4 Risk Bounds for an Infinite Class

In the previous section, a simple bound was obtained for the deviation of the empirical error from the true probability of error of any decision rule from a finite class $\mathcal{F}$. The bound holds for any distribution $\mathcal{P}$. So, almost inevitably it may be very loose for a particular distribution $\mathcal{P}$ that defines a given classification problem. However loose the bound may be for any specific problem, it shows a theoretical guarantee that the optimal decision rule can be learned from data as long as the sample size is sufficiently large when there are finitely many candidate rules.

In practice, there are many situations where the class $\mathcal{F}$ of our choice is infinite, and it is desirable to extend the theory to the infinite case. This section regards such a theoretical extension known as the Vapnik-Chervonenkis theory. As a starter, the convergence of the empirical distribution $F_n(x)$ to the actual distribution function $F(x)$ is discussed. The proof of the following theorem (due to D. Pollard) contains the essence of the idea and techniques useful for the desired extension to the infinite class.

4.1 Glivenko-Cantelli Theorem

Theorem 5 (Glivenko-Cantelli Theorem)**.** *Let $Z_1, \ldots, Z_n$ be iid with distribution function $F(\cdot)$. Let $F_n(z) = \frac{1}{n}\sum_{i=1}^n I(Z_i \le z)$ be the empirical distribution function. Then*

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

and in particular, by the Borel Cantelli lemma,

$$\sup_{z\in\mathbb{R}} |F_n(z) - F(z)| \to 0 \quad w.p.1.$$

Proof. Let Z denote the *real* sample of $Z_1, \ldots, Z_n$. Assume $n\epsilon^2 > 2$. (If $n\epsilon^2 \le 2$, $8(n+1)\exp\{-n\epsilon^2/32\} \ge 1$.) Consider a *ghost* sample of $Z_1', \ldots, Z_n'$ iid with $F(\cdot)$ and independent of $Z_1, \ldots, Z_n$. Let $F_n'(z) = \frac{1}{n}\sum_{i=1}^n I(Z_i' \le z)$.

The following inequalities will be proved step by step.

$$\begin{aligned}
&P_Z\left(\sup_{z\in\mathbb{R}} |F_n(z) - F(z)| > \epsilon\right) \\
&\le 2\,P_{Z,Z'}\left(\sup_{z\in\mathbb{R}} |F_n(z) - F_n'(z)| > \frac{\epsilon}{2}\right) && (4.1)\\
&\le 4\,P_{\sigma,Z}\left(\sup_{z\in\mathbb{R}} \frac{1}{n}\left|\sum_{i=1}^n \sigma_i I(Z_i \le z)\right| > \frac{\epsilon}{4}\right) && (4.2)\\
&\le 4\,E_Z\left\{(n+1)\sup_{z\in\mathbb{R}} P_\sigma\left(\frac{1}{n}\left|\sum_{i=1}^n \sigma_i I(Z_i \le z)\right| > \frac{\epsilon}{4}\Big| Z\right)\right\} && (4.3)\\
&\le 4\,E_Z\left\{(n+1)\,2\,e^{-n\epsilon^2/32}\right\} && (4.4)\\
&= 8(n+1)e^{-n\epsilon^2/32}. && (4.5)
\end{aligned}$$

Step 1: *Symmetrization by a ghost sample* for proving (4.1).

If $\sup_{z\in\mathbb{R}}|F_n(z) - F(z)| > \epsilon$, there exists z^* depending on $Z_1, \ldots, Z_n$ such that $|F_n(z^*) - F(z^*)| > \epsilon$. Consider

$$\Delta := P\left(|F_n(z^*) - F(z^*)| > \epsilon \text{ and } |F'_n(z^*) - F(z^*)| < \epsilon/2\right).$$

If $|F_n(z^*) - F(z^*)| > \epsilon$ and $|F'_n(z^*) - F(z^*)| < \epsilon/2$, then $|F_n(z^*) - F'_n(z^*)| > \epsilon/2$. So, on one hand, we have

$$\Delta \le P\left(|F_n(z^*) - F'_n(z^*)| > \frac{\epsilon}{2}\right) \le P\left(\sup_{z\in\mathbb{R}}|F_n(z) - F'_n(z)| > \frac{\epsilon}{2}\right). \tag{4.6}$$

On the other hand, by Chebyshev's inequality,

$$P_{Z'}\left(|F'_n(z^*) - F(z^*)| \ge \frac{\epsilon}{2}\Big|Z\right) \le \frac{\text{var}_{Z'}(F'_n(z^*)|Z)}{\epsilon^2/4} = \frac{4F(z^*)(1-F(z^*))}{n\epsilon^2} \le \frac{1}{n\epsilon^2} \le \frac{1}{2},$$

which implies

$$P_{Z'}\left(|F'_n(z^*) - F(z^*)| < \frac{\epsilon}{2}\Big|Z\right) \ge \frac{1}{2}.$$

Thus,

$$\begin{aligned}\Delta &= E_Z\left\{I\big(|F_n(z^*) - F(z^*)| > \epsilon\big)P_{Z'}\left(|F'_n(z^*) - F(z^*)| < \frac{\epsilon}{2}\Big|Z\right)\right\}\\ &\ge E_Z\left\{I\big(|F_n(z^*) - F(z^*)| > \epsilon\big)\frac{1}{2}\right\}\\ &= \frac{1}{2}P_Z\left(|F_n(z^*) - F(z^*)| > \epsilon\right). \end{aligned} \tag{4.7}$$

