Probabilistic Convergence and Bounds

Junier Oliva

10-701

2/19/2013

Useful Inequalities

- Markov's Inequality:

Non-negative r.v. $Z \geq 0$ and real $a > 0$

$$\Pr(Z \geq a) \leq \frac{\mathbb{E}[Z]}{a}$$

- Chebyshev's Inequality:

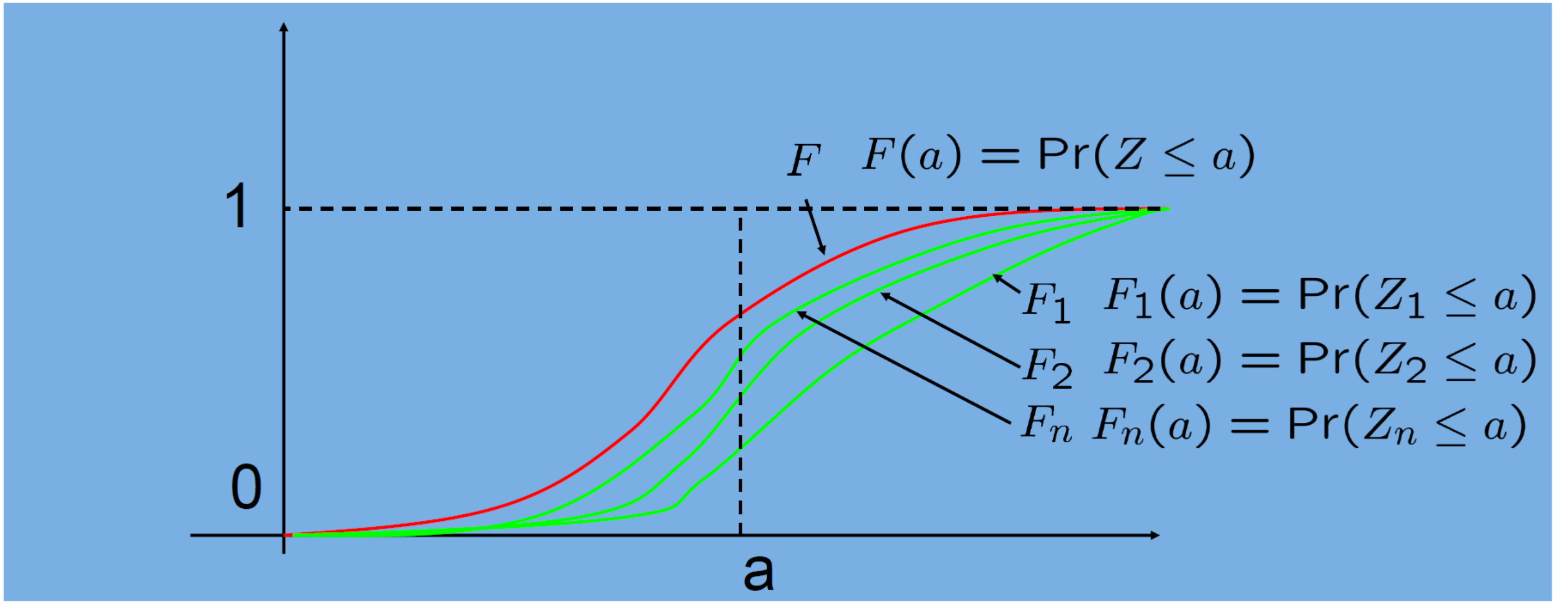
R.v. X with $\mathbb{E}[X] < \infty$, real $a > 0$

$$\Pr(|X - \mathbb{E}[X]| \geq a) \leq \frac{\mathrm{Var}(X)}{a^2} \qquad (\text{Set } Z \equiv |X - \mathbb{E}[X]|^2.)$$

Convergence in Distribution

- Notation: $Z_n \xrightarrow{d} Z,\ Z_n \xrightarrow{D} Z$
- Definition: Let F_n, F be cdfs of Z_n, and Z:

  $\lim_{n \to \infty} F_n(\omega) = F(\omega)\ \forall\ \omega \in \mathbb{R}$ s.t. F is continuous



Convergence in P^{th} Mean, L_p norm

- Notation: $Z_n \overset{L_p}{\to} Z$. If $p = 2$, $Z_n \overset{q.m.}{\to} Z$ (in quadratic mean)
- Definition: For fixed $p \geq 1$, $\lim_{n \to \infty} \mathbb{E}[|Z_n(\omega) - Z(\omega)|^p] = 0$
- Intuition:

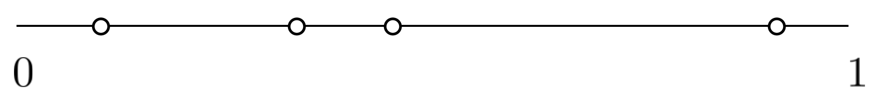
Note that for fixed n, $\mathbb{E}[|Z_n(\omega) - Z(\omega)|^p]$ is a determenistic value.
Let $a_n \equiv \mathbb{E}[|Z_n(\omega) - Z(\omega)|^p]$
Hence, $a_n \to 0$

Almost Sure Convergence

- Notation: $Z_n \overset{a.s.}{\to} Z$, $Z_n \to Z$ w.p. 1
- Definition: $P\left(\omega \in \Omega : \lim_{n\to\infty} Z_n(\omega) = Z(\omega)\right) = 1$
- Intuition:

Let $A \equiv \left\{\omega \in \Omega : \lim_{n\to\infty} Z_n(\omega) = Z(\omega)\right\}$, then $P(A) = 1$.

