorem is not surprising; the connections to naive stochastic search; which easy generalizations are available and which are not; how translation invariance relates to the gap between prevalence and probability; and why a measures-on-measures approach does not lead to shyness.

**3.1. In retrospect.** One intuition comes from how small approximately flat sets are. If  $W^d$  is a  $d$ -dimensional subspace of  $\mathbb{R}^k$ , then, as a proportion of the unit ball,  $W^d + B(0, \epsilon)$  is on the order of  $\epsilon^{k-d}$ . This leads one to suspect that approximately flat sets are “small” in infinite dimensional  $V$ 's. The Theorem builds on the fact that for any probability  $\mu$  and any  $\epsilon > 0$ , there is an approximately flat set  $F_\epsilon$  such that  $\mu(F_\epsilon) > 1 - \epsilon$ .

A second intuition comes from the observation that every  $\delta$ -ball in  $V$  contains infinitely many disjoint  $\delta/4$ -balls. Thus, the “proportion” of any given  $B(v, \delta)$  that has any significant proportion of the mass in  $B(v, \delta)$  is vanishingly small. Millar [6] used this intuition in a study of naive stochastic search.

**3.2. Naive stochastic search.** Let  $\Theta$  denote the closed unit ball in  $V$  where  $V$  is a Banach space with norm  $\|\cdot\|$ . Millar [6] examines a number of problems from

mathematical statistics, e.g. non-parametric goodness of fit statistics, minimum distance estimators and maximum likelihood estimates, that involve the computation of quantities of the form  $\inf_{\theta \in \Theta} \widehat{D}_n(\theta)$  where  $n$  is the number of available observations. For many  $\widehat{D}_n(\cdot)$ , finding the infimum may not be feasible, and stochastic search methods may be tempting.

After taking  $\lambda_n$  independent draws from some probability  $\mu$  with  $\text{supp}(\mu) = \Theta$ , one can approximate the complicated infimum problem by the finite problem  $\min_{i \leq \lambda_n} \widehat{D}_n(Y_i)$ . With probability 1, the sequence  $Y_1(\omega), Y_2(\omega), \dots$  comes within any  $\epsilon > 0$  of any  $\theta_0$ . Provided  $\widehat{D}_n(\cdot)$  is continuous and  $\lambda_n \uparrow \infty$  sufficiently quickly, this seems promising.

Unfortunately, it is typical that  $\widehat{D}_n$  changes with  $n$  so as to confound the naive stochastic search just described. First, if for all constants  $c$  there exists  $N$  such that for all  $n \geq N$ ,  $\theta_n \notin B(\theta_0, cn^{-\gamma})$  (where  $\gamma$  is in  $(0, 1/2]$ , usually  $\gamma = 1/2$ ), then typically  $\widehat{D}_n(\theta_n) \rightarrow \infty$ . Second, if there exists a  $c$  and an  $N$  such that  $\theta_n \in B(\theta_0, cn^{-\gamma})$  for all  $n \geq N$ , then typically  $\widehat{D}_n(\theta_n)$  is bounded.

The direct implication of the Theorem presented here and the two properties of  $\widehat{D}_n$  just described is that for all  $\theta_0$  in a prevalent set, with probability 1,  $\min_{i \leq \lambda_n} \widehat{D}_n(Y_i) \rightarrow \infty$  even when  $\inf_{\theta \in \Theta} \widehat{D}_n(\theta)$  is bounded. Millar [6] showed that there is a countably infinite set,  $\{\theta_{0,j}\} \subset \Theta$ , having this and a stronger property — on a set of  $\omega$  having probability 1,

$$\liminf_n \inf_j \min_{i \leq \lambda_n} \|Y_i - \theta_{0,j}\|/r_n \geq 1.$$

With probability 1,  $Y_i$ ,  $i = 1, \dots, \lambda_n$  not only keeps at distance  $r_n$  (or further) away from each  $\theta_{0,j}$ , it simultaneously stays  $r_n$  (or further) away from all of them.

The denseness of prevalent sets implies that the countably infinite set  $\{\theta_{0,j}\}$  in Millar’s result cannot be replaced by a prevalent set. In a more positive vein, he also demonstrates that there are a variety of more sophisticated, non-i.i.d. stochastic search schemes with better behavior.

**3.3. Generalizations.** Suppose that  $S_n$  is a sequence of sets in a class  $\mathcal{S}$  that is closed under finite union and that  $r_n \rightarrow 0$ . It is possible that  $\bigcup_n S_n$  is a strict subset of  $\bigcap_m \bigcup_{n \geq m} S_n + B(0, r_n)$ , which is in turn a strict subset of  $\text{cl} \bigcup_n S_n$ . Lemma 1 showed that if  $\mathcal{S}$  is the set of approximately flat sets, then  $\bigcap_m \bigcup_{n \geq m} S_n + B(0, r_n)$  is shy. It would be nice to know how large  $\mathcal{S}$  can be and still have  $\bigcap_m \bigcup_{n \geq m} S_n + B(0, r_n)$  shy. The answer cannot be all shy sets — the countable sets, including the countable dense ones, are shy.

As discussed above, there is no countably additive probability,  $\mu$ , assigning 0 to shy sets and 1 to prevalent sets. It is easy to show that there are finitely additive  $\mu$  that do this.<sup>4</sup> The class  $\mathcal{H}$  consisting of shy sets and their complements is a  $\sigma$ -field, define  $\mu'(H) = 0$  if  $H$  is shy and  $\mu'(H) = 1$  if the complement of  $H$  is shy.<sup>5</sup> By the Hahn-Banach Theorem,  $\mu'$  has an extension,  $\mu$ , to the Borel  $\sigma$ -field on  $V$ . Since every compact set is shy,  $\mu$  is 0 on the compacts; hence purely finitely additive. Being purely finitely additive is a drawback for both Bayesian and frequentist interpretations of  $\mu$ -likelihood. Being purely finitely additive, there exists a nested sequence  $E_n \downarrow \emptyset$  with  $E_1 = V$  and  $\mu(E_n) \equiv 1$ . Setting  $D_n =$

<sup>4</sup>This is the only paragraph in which the assumption of countable additivity is dropped.

<sup>5</sup>An analogous  $\sigma$ -field is the class of all countable subsets and their complements in an uncountable space.

$E_n \setminus E_{n+1}$  expresses  $V$  as a countable union of disjoint sets each of which has  $\mu$ -probability 0. In this sense,  $\mu$  is not supported anywhere in  $V$ . Another version of  $\mu$  not being supported anywhere in  $V$  is the observation is that if  $V$  is imbedded in its compact Stone space,  $\widehat{V}$ , the extension of  $\mu$ ,  $\widehat{\mu}$ , satisfies  $\widehat{\mu}(V) = 0$  (Kingman [5] discusses the implications of this observation when  $V$  is a set of possible sample paths and  $\mu$  is a stochastic process).

**3.4. Translation invariance and the gap.** Another way to understand the gap is that the requirement of translation invariance reverses the intuitive link between characteristics of a probability measure and the assignment of zero mass indicating a ‘small’ set. If a probability  $\mu_1$  assigns positive mass to many sets, e.g. a non-degenerate Gaussian distribution in  $\mathbb{R}^k$ , then  $\mu_1(S) = 0$  is a strong statement that  $S$  is ‘small’. On the other hand, if  $\mu_2$  assigns 0 mass to most sets, e.g. if  $\mu_2(\{x\}) = 1$  for some  $x$ , then  $\mu_2(S) = 0$  has very few implications for the ‘size’ of  $S$ . When translation invariance is added to this we reverse the strength of the conclusions:  $\mu_2(S + v) \equiv 0$  if and only if  $S = \emptyset$ , an extremely strong conclusion about the smallness of  $S$ ;  $\mu_1(S + v) \equiv 0$  if and only if  $\mu_1(S) = 0$  if and only if  $S$  is shy in  $\mathbb{R}^k$ .

Translation invariance also means that only a shy set of measures matter in the definition of shyness. A set  $S$  is  $\mu$ -shy if it is a subset of a universally measurable set  $S'$  satisfying  $\mu(S' + v) \equiv 0$ . The set of  $\mu$ -shy sets is  $\mathcal{S}(\mu)$ , the class of shy sets is  $\mathcal{S} = \bigcup_{\mu} \mathcal{S}(\mu)$ . If  $\mu$  has any atoms, then  $\mathcal{S}(\mu) = \{\emptyset\}$ . This means that the  $\mu$  that matter in defining shy sets are the ones without atoms. The probabilities without atoms, the ones that matter, are, relative to the set of probabilities on  $V$ , a shy set.

Let  $M(V)$  be the Borel measures on  $V$  with the (metrisable) weak\* topology. The set of probabilities,  $\Delta(V)$ , is complete in the relative weak\* topology and is itself a shy subset of the completion of  $M(V)$ . Anderson and Zame [1] address the rather subtle problems in extending the definitions of shy sets to shy subsets of sets that are themselves shy. Fix a metrisable, topological vector space  $\mathbb{X}$ . The following is their simple sufficient condition for shyness relative to a convex set  $C \subset \mathbb{X}$ , with  $C$  being a Borel subset of  $\mathbb{X}$  that is topologically complete in the relative topology.  $W$  denotes a finite dimensional subspace of  $\mathbb{X}$  and  $\lambda_W$  denotes Lebesgue measure<sup>6</sup> on  $W$ .

