2.4  
Glivenko-Cantelli Theorems

In this chapter we prove two types of Glivenko-Cantelli theorems. The first theorem is the simplest and is based on entropy with bracketing. Its proof relies on finite approximation and the law of large numbers for real variables. The second theorem uses random $L_1$-entropy numbers and is proved through symmetrization followed by a maximal inequality.

Recall Definition 2.1.6 of the bracketing numbers of a class $\mathcal{F}$ of functions.

2.4.1 Theorem. Let $\mathcal{F}$ be a class of measurable functions such that $N_{||}(\varepsilon, \mathcal{F}, L_1(P)) < \infty$ for every $\varepsilon > 0$. Then $\mathcal{F}$ is Glivenko-Cantelli.

Proof. Fix $\varepsilon > 0$. Choose finitely many $\varepsilon$-brackets $[l_i, u_i]$ whose union contains $\mathcal{F}$ and such that $P(u_i - l_i) < \varepsilon$ for every $i$. Then, for every $f \in \mathcal{F}$, there is a bracket such that

$$(\mathbb{P}_n - P)f \leq (\mathbb{P}_n - P)u_i + P(u_i - f) \leq (\mathbb{P}_n - P)u_i + \varepsilon.$$ 

Consequently,

$$\sup_{f \in \mathcal{F}}(\mathbb{P}_n - P)f \leq \max_i(\mathbb{P}_n - P)u_i + \varepsilon.$$ 

The right side converges almost surely to $\varepsilon$ by the strong law of large numbers for real variables. Combination with a similar argument for $\inf_{f \in \mathcal{F}}(\mathbb{P}_n - P)f$ yields that $\limsup \|\mathbb{P}_n - P\|^* \leq \varepsilon$ almost surely, for every $\varepsilon > 0$. Take a sequence $\varepsilon_m \downarrow 0$ to see that the limsup must actually be zero almost surely. □
2.4.2 Example. The previous proof generalizes a well-known proof of the Glivenko-Cantelli theorem for the empirical distribution function on the real line. Indeed, the set of indicator functions of cells \((-\infty, c]\) possesses finite bracketing numbers for any underlying distribution. Simply use the brackets \([1\{(-\infty, t_i]\}, 1\{(-\infty, t_{i+1})]\})\) for a grid of points \(-\infty = t_0 < t_1 < \cdots < t_m = \infty\) with the property \(P(t_i, t_{i+1}) < \varepsilon\) for each \(i\). Bracketing numbers of many other classes of functions are discussed in Chapter 2.7.

Both the statement and the proof of the following theorem are more complicated than the previous bracketing theorem. However, its sufficiency condition for the Glivenko-Cantelli property can be checked for many classes of functions by elegant combinatorial arguments, as discussed in a later chapter. Another important note: its random entropy condition is necessary, a fact that is not proved here.

2.4.3 Theorem. Let \(\mathcal{F}\) be a \(P\)-measurable class of measurable functions with envelope \(F\) such that \(P^* F < \infty\). Let \(\mathcal{F}_M\) be the class of functions \(f1\{F \leq M\}\) when \(f\) ranges over \(\mathcal{F}\). If \(\log N(\varepsilon, \mathcal{F}_M, L_1(\mathbb{P}_n)) = o_p(n)\) for every \(\varepsilon\) and \(M > 0\), then \(\|\mathbb{P}_n - P\|_{\mathcal{F}} \to 0\) both almost surely and in mean. In particular, \(\mathcal{F}\) is Glivenko-Cantelli.

Proof. By the symmetrization Lemma 2.3.1, measurability of the class \(\mathcal{F}\), and Fubini's theorem,

\[
E^* \|\mathbb{P}_n - P\|_{\mathcal{F}} \leq 2E_x E_\varepsilon \left\| \frac{1}{n} \sum_{i=1}^n \varepsilon_i f(X_i) \right\|_{\mathcal{F}} \leq 2E_x E_\varepsilon \left\| \frac{1}{n} \sum_{i=1}^n \varepsilon_i f(X_i) \right\|_{\mathcal{F}_M} + 2P^* F\{F > M\},
\]

by the triangle inequality, for every \(M > 0\). For sufficiently large \(M\), the last term is arbitrarily small. To prove convergence in mean, it suffices to show that the first term converges to zero for fixed \(M\). Fix \(X_1, \ldots, X_n\). If \(G\) is an \(\varepsilon\)-net in \(L_1(\mathbb{P}_n)\) over \(\mathcal{F}_M\), then

\[
E_\varepsilon \left\| \frac{1}{n} \sum_{i=1}^n \varepsilon_i f(X_i) \right\|_{\mathcal{F}_M} \leq E_\varepsilon \left\| \frac{1}{n} \sum_{i=1}^n \varepsilon_i f(X_i) \right\|_G + \varepsilon.
\]

The cardinality of \(G\) can be chosen equal to \(N(\varepsilon, \mathcal{F}_M, L_1(\mathbb{P}_n))\). Bound the \(L_1\)-norm on the right by the Orlicz-norm for \(\psi_2(x) = \exp(x^2) - 1\), and use the maximal inequality Lemma 2.2.2 to find that the last expression does not exceed a multiple of

\[
\sqrt{1 + \log N(\varepsilon, \mathcal{F}_M, L_1(\mathbb{P}_n))} \sup_{f \in G} \left\| \frac{1}{n} \sum_{i=1}^n \varepsilon_i f(X_i) \right\|_{\psi_2, \mathcal{X}} + \varepsilon,
\]

Theorem 2.4.3

2.4 Glivenko-Cantelli Theorems

123