E.g. $\Omega = [0,1]$, $P = \text{Unif}[0,1]$.

$A =$ 

Convergence in Probability

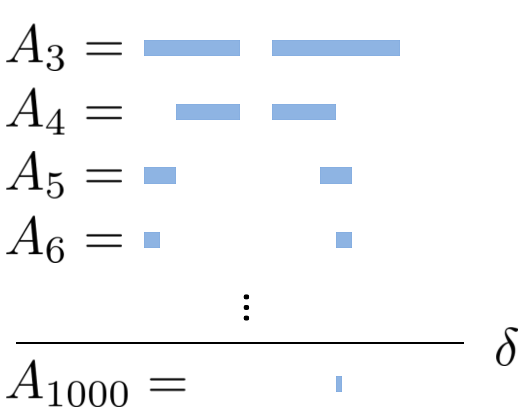
- **Notation:** $Z_n \xrightarrow{p} Z$
- **Definition:** $\forall \epsilon > 0 \;\; \lim_{n \to \infty} P\left(\omega \in \Omega : |Z_n(\omega) - Z(\omega)| \geq \epsilon\right) = 0$

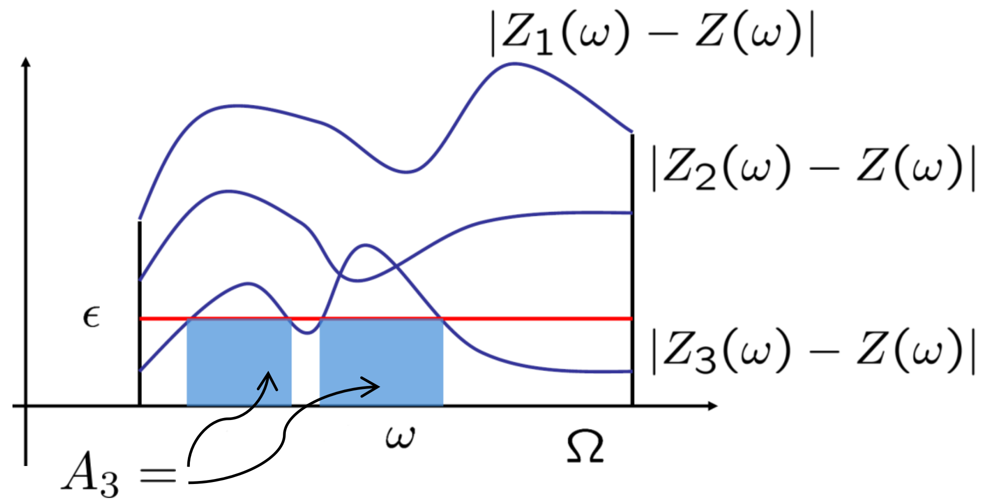
$$\forall \epsilon > 0 \;\; \lim_{n \to \infty} P\left(|Z_n - Z| \geq \epsilon\right) = 0$$

- **Intuition:** $\forall \epsilon > 0 \; \forall \delta > 0 \; \exists N \text{ s.t. } n > N \implies P\left(|Z_n - Z| \geq \epsilon\right) < \delta$

Let $\epsilon > 0$, $\delta > 0$ be given

Define $A_n \equiv \{\omega \in \Omega : |Z_n(\omega) - Z(\omega)| \geq \epsilon\}$





