Stat210B: Theoretical Statistics **Lecture Date: February 13, 2007**

P-Glivenko-Cantelli

Lecturer: Michael I. Jordan *Scribe: Christopher Hundt*

1 P-Glivenko-Cantelli

Definition 1 (P-Glivenko-Cantelli)**.** A class $\mathcal{F}$ is P-Glivenko-Cantelli if

$$\sup_{f\in\mathcal{F}} |P_n f - Pf| \xrightarrow{\text{a.s.}} 0.$$

Definition 2 (envelope)**.** An envelope for a class $\mathcal{F}$ of functions is a function F such that $PF < \infty$ and, for all $f \in \mathcal{F}$, $|f| \le F$.

Theorem 3. *(Pollard, 1984, Theorem 24) Let $\mathcal{F}$ be a permissible*[1] *class of functions with envelope F. If $\frac{1}{n}H_1(\varepsilon, P_n, \mathcal{F}) \xrightarrow{P} 0$ for all $\varepsilon > 0$ then $\|P_n - P\| \overset{\Delta}{=} \sup_{f\in\mathcal{F}} |P_n f - Pf| \xrightarrow{a.s.} 0$.*

Remark 4. The condition that $\frac{1}{n}H_1(\varepsilon, P_n, \mathcal{F}) \xrightarrow{P} 0$ is natural in the sense that we want to make sure that the covering number does not grow exponentially fast. See Pollard (1984) for more discussion of this theorem and its conditions.

Proof. In lectures 5 and 6, we proved Glivenko-Cantelli for a special class of functions, namely indicators. This proof extends it to more general classes of functions. The proof will be similar, but some changes will need to be made.

As before, we will prove convergence in probability. A reverse-martingale argument can be used to extend the proof to show convergence almost surely.

Since $PF < \infty$, for any $\varepsilon > 0$ there exists a K such that $PF\{F > K\} < \varepsilon$. It follows that

$$\sup_{f\in\mathcal{F}} |P_n f - Pf| \le \sup_{f\in\mathcal{F}} |P_n f\{F \le K\} - Pf\{F \le K\}| + \sup_{f\in\mathcal{F}} |P_n f\{F > K\}| + \sup_{f\in\mathcal{F}} |Pf\{F > K\}|. \tag{1}$$

Furthermore, since F is an envelope,

$$\sup_{f\in\mathcal{F}} |P_n f\{F > K\}| + \sup_{f\in\mathcal{F}} |Pf\{F > K\}| \le P_n F\{F > K\} + PF\{F > K\} \xrightarrow{\text{a.s.}} 2PF\{F > K\} < 2\varepsilon.$$

Since this is true for all ε, inequality (1) means that

$$\sup_{f\in\mathcal{F}} |P_n f\{F \le K\} - Pf\{F \le K\}| \xrightarrow{P} 0 \implies \sup_{f\in\mathcal{F}} |P_n f - Pf| \xrightarrow{P} 0.$$

This tells us that we can proceed under the assumption that $|f| \le K$ for all $f \in \mathcal{F}$.

[1] Permissibility is a concept from measure theory that is not important for this class; see Pollard (1984, Appendix C, Definition 1) for details.

Now lecture 5 used two symmetrization arguments to establish bounds that helped in proving Glivenko-Cantelli for indicator functions. Both these bounds apply in this more general case, and the proofs are similar, so we will repeat only the conclusion from lecture 6, which is

$$P\{\|P_n - P\| > \varepsilon\} \leq 4P\{\|P_n^0\| > \tfrac{\varepsilon}{4}\} \text{ for } n \geq \frac{2}{\varepsilon^2},$$

where P_n^0 is the signed measure putting mass $\pm\frac{1}{n}$ at each of the observed data points $\xi = \{\xi_1, \ldots, \xi_n\}$. We will now continue, working conditionally with ξ.