Combining (4.6) and (4.7) we get,

$$\begin{aligned}&P_Z\left(\sup_{z\in\mathbb{R}}|F_n(z) - F(z)| > \epsilon\right) = P_Z\left(|F_n(z^*) - F(z^*)| > \epsilon \text{ for some } z^* \in \mathbb{R}\right)\\ &\le 2\,P_{Z,Z'}\left(\sup_{z\in\mathbb{R}}|F_n(z) - F'_n(z)| > \frac{\epsilon}{2}\right)\end{aligned}$$

Step 2: *Symmetrization by random signs* for proving (4.2).

$$F_n(z) - F'_n(z) = \frac{1}{n}\sum_{i=1}^n I(Z_i \le z) - I(Z'_i \le z).$$

Let $\sigma_1, \ldots, \sigma_n$ be iid random variables (Rademacher random variables) independent of both the real and ghost samples with $P(\sigma_1 = -1) = P(\sigma_1 = 1) = 1/2$. Let $D_i := I(Z_i \le z) - I(Z'_i \le z)$. An easy claim is that D_i and $\sigma_i D_i$ are identically distributed for all i. To see this, first note that $\sigma_i \in \{-1, 1\}$ and $D_i \in \{-1, 0, 1\}$ which implies $\sigma_i D_i \in \{-1, 0, 1\}$. Also note that $P(D_i = -1) = P(D_i = 1)$. Therefore,

$$\begin{aligned}P(\sigma_i D_i = -1) &= P(\sigma_i = -1, D_i = 1) + P(\sigma_i = 1, D_i = -1)\\ &= P(\sigma_i = -1)P(D_i = 1) + P(\sigma_i = 1)P(D_i = -1) = P(D_i = -1).\end{aligned}$$

Similarly, it is easy to see that $P(\sigma_i D_i = 1) = P(D_i = 1)$ and $P(\sigma_i D_i = 0) = P(D_i = 0)$. Now,

$$\begin{aligned} & P_{Z,Z'}\left(\sup_{z\in\mathbb{R}}|F_n(z) - F'_n(z)| > \frac{\epsilon}{2}\right) \\ & = P_{Z,Z'}\left(\sup_{z\in\mathbb{R}}\frac{1}{n}\left|\sum_{i=1}^{n}\{I(Z_i \le z) - I(Z'_i \le z)\}\right| > \frac{\epsilon}{2}\right) \\ & = P_{\sigma,Z,Z'}\left(\sup_{z\in\mathbb{R}}\frac{1}{n}\left|\sum_{i=1}^{n}\sigma_i\{I(Z_i \le z) - I(Z'_i \le z)\}\right| > \frac{\epsilon}{2}\right) \\ & \le P_{\sigma,Z,Z'}\left(\sup_{z\in\mathbb{R}}\frac{1}{n}\left|\sum_{i=1}^{n}\sigma_i I(Z_i \le z)\right| + \sup_{z\in\mathbb{R}}\frac{1}{n}\left|\sum_{i=1}^{n}\sigma_i I(Z'_i \le z)\right| > \frac{\epsilon}{2}\right) \\ & \le 2\, P_{\sigma,Z}\left(\sup_{z\in\mathbb{R}}\frac{1}{n}\left|\sum_{i=1}^{n}\sigma_i I(Z_i \le z)\right| > \frac{\epsilon}{4}\right). \end{aligned}$$

Step 3: *Conditioning* for proving (4.3).

Examine $\sum_{i=1}^{n}\sigma_i I(Z_i \le z)$ conditional on the real sample $z_1, \dots, z_n$. For fixed $z_1, \dots, z_n$, there are at most $(n+1)$ different vectors of $(I(z_1 \le z), \dots, I(z_n \le z))$ as z runs through $\mathbb{R}$. Thus using the usual union bound, $P(A \cup B) \le P(A) + P(B) \le 2 \sup_{C\in\{A,B\}} P(C)$, we obtain

$$\begin{aligned} & P_{\sigma,Z}\left(\sup_{z\in\mathbb{R}}\frac{1}{n}\left|\sum_{i=1}^{n}\sigma_i I(Z_i \le z)\right| > \frac{\epsilon}{4}\right) \\ & = E_Z P_\sigma\left(\sup_{z\in\mathbb{R}}\frac{1}{n}\left|\sum_{i=1}^{n}\sigma_i I(z_i \le z)\right| > \frac{\epsilon}{4}\middle| Z = (z_1, \dots, z_n)\right) \\ & \le E_Z\left\{(n+1)\sup_{z\in\mathbb{R}} P_\sigma\left(\frac{1}{n}\left|\sum_{i=1}^{n}\sigma_i I(z_i \le z)\right| > \frac{\epsilon}{4}\middle| Z\right)\right\}. \end{aligned}$$

Step 4: *Using Hoeffding's inequality* for proving (4.4).

Again given the sample $z_1, \dots, z_n$ for a fixed $z \in \mathbb{R}$, $\sigma_i I(z_i \le z)$ are independent bounded ($[-1, 1]$) random variables with mean

$$E_{\sigma_i}\{\sigma_i I(z_i \le z)|Z_i = z_i\} = I(z_i \le z)E\{\sigma_i\} = 0.$$

Thus Hoeffding's inequality yields

$$P_\sigma\left(\left|\sum_{i=1}^{n}\sigma_i I(z_i \le z)\right| > \frac{n\epsilon}{4}\middle| Z\right) \le 2\,\exp\left\{-2\frac{(n\epsilon/4)^2}{\sum_{i=1}^{n}(1-(-1))^2}\right\} = 2\,e^{-n\epsilon^2/32}.$$

Hence,

$$\sup_{z\in\mathbb{R}} P_\sigma\left(\left|\sum_{i=1}^{n}\sigma_i I(z_i \le z)\right| > \frac{n\epsilon}{4}\middle| Z\right) \le 2\,e^{-n\epsilon^2/32}.$$

This completes the proof. □