Relation Among Convergences

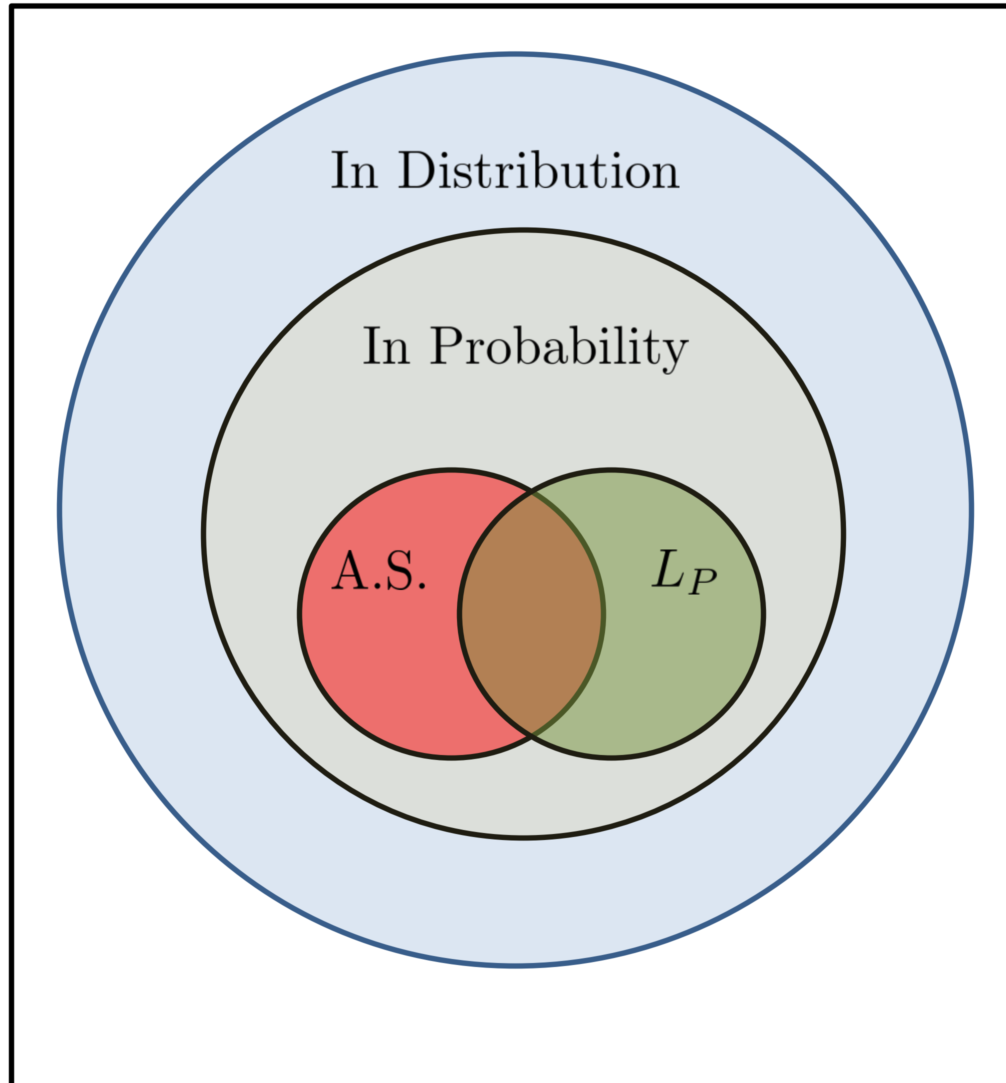
$$Z_n \xrightarrow{a.s.} Z \qquad Z_n \xrightarrow{L_p} Z$$

$$\Downarrow$$

$$Z_n \xrightarrow{p} Z$$

$$\Downarrow$$

$$Z_n \xrightarrow{d} Z$$



Convergence in Quadratic Mean => Convergence in Probability

Suppose $Z_n \stackrel{q.m.}{\to} Z$; i.e., $\lim_{n\to\infty} \mathbb{E}[|Z_n - Z|^2] = 0$.

Convergence in Quadratic Mean => Convergence in Probability

Suppose $Z_n \stackrel{q.m.}{\to} Z$; i.e., $\lim_{n\to\infty} \mathbb{E}[|Z_n - Z|^2] = 0$.

Let $\epsilon > 0$, and $\delta > 0$ be given.

Convergence in Quadratic Mean => Convergence in Probability

Suppose $Z_n \overset{q.m.}{\to} Z$; i.e., $\lim_{n\to\infty} \mathbb{E}[|Z_n - Z|^2] = 0$.

Let $\epsilon > 0$, and $\delta > 0$ be given.

Let N be s.t. $n > N \implies \mathbb{E}[|Z_n - Z|^2] \leq \epsilon^2\delta$

Convergence in Quadratic Mean => Convergence in Probability

Suppose $Z_n \overset{q.m.}{\to} Z$; i.e., $\lim_{n\to\infty} \mathbb{E}[|Z_n - Z|^2] = 0$.

Let $\epsilon > 0$, and $\delta > 0$ be given.

Let N be s.t. $n > N \implies \mathbb{E}[|Z_n - Z|^2] \leq \epsilon^2\delta$

Then, by Markov's: $\Pr(|Z_n - Z|^2 > \epsilon^2) \leq \frac{\mathbb{E}[|Z_n - Z|^2]}{\epsilon^2}$

Convergence in Quadratic Mean => Convergence in Probability

Suppose $Z_n \stackrel{q.m.}{\rightarrow} Z$; i.e., $\lim\limits_{n\rightarrow\infty} \mathbb{E}[|Z_n - Z|^2] = 0$.

Let $\epsilon > 0$, and $\delta > 0$ be given.

Let N be s.t. $n > N \implies \mathbb{E}[|Z_n - Z|^2] \leq \epsilon^2\delta$

Then, by Markov's: $\Pr(|Z_n - Z|^2 > \epsilon^2) \leq \frac{\mathbb{E}[|Z_n - Z|^2]}{\epsilon^2}$

So, $n > N \implies \Pr(|Z_n - Z|^2 > \epsilon^2) \leq \frac{\mathbb{E}[|Z_n - Z|^2]}{\epsilon^2} \leq \frac{\epsilon^2\delta}{\epsilon^2} = \delta$

Convergence in Quadratic Mean => Convergence in Probability

Suppose $Z_n \stackrel{q.m.}{\to} Z$; i.e., $\lim_{n\to\infty} \mathbb{E}[|Z_n - Z|^2] = 0$.