Given any ξ, choose $g_1, \ldots, g_M$, where $M = N_1(\frac{\varepsilon}{8}, P_n, \mathcal{F})$ such that $\min_j P_n|f - g_j| < \frac{\varepsilon}{8}$ for all $f \in \mathcal{F}$. Denote f^* as the g_j that achieves the minimal P_n-norm distance from f. Now

$$\begin{aligned}
P\{\|P_n^0\| > \tfrac{\varepsilon}{4}|\xi\} &\leq P\{\sup_{f\in\mathcal{F}}(|P_n^0 f^*| + |P_n^0(f - f^*)|) > \tfrac{\varepsilon}{4}|\xi\} \\
&\leq P\{\sup_{f\in\mathcal{F}}(|P_n^0 f^*| + P_n|f - f^*|) > \tfrac{\varepsilon}{4}|\xi\} \\
&\leq P\{\max_j |P_n^0 g_j| > \tfrac{\varepsilon}{8}|\xi\} && \text{since } P_n|f - f^*| < \tfrac{\varepsilon}{8} \\
&= P\{\bigcup_{j=1}^{M} |P_n^0 g_j| > \tfrac{\varepsilon}{8}|\xi\} \\
&\leq \sum_{j=1}^{M} P\{|P_n^0 g_j| > \tfrac{\varepsilon}{8}|\xi\} \\
&\leq N_1(\tfrac{\varepsilon}{8}, P_n, \mathcal{F}) \max_j P\{|P_n^0 g_j| > \tfrac{\varepsilon}{8}|\xi\} \\
&\leq N_1(\tfrac{\varepsilon}{8}, P_n, \mathcal{F}) \max_j 2\exp\left(-2\frac{\left(\frac{n\varepsilon}{8}\right)^2}{\sum_{i=1}^{n}(2g_j(\xi_i))^2}\right) && \text{by Hoeffding} \\
&\leq 2N_1(\tfrac{\varepsilon}{8}, P_n, \mathcal{F}) \exp\left(-\frac{n\varepsilon^2}{128K^2}\right) && \text{since } |g_j| \leq K
\end{aligned}$$

Note that this bound does not depend on the data!

To complete the proof we must integrate over ξ: for the event $\{\log N_1(\frac{\varepsilon}{8}, P_n, \mathcal{F}) \leq \frac{n\varepsilon^2}{256K^2}\}$ we can use the bound just obtained, replacing $N_1(\frac{\varepsilon}{8}, P_n, \mathcal{F})$ with the upper bound $e^{n\varepsilon^2/256K^2}$. Otherwise, we will use 1 as a bound. That is,

$$\begin{aligned}
P\{\|P_n^0\| > \tfrac{\varepsilon}{4}\} &\leq P\{\log N_1(\tfrac{\varepsilon}{8}, P_n, \mathcal{F}) \leq \tfrac{n\varepsilon^2}{256K^2}\} 2\exp\left(-\frac{n\varepsilon^2}{256K^2}\right) + P\{\log N_1(\tfrac{\varepsilon}{8}, P_n, \mathcal{F}) > \tfrac{n\varepsilon^2}{256K^2}\} \\
&\leq \underbrace{2\exp\left(-\frac{n\varepsilon^2}{256K^2}\right)}_{\to 0} + \underbrace{P\{\log N_1(\tfrac{\varepsilon}{8}, P_n, \mathcal{F}) > \tfrac{n\varepsilon^2}{256K^2}\}}_{\xrightarrow{P} 0}.
\end{aligned}$$

□

Example 5 (A non-GC class)**.** Suppose $\mathcal{F} = \{1_A : A \subset \mathbf{R}\}$, $P = U(0,1)$, and $\mathcal{X} = (0,1)$. Consider $A = \{x_1, \ldots, x_n\}$. Then $P1_A \equiv 0$, but $P_n 1_A = 1$ for some subsets.

2 Glivenko-Cantelli and VC dimension

Lemma 6 (Approximation Lemma)**.** *(Pollard, 1984, Lemma 25) Let $\mathcal{F}$ be a class of functions with envelope F and let $\mathcal{Q}$ be a probability measure such that $\mathcal{Q}F < \infty$. Suppose graphs of $\mathcal{F}$ have finite VC dimension $\mathcal{V}$. Then*

$$N_1(\varepsilon \mathcal{Q}F, \mathcal{Q}, \mathcal{F}) \leq A\mathcal{V}(16e)^{\mathcal{V}}\varepsilon^{-(\mathcal{V}-1)}.$$

Remark 7. The exponential dependence of N_1 on $\mathcal{V}$ shown in this lemma gives an intuition for the use of the word "dimension" in *VC dimension*.

Remark 8. This lemma implies that $H_1 \leq C + (\mathcal{V} - 1) \log \frac{1}{\varepsilon}$.

Remark 9. See van der Vaart (1998, Lemma 19.15) for a tighter result.

References

Pollard, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge.