Recall that if we use empirical minimization to obtain our predictor

\[ \hat{h}_n = \arg \min_{h \in H} \hat{R}_n(h), \]

then in order to bound the quantity \( R(\hat{h}_n) - \inf_{h \in H} R(h) \), it suffices to bound the quantity

\[ \sup_{h \in H} |R(h) - \hat{R}_n(h)|. \]

Thus the uniform bound plays an important role in statistical learning theory. The Glivenko-Cantelli class is defined such that the above property holds as \( n \to \infty \).

**Definition.** \( H \) is a Glivenko-Cantelli class with respect to a probability measure \( P \) if for all \( \epsilon > 0 \),

\[ P \left( \lim_{n \to \infty} \sup_{h \in H} |P f - P_n f| = 0 \right) = 1, \]

i.e. \( \sup_{h \in H} |P f - P_n f| \) converges to zero almost surely (with probability 1). \( H \) is said to be a uniformly GC Class if the convergence is uniformly over all probability measures \( P \).

Note that Vapnik and Chervonenkis have shown that a function class is a uniformly GC class if and only if it is a VC class.

Given a set of iid real-valued random variables \( Z_1, \ldots, Z_n \) and any \( z \in \mathbb{R} \), we know that the quantity \( I(Z_i \leq z) \) is a Bernoulli random variable with mean \( P(Z \leq z) = F(z) \), where \( F(.) \) is the CDF. Furthermore, by strong law of large numbers, we know that

\[ \frac{1}{n} \sum_{i=1}^{n} I(Z_i \leq z) \to F(z) \]

almost surely. The following theorem is one of the most fundamental theorems in mathematical statistics, which generalizes the strong law of large numbers: the empirical distribution function uniformly almost surely converges to the true distribution function.

**Theorem (Glivenko-Cantelli).** Let \( Z_1, \ldots, Z_n \) be iid real-valued random variables with distribution function \( F(z) = P(Z \leq z) \). Denote the standard empirical distribution function by

\[ F_n(z) = \frac{1}{n} \sum_{i=1}^{n} I(Z_i \leq z). \]

Then

\[ P \left( \sup_{z \in \mathbb{R}} |F(z) - F_n(z)| > \epsilon \right) \leq 8(n + 1) \exp \left( -\frac{n \epsilon^2}{32} \right), \]

and in particular, by the Borel-Cantelli lemma, we have

\[ \lim_{n \to \infty} \sup_{z \in \mathbb{R}} |F(z) - F_n(z)| = 0 \text{ almost surely.} \]

**Proof.**

We use the notation \( \nu(A) := P(Z \in A) \) and \( \nu_n(A) = \frac{1}{n} \sum_{i=1}^{n} I(Z_i \in A) \) for any measurable set \( A \subset \mathbb{R} \). If we let \( \mathcal{A} \) denote the class of sets of the form \( (-\infty, z] \) for all \( z \in \mathbb{R} \), then we have

\[ \sup_{z \in \mathbb{R}} |F(z) - F_n(z)| = \sup_{A \in \mathcal{A}} |\nu(A) - \nu_n(A)|. \]
We assume $nc^2 > 2$ since otherwise the result holds trivially. The proof consists of several key steps.

(1) **Symmetrization by a ghost sample:** Introduce a ghost sample $Z'_1, \ldots, Z'_n$ which are iid together with the original sample, and denote by $\nu'_n$ the empirical measure with respect to the ghost sample. Then for $nc^2 > 2$ we have (by the symmetrization lemma)

$$
P \left( \sup_{A \in \mathcal{A}} |\nu_n(A) - \nu(A)| > \epsilon \right) \leq 2P \left( \sup_{A \in \mathcal{A}} |\nu_n(A) - \nu'_n(A)| > \epsilon / 2 \right).
$$