Let $\epsilon > 0$, and $\delta > 0$ be given.

Let N be s.t. $n > N \implies \mathbb{E}[|Z_n - Z|^2] \leq \epsilon^2\delta$

Then, by Markov's: $\Pr(|Z_n - Z|^2 > \epsilon^2) \leq \frac{\mathbb{E}[|Z_n - Z|^2]}{\epsilon^2}$

So, $n > N \implies \Pr(|Z_n - Z|^2 > \epsilon^2) \leq \frac{\mathbb{E}[|Z_n - Z|^2]}{\epsilon^2} \leq \frac{\epsilon^2\delta}{\epsilon^2} = \delta$

Hence, $\forall \epsilon > 0 \; \forall \delta > 0 \; \exists N$ s.t. $n > N \implies P\left(|Z_n - Z| \geq \epsilon\right) < \delta$.

Convergence in Probability => Convergence in Distribution

Suppose $Z_n \xrightarrow{p} Z$; i.e., $\forall \epsilon > 0 \ \lim_{n\to\infty} P\left(|Z_n - Z| \geq \epsilon\right) = 0$

Convergence in Probability => Convergence in Distribution

Suppose $Z_n \xrightarrow{p} Z$; i.e., $\forall \epsilon > 0 \ \lim_{n\to\infty} P\left(|Z_n - Z| \geq \epsilon\right) = 0$

Let F be cdf of Z, F_n cdf of Z_n, z be a point where F is continuous.

Convergence in Probability => Convergence in Distribution

Suppose $Z_n \xrightarrow{p} Z$; i.e., $\forall \epsilon > 0 \ \lim_{n \to \infty} P\left(|Z_n - Z| \geq \epsilon\right) = 0$

Let F be cdf of Z, F_n cdf of Z_n, z be a point where F is continuous.

Fix $\epsilon > 0$

Convergence in Probability => Convergence in Distribution

Suppose $Z_n \xrightarrow{p} Z$; i.e., $\forall \epsilon > 0 \ \lim_{n \to \infty} P\left(|Z_n - Z| \geq \epsilon\right) = 0$

Let F be cdf of Z, F_n cdf of Z_n, z be a point where F is continuous.

Fix $\epsilon > 0$

$$F_n(z) = P(Z_n \leq z)$$

Convergence in Probability => Convergence in Distribution

Suppose $Z_n \xrightarrow{p} Z$; i.e., $\forall \epsilon > 0 \ \lim_{n \to \infty} P\left(|Z_n - Z| \geq \epsilon\right) = 0$

Let F be cdf of Z, F_n cdf of Z_n, z be a point where F is continuous.

Fix $\epsilon > 0$

$$\begin{aligned} F_n(z) =& P(Z_n \leq z) \\ =& P(Z_n \leq z, Z \leq z + \epsilon) + P(Z_n \leq z, Z > z + \epsilon) \end{aligned}$$

Convergence in Probability => Convergence in Distribution

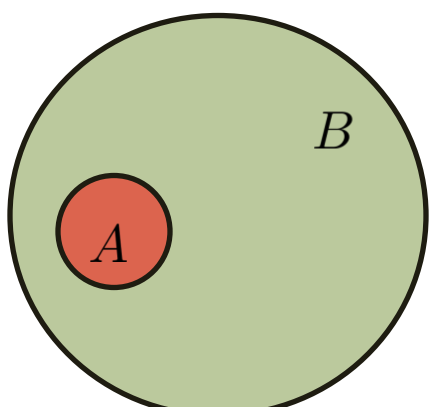
Suppose $Z_n \xrightarrow{p} Z$; i.e., $\forall \epsilon > 0 \ \lim_{n\to\infty} P\left(|Z_n - Z| \geq \epsilon\right) = 0$

Let F be cdf of Z, F_n cdf of Z_n, z be a point where F is continuous.

Fix $\epsilon > 0$

$$
\begin{aligned}
F_n(z) =& P(Z_n \leq z) \\
=& P(Z_n \leq z, Z \leq z+\epsilon) + P(Z_n \leq z, Z > z+\epsilon) \\
\leq& P(Z \leq z+\epsilon) + P(Z > Z_n + \epsilon)
\end{aligned}
$$

Since, If $A \implies B$, then $\Pr(A) \leq \Pr(B)$.



Convergence in Probability => Convergence in Distribution

Suppose $Z_n \xrightarrow{p} Z$; i.e., $\forall \epsilon > 0 \ \lim_{n \to \infty} P\left(|Z_n - Z| \geq \epsilon\right) = 0$

Let F be cdf of Z, F_n cdf of Z_n, z be a point where F is continuous.

