Donsker implies VC

Wellner; 2/3/99

Proposition 1. Let $\mathcal{C}$ be a countable class of sets in $\mathcal{X}$. Suppose that there is a sequence $a_n \to 0$ such that for every P on $(\mathcal{X}, A)$, $\|\mathrm{P}_n - P\|_{\mathcal{C}}/a_n = O_p(1)$. Then $\mathcal{C}$ is a VC-class. In particular, if $\mathcal{C}$ is Donsker for every P, then $\mathcal{C}$ is VC.

Proof. We may assume that $\sup_n (a_n\sqrt{n})^{-1} < \infty$; thus $a_n \ge Kn^{-1/2}$ for some $K > 0$. By Hoffmann-Jørgensen's inequality

$$\sup_n \frac{1}{a_n} E\|\mathrm{P}_n - P\|_{\mathcal{C}} < \infty\,.$$

By the symmetrization inequalities,

$$\frac{1}{a_n} E\|\frac{1}{n}\sum_{i=1}^n \epsilon_i(1_C(X_i) - P(C))\|_{\mathcal{C}} \le \frac{2}{a_n} E\|\mathrm{P}_n - P\|_{\mathcal{C}} < \infty\,,$$

so

$$\begin{aligned}
\Phi(P) &\equiv \sup_n \frac{1}{na_n} E\|\sum_{i=1}^n \epsilon_i 1_C(X_i)\|_{\mathcal{C}} \\
&\le \frac{2}{a_n} E\|\mathrm{P}_n - P\|_{\mathcal{C}} + \sup_n \frac{1}{na_n} E|\sum_{i=1}^n \epsilon_i| \\
&\le \frac{2}{a_n} E\|\mathrm{P}_n - P\|_{\mathcal{C}} + \sup_n \frac{1}{a_n\sqrt{n}} n^{-1/2} E|\sum_{i=1}^n \epsilon_i| \\
&\le \frac{2}{a_n} E\|\mathrm{P}_n - P\|_{\mathcal{C}} + 1/K < \infty\,.
\end{aligned}$$

Thus $\Phi(P) < \infty$ for every P. We now show that there exists an $M < \infty$ such that

$$(1) \qquad \Phi(P) \le M \qquad \text{for all } P\,.$$

To this end, we first show that if P^0, P^1 are two measures on $(\mathcal{X}, \mathcal{A})$, and $P = \alpha P^0 + (1-\alpha)P^1$, then $\Phi(P) \ge \alpha\Phi(P^0)$. To see this, let X_i^0, X_i^1 respectively, be i.i.d. P^0, P^1 respectively. Let λ_i be i.i.d. Bernoulli$(1-\alpha)$

random variables independent of the X_i^0's and X_i^1's. Then $X_i =_d X_i^{\lambda_i}$, and by the contraction principle

$$E\|\sum_{i=1}^n \epsilon_i 1_C(X_i)\|_{\mathcal{C}} \geq E\|\sum_{i=1}^n \epsilon_i 1_C(X_i^0) 1_{[\lambda_i=0]}\|_{\mathcal{C}}\,.$$

By Jensen's inequality this yields

$$E\|\sum_{i=1}^n \epsilon_i 1_C(X_i)\|_{\mathcal{C}} \geq \alpha E\|\sum_{i=1}^n \epsilon_i 1_C(X_i^0)\|_{\mathcal{C}}\,,$$

and hence $\Phi(P) \geq \alpha\Phi(P^0)$. Now suppose that (1) is false. Then there exists a sequence of measures P_k on $(\mathcal{X}, \mathcal{A})$ such that $\Phi(P_k) \geq 4^k$ for every k. Then, defining P

$$P = \sum_{j=1}^\infty 2^{-j} P_j = 2^{-k} P_k + (1 - 2^{-k}) \sum_{j \neq k} 2^{-j} P_j\,,$$

we find that P has $\Phi(P) \geq 2^{-k}\Phi(P_k) \geq 2^k$ for every k, and this yields $\Phi(P) = \infty$, contradicting $\Phi(P) < \infty$ for all P. Thus (1) holds.

Now suppose that $\mathcal{C}$ is not VC. Then for every k there is a set $A = A_k = \{x_1, \ldots, x_k\} \subset \mathcal{X}$ such that $\mathcal{C}$ shatters A; i.e. $\#\{C \cap A : C \in \mathcal{C}\} = 2^k$. Then for each $\alpha \in R^k$ we have

$$\begin{aligned} \sum_{i=1}^k |\alpha_i| &= \sum \alpha_i^+ + \sum \alpha_i^- \\ &\leq 2\max\left\{\sum \alpha_i^+, \sum \alpha_i^-\right\} \\ &\leq 2\|\sum_{i=1}^k \alpha_i 1_C(x_i)\|_{\mathcal{C}}\,; \end{aligned} \tag{2}$$

note that the last inequality holds equality when C picks out the set of x_i's corresponding to those $\alpha_i's$ yielding the maximum of $\sum \alpha_i^+$ and $\sum \alpha_i^-$. Now take $P = k^{-1}\sum_{i=1}^k \delta_{x_i}$. Choose n so large that $n > 4Mna_n$; this is possible since $a_n \to 0$. Then choose $k > 2n^2$; with this choice of k it follows that the set $\Omega_0 \equiv \cap_{i\neq j}[X_i \neq X_j]$ has $P(\Omega_0) \geq 1/2$: note that

$$P(\Omega_0^c) \;=\; P(\cup_{i \neq j \leq n}[X_i = X_j])$$

$$\begin{aligned} &\leq \sum_{i \neq j \leq n} P(X_i = X_j) \\ &\leq n^2 k^{-1} < 1/2\,. \end{aligned}$$

Thus, since $\Phi(P) \leq M$, (2) yields

$$\begin{aligned} Mna_n &\geq E\|\sum_{i=1}^{n} \epsilon_i 1_C(X_i)\|_{\mathcal{C}} \\ &\geq E\left\{\|\sum_{i=1}^{n} \epsilon_i 1_C(X_i)\|_{\mathcal{C}} 1_{\Omega_0}\right\} \\ &\geq \frac{n}{2} P(\Omega_0) \geq \frac{n}{4}\,. \end{aligned}$$

This contradicts our choice of $n > 4Mna_n$. It follows that $\mathcal{C}$ is VC. □