(2) **Symmetrization by Rademacher variables:** Let $\sigma_1, \ldots, \sigma_n$ be iid random variables, independent of $Z_1, \ldots, Z_n, Z'_1, \ldots, Z'_n$, with $P(\sigma_i = 1) = P(\sigma_i = -1) = 1/2$. Such random variables are called *Rademacher random variables*. Observe that the distribution of $Z_1, \ldots, Z_n; Z'_1, \ldots, Z'_n$ and $\sigma_1, \ldots, \sigma_n$. Thus we have

$$
P \left( \sup_{A \in \mathcal{A}} |\nu_n(A) - \nu(A)| > \epsilon \right) \leq 2P \left( \sup_{A \in \mathcal{A}} |\nu_n(A) - \nu'_n(A)| > \epsilon / 2 \right)
= 2P \left( \sup_{A \in \mathcal{A}} \frac{1}{n} \sum_{i=1}^{n} \sigma_i (I(Z_i \in A) - I(Z'_i \in A)) > \epsilon / 2 \right)
\leq 2P \left( \sup_{A \in \mathcal{A}} \frac{1}{n} \sum_{i=1}^{n} \sigma_i I(Z_i \in A) > \frac{\epsilon}{4} \right) + 2P \left( \sup_{A \in \mathcal{A}} \frac{1}{n} \sum_{i=1}^{n} \sigma_i I(Z'_i \in A) > \frac{\epsilon}{4} \right)
= 4P \left( \sup_{A \in \mathcal{A}} \frac{1}{n} \sum_{i=1}^{n} \sigma_i I(Z_i \in A) > \frac{\epsilon}{4} \right).
$$

(3) **Conditioning:** To bound the probability

$$
P \left( \sup_{A \in \mathcal{A}} \frac{1}{n} \sum_{i=1}^{n} \sigma_i I(Z_i \in A) > \frac{\epsilon}{4} \right) = P \left( \sup_{z \in \mathbb{R}} \frac{1}{n} \sum_{i=1}^{n} \sigma_i I(Z_i \leq z) > \frac{\epsilon}{4} \right)
$$

we condition on $Z_1, \ldots, Z_n$. Fix $z_1, \ldots, z_n \in \mathbb{R}$ and note that the vector $[I(z_1 \leq z), \ldots, I(z_n \leq z)]$ can take at most $(n + 1)$ possible values for any $z$. Thus conditioned on $Z_1, \ldots, Z_n$, the supremum is just a maximum over at most $n + 1$ random variables. Applying union bound we obtain

$$
P \left( \sup_{A \in \mathcal{A}} \frac{1}{n} \sum_{i=1}^{n} \sigma_i I(Z_i \in A) > \frac{\epsilon}{4} \bigg| Z_1, \ldots, Z_n \right) \leq (n + 1) \sup_{A \in \mathcal{A}} P \left( \frac{1}{n} \sum_{i=1}^{n} \sigma_i I(Z_i \in A) > \frac{\epsilon}{4} \bigg| Z_1, \ldots, Z_n \right)
$$

where the sup is outside of the probability. The next step is to find an exponential bound for the RHS.

(4) **Hoeffding’s inequality:** With $z_1, \ldots, z_n$ fixed, $\sum_{i=1}^{n} \sigma_i I(z_i \in A)$ is a sum of $n$ independent zero mean random variables between $[-1, 1]$. Thus, by Hoeffding’s inequality we have

$$
P \left( \sup_{A \in \mathcal{A}} \frac{1}{n} \sum_{i=1}^{n} \sigma_i I(Z_i \in A) > \frac{\epsilon}{4} \bigg| Z_1, \ldots, Z_n \right) \leq (n + 1) \sup_{A \in \mathcal{A}} P \left( \frac{1}{n} \sum_{i=1}^{n} \sigma_i I(Z_i \in A) > \frac{\epsilon}{4} \bigg| Z_1, \ldots, Z_n \right)
\leq 2(n + 1) \exp \left( \frac{n \epsilon^2}{32} \right).$$
Taking expectation on both side we obtain the claimed result

\[ P \left( \sup_{A \in \mathcal{A}} |\nu_n(A) - \nu(A)| > \epsilon \right) \leq 8(n + 1) \exp \left( -\frac{n\epsilon^2}{32} \right). \]