Fix $\epsilon > 0$

$$\begin{aligned}
F_n(z) &= P(Z_n \leq z) \\
&= P(Z_n \leq z, Z \leq z + \epsilon) + P(Z_n \leq z, Z > z + \epsilon) \\
&\leq P(Z \leq z + \epsilon) + P(Z > Z_n + \epsilon) \\
&\leq F(z + \epsilon) + P(Z - Z_n > \epsilon) + P(Z - Z_n < -\epsilon)
\end{aligned}$$

Convergence in Probability => Convergence in Distribution

Suppose $Z_n \xrightarrow{p} Z$; i.e., $\forall \epsilon > 0 \ \lim_{n\to\infty} P\left(|Z_n - Z| \geq \epsilon\right) = 0$

Let F be cdf of Z, F_n cdf of Z_n, z be a point where F is continuous.

Fix $\epsilon > 0$

$$\begin{aligned}
F_n(z) =& P(Z_n \leq z) \\
=& P(Z_n \leq z, Z \leq z + \epsilon) + P(Z_n \leq z, Z > z + \epsilon) \\
\leq& P(Z \leq z + \epsilon) + P(Z > Z_n + \epsilon) \\
\leq& F(z + \epsilon) + P(Z - Z_n > \epsilon) + P(Z - Z_n < -\epsilon) \\
=& F(z + \epsilon) + P(|Z - Z_n| > \epsilon)
\end{aligned}$$

Convergence in Probability => Convergence in Distribution

Suppose $Z_n \xrightarrow{p} Z$; i.e., $\forall \epsilon > 0 \ \lim_{n\to\infty} P\left(|Z_n - Z| \geq \epsilon\right) = 0$

Let F be cdf of Z, F_n cdf of Z_n, z be a point where F is continuous.

Fix $\epsilon > 0$

$$\begin{aligned}
F_n(z) =& P(Z_n \leq z) \\
=& P(Z_n \leq z, Z \leq z + \epsilon) + P(Z_n \leq z, Z > z + \epsilon) \\
\leq& P(Z \leq z + \epsilon) + P(Z > Z_n + \epsilon) \\
\leq& F(z + \epsilon) + P(Z - Z_n > \epsilon) + P(Z - Z_n < -\epsilon) \\
=& F(z + \epsilon) + P(|Z - Z_n| > \epsilon)
\end{aligned}$$

$$\begin{aligned}
F(z - \epsilon) =& P(Z \leq z - \epsilon) \\
=& P(Z \leq z - \epsilon, Z_n \leq z) + P(Z \leq z - \epsilon, Z_n > z) \\
\leq& F_n(z) + P(|Z_n - Z| > \epsilon)
\end{aligned}$$

Convergence in Probability => Convergence in Distribution

Hence,

$$F(z-\epsilon) - P(|Z - Z_n| > \epsilon) \leq F_n(z) \leq F(z+\epsilon) + P(|Z_n - Z| > \epsilon)$$

Convergence in Probability => Convergence in Distribution

Hence,

$$F(z-\epsilon) - P(|Z - Z_n| > \epsilon) \le F_n(z) \le F(z+\epsilon) + P(|Z_n - Z| > \epsilon)$$

Thus, $n \to \infty$:

$$F(z-\epsilon) \le \liminf_{n\to\infty} F_n(z) \le \limsup_{n\to\infty} F_n(z) \le F(z+\epsilon)$$

Recall:

$$\liminf_{n\to\infty} F_n(z) \equiv \lim_{n\to\infty}\left(\inf_{m\ge n} F_m(z)\right) \le \limsup_{n\to\infty} F_n(z) \equiv \lim_{n\to\infty}\left(\sup_{m\ge n} F_m(z)\right)$$

Convergence in Probability => Convergence in Distribution

Hence,

$$F(z-\epsilon) - P(|Z - Z_n| > \epsilon) \le F_n(z) \le F(z+\epsilon) + P(|Z_n - Z| > \epsilon)$$

Thus, $n \to \infty$:

$$F(z-\epsilon) \le \liminf_{n\to\infty} F_n(z) \le \limsup_{n\to\infty} F_n(z) \le F(z+\epsilon)$$

This holds for $\forall \epsilon > 0$, $\epsilon \to 0$: $F(z) = \lim_{n\to\infty} F_n(z)$

Hoeffding's Bound

Let X_i for $i \in \{1, .., n\}$ be independent r.v. with $a_i \leq X_i \leq b_i$, and $\overline{X}_n = \frac{1}{n}\sum_i X_i$, then:

$$\Pr(|\overline{X}_n - \mathrm{E}[\overline{X}_n]| \geq t) \leq 2\exp\left(-\frac{2n^2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right)$$

Union Bound

For events A_i, $\Pr\left(\cup_i A_i\right) \leq \sum_i \Pr(A_i)$

Empirical Optimization Set-up

Let $\mathcal{H} = \{h_1, ..., h_m\}$, $\mathcal{D} = \{(X_1, Y_1), \ldots, (X_n, Y_n)\}$

Suppose we wish to find the best classifier $h \in \mathcal{H}$ to minimize its true risk $R(h) \equiv \mathbb{E}[h(X) \neq Y]$.

Define:

$$\hat{R}_n(h) \equiv \tfrac{1}{n} \textstyle\sum_{i=1}^n I\{h(X_i) \neq Y_i\}$$

$$\hat{h} \equiv \text{argmin}_{h \in \mathcal{H}} \hat{R}_n(h)$$

$$h^* \equiv \text{argmin}_{h \in \mathcal{H}} R(h)$$

We want to bound $|R(\hat{h}) - R(h^*)|$ with high probability.

