

Friederike Horn & Bileam Scheuven

Convergence of random variables

Solution

- a) For any h we divide $[0, 1]$ into k subintervals of length/measure $\frac{1}{2^h} = 2^{-h}$.
 Since $2^h \geq \frac{n}{2}$, $h \geq \log_2(\frac{n}{2}) = \log_2(n) - 1$, we know that:

$$\lim_{n \rightarrow \infty} 2^{-h} = 0$$

Since the length of these intervals tends to 0, $\mathbb{P}(X_n(\omega)) \xrightarrow{p} 0$

However for $\frac{k}{2^h} \leq \omega \leq \frac{k+1}{2^h}$, we still have probability 1, and thus no convergence a.s.

- b) Let $(B, \mathcal{A}, \lambda)$ be a probability space. and X_n a sequence of random variables that satisfies $X_n \rightarrow X$ for $n \rightarrow \infty$.

In particular this means that we can select a subsequence n_k such that

$$A_k = \{\omega | \lambda((X_{n_k}(\omega) - X) > \frac{1}{2^k}) \leq \frac{1}{2^k}\}.$$

Furthermore, we can choose these sets such that $n_k + 1 > n_k$. In this way we can guarantee that the overall weight of all these sets exists, i.e.:

$$\sum_{k=0}^{\infty} \lambda A_k \leq \sum_{k=0}^{\infty} 2^{-k} < \infty.$$

Then the Borel-Cantelli lemma holds that

$$\mu(\limsup_{k \rightarrow \infty} A_k) = 0,$$

and therefore the subsequence X_{n_k} is almost surely convergent to X .

- c) In the example above we can easily find a subsequence by selecting $X_{n_k} = \mathbb{1}_{[0, \frac{1}{2^k}]}$. We can directly see that for this subsequence

$$\sup\{\omega \in \Omega : |X_{n_k}(\omega) - X(\omega)| > \epsilon\} = [0, \frac{1}{2^k}]$$

and thus

$$\mathbb{P}\left(\limsup_{n \rightarrow \infty} \{\omega \in \Omega : |X_{n_k}(\omega) - X(\omega)| > \epsilon\}\right) = \mathbb{P}\left(\lim_{n \rightarrow \infty} [0, \frac{1}{2^k}]\right) = 0.$$

Limit theorems

Solution

- a) Each sock $S_i \sim \text{Ber}(p)$, thus $\mathbb{E}S_i = p$, $X_n = \frac{\sum_{i=1}^n S_i}{n}$.

$$\mathbb{E}X_n = \frac{\sum_{i=1}^n \mathbb{E}S_i}{n}$$

$$\begin{aligned}
&= \frac{n \cdot p}{n} \\
&= p
\end{aligned}$$

By the strong law of large numbers (SLLN), $X_n \xrightarrow{a.s.} \mathbb{E}X_n = p$. If S_i is the indicator variable for whether the sock was black, it follows that $(1-S_i) = 1$ if the sock is white, and we find $\mathbb{E}(1-S_i) = 1 - \mathbb{E}S_1 = 1-p$. Therefore by the same logic $Y_i = \frac{\sum S_i}{\sum 1-S_i}$ converges to its expectation as $n \rightarrow \infty$, yielding $\frac{p}{1-p}$.

Estimation error in learning theory

Solution

- a) For a single predictor h , since the true risk is simply the expectation of the empirical risk, we can bound it with Hoeffdings inequality.

$$P(|R_n(h) - R(h)| \geq \epsilon) \leq 2e^{-2n\epsilon^2}$$

The probability that the supremum over all $h \in \mathcal{H}$ deviates by more than ϵ is the same as $1 - P(\text{no } h \text{ deviates by more than } \epsilon)$, therefore we can apply the Union bound over all h .

$$P(A_n) \leq 2|\mathcal{H}|e^{-2n\epsilon^2}$$

- b) We know that:

$$P(|R_n(h) - R(h)| \geq \epsilon) \leq 2e^{-2n\epsilon^2}$$

Since $\lim_{n \rightarrow \infty} 2e^{-2n\epsilon^2} = 0$, $R_n(h) \rightarrow R(h)$ almost surely .

- c) For any h including h^* , $|R_n(h) - R(h)| \leq \sup_{h \in \mathcal{H}} |R_n(h) - R(h)|$. Therefore the empirically minimizing h can be at most this far away from its true risk, and the true risk can be at most this far away from the empirical risk of the risk of the true minimizer, yielding the upper bound on the distance:

$$|R(H_n) - R(h^*)| \leq 2 \sup_{h \in \mathcal{H}} |R_n(h) - R(h)|$$

(The absolute value can be omitted since we know $R(h^*)$ to be minimal). Since we know from b) that this distance tends to 0, we conclude that $h_n \rightarrow h^*$ almost surely.

- d) If \mathcal{H} is not finite, the bound from b) becomes useless, as $M = 2|\mathcal{H}|$ is no longer finite, and the limit of $P(A_n)$ is indeterminate. Without this bound we cannot conclude c).