**Definition.** A universally measurable subset  $E \subset C$  is **finitely shy in  $C$**  if there is a finite-dimensional subspace  $W \subset \mathbb{X}$  such that  $\lambda_W(C + a) > 0$  for some  $a \in \mathbb{X}$  and  $\lambda_W(E + x) = 0$  for every  $x \in \mathbb{X}$ . A (not necessarily universally measurable) subset  $F \subset \mathbb{X}$  is finitely shy in  $C$  if it is contained in a finitely shy universally measurable set. A subset  $Y \subset C$  is **finitely prevalent in  $C$**  if  $C \setminus Y$  is finitely shy in  $C$ .

**Lemma 2.** *The set of probabilities on  $V$  without atoms is finitely shy in  $\Delta(V)$ .*

*Proof of Lemma 2.* Let  $\delta_1$  and  $\delta_2$  be point mass on two distinct points,  $v_1$  and  $v_2$ , in  $V$ . Take  $W$  to be the span of the point  $(\delta_1 - \delta_2)$  in  $M(V)$ . If we take  $a = \delta_2$ , then for all  $r \in [0, 1]$ ,  $r(\delta_1 - \delta_2) + a = r\delta_1 + (1 - r)\delta_2 \in \Delta(V)$  so that  $\lambda_W(\Delta(V) - a) > 0$ . Let  $E \subset \Delta(V)$  denote the set of probabilities on  $V$  without atoms. For every  $x \in M(V)$ , the set of  $r$  such that  $r(\delta_1 - \delta_2) + x$  has no atoms contains at most one point so that  $\lambda_W(E + x) = 0$ .  $\square$

<sup>6</sup>Since all that matters for the definition are the null sets, any linear parametrization of  $W$  can be used to define  $\lambda_W$ .

**3.5. Measures on measures.** Only a shy set of measures are used in defining shy sets. Because  $\Delta(V)$  is a complete, separable metric space, any probability on probabilities,  $P \in \Delta(\Delta(V))$ , is carried on a shy set of probabilities. One might imagine that there is a  $P$  carried on the set of  $\mu$  that matter with the property that  $S$  is shy if and only if  $P(\{\mu : \mu(S) = 0\}) > 0$ . Here  $P$  would play the role of translating the  $\mu$ . This will not work.

**Lemma 3.** *For every  $P \in \Delta(\Delta(V))$ , there is a shy set  $S$  such that*

$$P(\{\mu : \mu(S) = 1\}) = 1.$$

*Proof of Lemma 3.* As  $\Delta(V)$  is a complete metric space, for every  $\delta > 0$ , there is a compact set  $K_\delta$  such that  $P(K_\delta) > 1 - \delta$ . For every weak\* compact set,  $K$ , of measures on  $V$ , and every  $\epsilon > 0$ , there exists a compact  $C_\epsilon(K) \subset V$  such that for all  $\mu \in K$ ,  $\mu(C_\epsilon(K)) > 1 - \epsilon$ . Take  $S = \bigcup_{\epsilon, \delta} C_\epsilon(K_\delta)$  (where the union is over rational  $\epsilon, \delta > 0$ ). For  $P$ -almost every  $q \in \Delta(V)$ ,  $q(S) = 1$ , but  $S$  is shy.  $\square$

#### REFERENCES

- [1] Anderson, R. M. and W. R. Zame (1997). Genericity with Infinitely Many Parameters. Working Paper, Department of Economics, U. C. Berkeley.
- [2] Christensen, J. P. R. (1974). *Topology and Borel Structure*. Amsterdam: North-Holland Publishing Company. MR **50**:1221
- [3] Dudley, R. M. (1989). *Real Analysis and Probability*. Wadsworth & Brooks/Cole, Pacific Grove, California. MR **91g**:60001
- [4] Hunt, B. R., T. Sauer, and J. A. Yorke (1992). Prevalence: A Translation-Invariant ‘Almost Every’ on Infinite-Dimensional Spaces. *Bulletin (New Series) of the American Mathematical Society* 27, 217-238. MR **93k**:28019
- [5] Kingman, J. F. C (1967). Additive Set Functions and the Theory of Probability. *Proceedings of the Cambridge Philosophical Society* **63**, 767-775. MR **36**:3385
- [6] Millar, P. W. (1992). Stochastic Search in Banach Spaces. In *Probability in Banach Spaces*, R. M. Dudley, M. G. Hahn, and J. Kuelbs (eds.), 497-509. MR **94g**:62084

DEPARTMENT OF ECONOMICS, UNIVERSITY OF TEXAS AT AUSTIN, AUSTIN, TEXAS 78712-1173  
*E-mail address:* maxwell@eco.utexas.edu