To do so, we prove that $\sup_{h \in \mathcal{H}} |\hat{R}_n(h) - R(h)| < \epsilon$ with high probability.

Bounding Worst Case With High Probability

$$\Pr\left(\sup_{h\in\mathcal{H}} |\hat{R}_n(h) - R(h)| < \epsilon\right)$$

$$= \Pr\left(\forall i \in \{1,\ldots,m\},\ |\hat{R}_n(h_i) - R(h_i)| < \epsilon\right)$$

Bounding Worst Case With High Probability

$$\Pr\left(\sup_{h\in\mathcal{H}} |\hat{R}_n(h) - R(h)| < \epsilon\right)$$

$$= \Pr\left(\forall i \in \{1,\ldots,m\},\ |\hat{R}_n(h_i) - R(h_i)| < \epsilon\right)$$

$$= 1 - \Pr\left(\exists i\ |\hat{R}_n(h_i) - R(h_i)| \geq \epsilon\right)$$

Bounding Worst Case With High Probability

$$
\begin{aligned}
&\Pr\left(\sup_{h\in\mathcal{H}} |\hat{R}_n(h) - R(h)| < \epsilon\right)\\
&\qquad = \Pr\left(\forall i \in \{1,\ldots,m\},\ |\hat{R}_n(h_i) - R(h_i)| < \epsilon\right)\\
&\qquad = 1 - \Pr\left(\exists i\ |\hat{R}_n(h_i) - R(h_i)| \geq \epsilon\right)\\
&\qquad = 1 - \Pr\left(\cup_i\ |\hat{R}_n(h_i) - R(h_i)| \geq \epsilon\right)
\end{aligned}
$$

Bounding Worst Case With High Probability

$$
\begin{aligned}
&\Pr\left(\sup_{h\in\mathcal{H}} |\hat{R}_n(h) - R(h)| < \epsilon\right) \\
&\qquad = \Pr\left(\forall i \in \{1,\dots,m\},\ |\hat{R}_n(h_i) - R(h_i)| < \epsilon\right) \\
&\qquad = 1 - \Pr\left(\exists i\ |\hat{R}_n(h_i) - R(h_i)| \geq \epsilon\right) \\
&\qquad = 1 - \Pr\left(\cup_i\ |\hat{R}_n(h_i) - R(h_i)| \geq \epsilon\right) \\
&\qquad \geq 1 - \sum_{i=1}^{m} \Pr\left(|\hat{R}_n(h_i) - R(h_i)| \geq \epsilon\right)
\end{aligned}
$$

Bounding Worst Case With High Probability

$$
\begin{aligned}
&\Pr\left(\sup_{h\in\mathcal{H}} |\hat{R}_n(h) - R(h)| < \epsilon\right) \\
&\qquad = \Pr\left(\forall i \in \{1,\ldots,m\},\ |\hat{R}_n(h_i) - R(h_i)| < \epsilon\right) \\
&\qquad = 1 - \Pr\left(\exists i\ |\hat{R}_n(h_i) - R(h_i)| \geq \epsilon\right) \\
&\qquad = 1 - \Pr\left(\cup_i\ |\hat{R}_n(h_i) - R(h_i)| \geq \epsilon\right) \\
&\qquad \geq 1 - \sum_{i=1}^{m} \Pr\left(|\hat{R}_n(h_i) - R(h_i)| \geq \epsilon\right) \\
&\qquad = 1 - \sum_{i=1}^{m} \Pr\left(\left|\frac{1}{n}\sum_{j=1}^{n} I\{h_i(X_j) \neq Y_i\} - \mathbb{E}[I\{h_i(X) \neq Y\}]\right| \geq \epsilon\right)
\end{aligned}
$$

Bounding Worst Case With High Probability

$$
\begin{aligned}
&\Pr\left(\sup_{h\in\mathcal{H}} |\hat{R}_n(h) - R(h)| < \epsilon\right) \\
&\qquad = \Pr\left(\forall i \in \{1,\ldots,m\},\ |\hat{R}_n(h_i) - R(h_i)| < \epsilon\right) \\
&\qquad = 1 - \Pr\left(\exists i\ |\hat{R}_n(h_i) - R(h_i)| \geq \epsilon\right) \\
&\qquad = 1 - \Pr\left(\cup_i\ |\hat{R}_n(h_i) - R(h_i)| \geq \epsilon\right) \\
&\qquad \geq 1 - \sum_{i=1}^{m} \Pr\left(|\hat{R}_n(h_i) - R(h_i)| \geq \epsilon\right) \\
&\qquad = 1 - \sum_{i=1}^{m} \Pr\left(\left|\frac{1}{n}\sum_{j=1}^{n} I\{h_i(X_j) \neq Y_i\} - \mathbb{E}[I\{h_i(X) \neq Y\}]\right| \geq \epsilon\right) \\
&\qquad \geq 1 - 2m\exp\left(-2n\epsilon^2\right)
\end{aligned}
$$

Nearly Optimal Estimator

Suppose $\sup_{h\in\mathcal{H}} |\hat{R}_n(h) - R(h)| < \epsilon$ holds.

