

Machine Learning

Notes taken by Runqiu Ye
Carnegie Mellon University

Spring 2025

Contents

1	Probability and Statistical Inference	3
1.1	Probability	3
1.2	Statistical Inference	4

1 Probability and Statistical Inference

1.1 Probability

Definition (Types of convergence). Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of random variables and X be another random variable. Let F_n be the CDF of X_n for each $n \in \mathbb{N}$ and F be the CDF of X .

1. X_n converges to X in probability and write $X_n \xrightarrow{P} X$ if for arbitrary $\varepsilon > 0$,

$$\mathbb{P}[|X_n - X| > \varepsilon] \rightarrow 0$$

as $n \rightarrow \infty$.

2. X_n converges to X in distribution and write $X_n \rightsquigarrow X$ if

$$\lim_{n \rightarrow \infty} F_n(t) = F(t)$$

for all t where F is continuous.

3. X_n converges to X in L^p if

$$\mathbb{E}[|X_n - X|^p] \rightarrow 0$$

as $n \rightarrow \infty$. In particular, say X_n converges to X in quadratic mean and write $X_n \xrightarrow{qm} X$ if X_n converges to X in L^2 .

4. X_n converges to X almost surely and write $X_n \xrightarrow{as} X$ if

$$\mathbb{P}\left[\lim_{n \rightarrow \infty} X_n = X\right] = 1.$$

Theorem. The following implication holds:

1. If X_n converges to X almost surely, then X_n converges to X in probability.
2. If X_n converges to X in L^p , then X_n converges to X in probability.

Proof. 1. If X_n converges to X almost surely, the set of points $O = \{\omega : \lim_{n \rightarrow \infty} X_n(\omega) \neq X(\omega)\}$ has measure zero. Now fix $\varepsilon > 0$ and consider the sequence of sets

$$A_n = \bigcup_{m=n}^{\infty} \{|X_m - X| > \varepsilon\}.$$

Note that $A_n \supset A_{n+1}$ for each $n \in \mathbb{N}$ and let $A_{\infty} = \bigcap_{n=1}^{\infty} A_n$. Now show $\mathbb{P}[A_{\infty}] = 0$. If $\omega \notin O$, then $\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)$ and thus $|X_n(\omega) - X(\omega)| < \varepsilon$ for some $n \in \mathbb{N}$. Therefore, $\omega \notin A_{\infty}$. It follows that $A_{\infty} \subset O$ and $\mathbb{P}[A_{\infty}] = 0$.

By monotone continuity, we have $\lim_{n \rightarrow \infty} \mathbb{P}[A_n] = \mathbb{P}[A_{\infty}]$. It follows that

$$\mathbb{P}[|X_n - X| > \varepsilon] \leq \mathbb{P}[A_n] \rightarrow 0$$

as $n \rightarrow \infty$. This completes the proof.

2. From Chebyshev's inequality, we have

$$\mathbb{P}[|X - X_n| > \varepsilon] \leq \frac{1}{\varepsilon^p} \mathbb{E}[|X - X_n|^p].$$

The claim follows directly. □

Theorem (Central Limit Theorem). Let X_1, \dots, X_n be i.i.d with mean μ and variance σ^2 . Let $S_n = \frac{1}{n} \sum_{i=1}^n X_i$. Then

$$Z_n = \frac{S_n - \mu}{\sqrt{\text{var } S_n}} = \frac{\sqrt{n}(S_n - \mu)}{\sigma} \rightsquigarrow Z,$$

where $Z \sim N(0, 1)$. In other words,

$$\lim_{n \rightarrow \infty} \mathbb{P}[Z_n < z] = \Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

Also write $Z_n \approx N(0, 1)$.

1.2 Statistical Inference

Definition. A $1 - \alpha$ interval for a parameter θ is an interval $C_n = (a, b)$ where $a = a(X_1, \dots, X_n)$ and $b = b(X_1, \dots, X_n)$ are functions of data such that

$$\mathbb{P}_\theta[\theta \in C_n] \geq 1 - \alpha \text{ for all } \theta \in \Theta.$$

In other word, (a, b) traps θ with probability $1 - \alpha$. **Warning!** C_n is random and θ is fixed.