$R(h^*) \leq R(\hat{h})$

$$R(h) \equiv \mathbb{E}[h(X) \neq Y] \qquad \hat{R}_n(h) \equiv \frac{1}{n}\sum_{i=1}^{n} I\{h(X_i) \neq Y_i\}$$

$$h^* \equiv \mathrm{argmin}_{h\in\mathcal{H}} R(h) \qquad \hat{h} \equiv \mathrm{argmin}_{h\in\mathcal{H}} \hat{R}_n(h)$$

Nearly Optimal Estimator

Suppose $\sup_{h \in \mathcal{H}} |\hat{R}_n(h) - R(h)| < \epsilon$ holds.

$$\begin{aligned} R(h^*) &\leq R(\hat{h}) \\ &\leq \hat{R}_n(\hat{h}) + \epsilon \quad \text{since } R(\hat{h}) - \hat{R}_n(\hat{h}) < \epsilon \end{aligned}$$

$$R(h) \equiv \mathbb{E}[h(X) \neq Y] \qquad \hat{R}_n(h) \equiv \tfrac{1}{n} \textstyle\sum_{i=1}^n I\{h(X_i) \neq Y_i\}$$

$$h^* \equiv \text{argmin}_{h \in \mathcal{H}} R(h) \qquad \hat{h} \equiv \text{argmin}_{h \in \mathcal{H}} \hat{R}_n(h)$$

Nearly Optimal Estimator

Suppose $\sup_{h \in \mathcal{H}} |\hat{R}_n(h) - R(h)| < \epsilon$ holds.

$$
\begin{aligned}
R(h^*) &\leq R(\hat{h}) \\
&\leq \hat{R}_n(\hat{h}) + \epsilon \\
&\leq \hat{R}_n(h^*) + \epsilon
\end{aligned}
$$

$$R(h) \equiv \mathbb{E}[h(X) \neq Y] \qquad \hat{R}_n(h) \equiv \frac{1}{n} \sum_{i=1}^n I\{h(X_i) \neq Y_i\}$$

$$h^* \equiv \text{argmin}_{h \in \mathcal{H}} R(h) \qquad \hat{h} \equiv \text{argmin}_{h \in \mathcal{H}} \hat{R}_n(h)$$

Nearly Optimal Estimator

Suppose $\sup_{h \in \mathcal{H}} |\hat{R}_n(h) - R(h)| < \epsilon$ holds.

$$
\begin{aligned}
R(h^*) &\leq R(\hat{h}) \\
&\leq \hat{R}_n(\hat{h}) + \epsilon \\
&\leq \hat{R}_n(h^*) + \epsilon \\
&\leq R(h^*) + 2\epsilon \quad \text{since } \hat{R}_n(h^*) - R(h^*) < \epsilon
\end{aligned}
$$

$$R(h) \equiv \mathbb{E}[h(X) \neq Y] \qquad \hat{R}_n(h) \equiv \frac{1}{n} \sum_{i=1}^{n} I\{h(X_i) \neq Y_i\}$$

$$h^* \equiv \text{argmin}_{h \in \mathcal{H}} R(h) \qquad \hat{h} \equiv \text{argmin}_{h \in \mathcal{H}} \hat{R}_n(h)$$

Nearly Optimal Estimator

Suppose $\sup_{h \in \mathcal{H}} |\hat{R}_n(h) - R(h)| < \epsilon$ holds.

$$\begin{aligned} R(h^*) &\le R(\hat{h}) \\ &\le \hat{R}_n(\hat{h}) + \epsilon \\ &\le \hat{R}_n(h^*) + \epsilon \\ &\le R(h^*) + 2\epsilon \end{aligned}$$

Thus, $|R(\hat{h}) - R(h^*)| < 2\epsilon$

$$R(h) \equiv \mathbb{E}[h(X) \neq Y] \qquad \hat{R}_n(h) \equiv \frac{1}{n}\sum_{i=1}^{n} I\{h(X_i) \neq Y_i\}$$

$$h^* \equiv \text{argmin}_{h \in \mathcal{H}} R(h) \qquad \hat{h} \equiv \text{argmin}_{h \in \mathcal{H}} \hat{R}_n(h)$$

All Together

Thus, $|R(\hat{h}) - R(h^*)| < 2\epsilon$, with probablity atleast $1 - 2m \exp\left(-2n\epsilon^2\right)$

All Together

Thus, $|R(\hat{h}) - R(h^*)| < 2\epsilon$, with probablity atleast $1 - 2m\exp\left(-2n\epsilon^2\right)$

And, $|R(\hat{h}) - R(h^*)| < 2\sqrt{\frac{\log(2m) - log(\delta)}{2n}}$, with probablity atleast $1 - \delta$

Thank You

Convergence in Probability => Convergence in Distribution

Suppose $Z_n \xrightarrow{p} Z$; i.e., $\forall \epsilon > 0 \ \lim_{n\to\infty} P\left(|Z_n - Z| \geq \epsilon\right) = 0$

Let F be cdf of Z, F_n cdf of Z_n, z be a point where F is continuous.

Fix $\epsilon > 0$

$$\begin{aligned}
F_n(z) =& P(Z_n \leq z) \\
=& P(Z_n \leq z, Z \leq z+\epsilon) + P(Z_n \leq z, Z > z+\epsilon) \\
\leq& P(Z \leq z+\epsilon) + P(Z > Z_n + \epsilon) \\
\leq& F(z+\epsilon) + P(Z - Z_n > \epsilon) + P(Z - Z_n < -\epsilon) \\
\leq& F(z+\epsilon) + P(|Z - Z_n| > \epsilon)
\end{aligned}$$

$$\begin{aligned}
F(z-\epsilon) =& P(Z \leq z-\epsilon) \\
=& P(Z \leq z-\epsilon, Z_n \leq z) + P(Z \leq z-\epsilon, Z_n > z) \\
\leq& F_n(z) + P(|Z_n - Z| > \epsilon)
\end{aligned}$$

Convergence in Probability => Convergence in Distribution

Hence,

$$F(z-\epsilon) - P(|Z - Z_n| > \epsilon) \leq F_n(z) \leq F(z+\epsilon) + P(|Z_n - Z| > \epsilon)$$

Thus, $n \to \infty$:

$$F(z-\epsilon) \leq \liminf_{n\to\infty} F_n(z) \leq \limsup_{n\to\infty} F_n(z) \leq F(z+\epsilon)$$

This holds for $\forall \epsilon > 0$, $\epsilon \to 0$: $F(z) = \lim_{n\to\infty} F_n(z)$

Bounding Worst Case With High Probability

$$
\begin{aligned}
&\Pr\left(\sup_{h\in\mathcal{H}} |\hat{R}_n(h) - R(h)| < \epsilon\right) \\
&\qquad = \Pr\left(\forall i \in \{1,\ldots,m\},\ |\hat{R}_n(h_i) - R(h_i)| < \epsilon\right) \\
&\qquad = 1 - \Pr\left(\exists i\ |\hat{R}_n(h_i) - R(h_i)| \geq \epsilon\right) \\
&\qquad = 1 - \Pr\left(\cup_i\ |\hat{R}_n(h_i) - R(h_i)| \geq \epsilon\right) \\
&\qquad \geq 1 - \sum_{i=1}^{m} \Pr\left(|\hat{R}_n(h_i) - R(h_i)| \geq \epsilon\right) \\
&\qquad = 1 - \sum_{i=1}^{m} \Pr\left(\left|\frac{1}{n}\sum_{j=1}^{n} I\{h_i(X_j) \neq Y_i\} - \mathbb{E}[I\{h_i(X) \neq Y\}]\right| \geq \epsilon\right) \\
&\qquad \geq 1 - 2m\exp\left(-2n\epsilon^2\right)
\end{aligned}
$$

Bounding Worst Case With High Probability

$$
\begin{aligned}
&\Pr\left(\sup_{h\in\mathcal{H}} |\hat{R}_n(h) - R(h)| < \epsilon\right) \\
&\quad = \Pr\left(\forall i \in \{1,\ldots,m\},\ |\hat{R}_n(h_i) - R(h_i)| < \epsilon\right) \\
&\quad = 1 - \Pr\left(\exists i\ |\hat{R}_n(h_i) - R(h_i)| \geq \epsilon\right) \\
&\quad = 1 - \Pr\left(\cup_i\ |\hat{R}_n(h_i) - R(h_i)| \geq \epsilon\right) \\
&\quad \geq 1 - \sum_{i=1}^{m} \Pr\left(|\hat{R}_n(h_i) - R(h_i)| \geq \epsilon\right) \\
&\quad = 1 - \sum_{i=1}^{m} \Pr\left(\left|\frac{1}{n}\sum_{j=1}^{n} I\{h_i(X_j) \neq Y_i\} - \mathbb{E}[I\{h_i(X) \neq Y\}]\right| \geq \epsilon\right) \\
&\quad \geq 1 - 2m\exp\left(-2n\epsilon^2\right)
\end{aligned}
$$

Nearly Optimal Estimator

Suppose $\sup_{h\in\mathcal{H}} |\hat{R}_n(h) - R(h)| < \epsilon$ holds.

$$
\begin{aligned}
R(h^*) &\le R(\hat{h}) \\
&\le \hat{R}_n(\hat{h}) + \epsilon \\
&\le \hat{R}_n(h^*) + \epsilon \\
&\le R(h^*) + 2\epsilon
\end{aligned}
$$

$$R(h) \equiv \mathbb{E}[h(X) \neq Y] \qquad \hat{R}_n(h) \equiv \tfrac{1}{n}\textstyle\sum_{i=1}^n I\{h(X_i) \neq Y_i\}$$

$$h^* \equiv \mathrm{argmin}_{h\in\mathcal{H}} R(h) \qquad \hat{h} \equiv \mathrm{argmin}_{h\in\mathcal{H}} \hat{R}_n(h)$$