

Lecture 2

Lecturer: Michael I. Jordan

Scribe: Ariel Kleiner

Lemma 1 (Fatou). If $X_n \xrightarrow{a.s.} X$ and $X_n \geq Y$ with $E[|Y|] < \infty$, then

$$\liminf_{n \rightarrow \infty} E[X_n] \geq E[X].$$

Theorem 2 (Monotone Convergence Theorem). If $0 \leq X_1 \leq X_2 \leq \dots$ and $X_n \xrightarrow{a.s.} X$, then

$$E[X_n] \rightarrow E[X].$$

Note that the Monotone Convergence Theorem can be proven from Fatou's Lemma.

Theorem 3 (Dominated Convergence Theorem). If $X_n \xrightarrow{a.s.} X$ and $|X_n| \leq Y$, $E[|Y|] < \infty$, then

$$E[X_n] \rightarrow E[X].$$

Theorem 4 (Weak Law of Large Numbers). If $X_i \stackrel{i.i.d.}{\sim} X$ and $E[|X|] < \infty$, then

$$\bar{X}_n \xrightarrow{P} E[X],$$

where $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$.

Theorem 5 (Strong Law of Large Numbers). If $X_i \stackrel{i.i.d.}{\sim} X$ and $E[|X|] < \infty$, then

$$\bar{X}_n \xrightarrow{a.s.} E[X].$$

Definition 6 (Empirical Distribution Function). Given n i.i.d. data points $X_i \stackrel{i.i.d.}{\sim} F$, the empirical distribution function is defined as

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n 1_{[X_i, \infty)}(x).$$

Note that $F_n(x) \xrightarrow{a.s.} F(x)$, for each x .

Theorem 7 (Glivenko-Cantelli). Given n i.i.d. data points $X_i \stackrel{i.i.d.}{\sim} F$,

$$P\{\sup_x |F_n(x) - F(x)| \rightarrow 0\} = 1$$

That is, the random variable $\sup_x |F_n(x) - F(x)|$ converges to 0, almost surely.

Theorem 8 (Central Limit Theorem). Given n i.i.d. random variables X_i from some distribution with mean μ and covariance Σ (which are assumed to exist),

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \Sigma).$$

The following theorem is a generalization of the Central Limit Theorem. It applies to non-i.i.d. (i.e., independent but not identically distributed) random variables as might be arranged in a triangular array as follows, where the random variables within each row are independent:

$$\begin{array}{ccc} Y_{11} & & \\ Y_{21} & Y_{22} & \\ Y_{31} & Y_{32} & Y_{33} \\ \vdots & & \end{array}$$

Theorem 9 (Lindeberg-Feller). For each n , let $Y_{n1}, Y_{n2}, \dots, Y_{nk_n}$ be independent random variables with finite variance such that $\sum_{i=1}^{k_n} \text{Var}(Y_{ni}) \rightarrow \Sigma$ and

$$\sum_{i=1}^{k_n} E [\|Y_{ni}\|^2 \mathbf{1}(\|Y_{ni}\| > \varepsilon)] \xrightarrow{n \rightarrow \infty} 0, \quad \forall \varepsilon > 0.$$

Then,

$$\sum_{i=1}^{k_n} (Y_{ni} - E[Y_{ni}]) \xrightarrow{d} N(0, \Sigma).$$

We now consider an example illustrating application of the Lindeberg-Feller theorem.

Example 10 (Permutation Tests). Consider $2n$ paired experimental units in which we observe the results of n treatment experiments X_{nj} and n control experiments W_{nj} . Let $Z_{nj} = X_{nj} - W_{nj}$. We would like to determine whether or not the treatment has had any effect. That is, are the Z_{nj} significantly non-zero? To test this, we condition on $|Z_{nj}|$. This conditioning effectively causes us to discard information regarding the magnitude of Z_{nj} and leaves us to consider only signs. Thus, under the null hypothesis H_0 , there are 2^n possible outcomes, all equally probable. We now consider the test statistic

$$Z_n = \frac{1}{n} \sum_{i=1}^n Z_{ni}$$

and show that, under H_0 ,

$$\frac{\sqrt{n}Z_n}{\sigma_n} \xrightarrow{d} N(0, 1),$$

where $\sigma_n^2 = \frac{1}{n} \sum_{i=1}^n Z_{ni}^2$, and we assume that

$$\max_j \frac{Z_{nj}^2}{\sum_i Z_{ni}^2} \rightarrow 0.$$

Proof. Let

$$Y_{nj} = \frac{Z_{nj}}{(\sum_i Z_{ni}^2)^{1/2}}.$$

Note that, under H_0 , $E[Y_{nj}] = 0$ because H_0 states that X_j and Y_j are identically distributed. Additionally,

we have $\sum_j \text{Var}(Y_{nj}) = 1$. Now observe that, $\forall \varepsilon > 0$,

$$\begin{aligned} \sum_j E[|Y_{nj}|^2 \mathbf{1}\{|Y_{nj}| > \varepsilon\}] &= \sum_j \frac{Z_{nj}^2}{\sum_i Z_{ni}^2} \mathbf{1}\left\{\frac{Z_{nj}^2}{\sum_i Z_{ni}^2} > \varepsilon^2\right\} \\ &\leq \left(\sum_j \frac{Z_{nj}^2}{\sum_i Z_{ni}^2}\right) \mathbf{1}\left\{\max_j \frac{Z_{nj}^2}{\sum_i Z_{ni}^2} > \varepsilon^2\right\} \\ &= \mathbf{1}\left\{\max_j \frac{Z_{nj}^2}{\sum_i Z_{ni}^2} > \varepsilon^2\right\} \\ &\rightarrow 0 \end{aligned}$$

where the equality in the first line follows from the definition of Y_{nj} and the fact that we are conditioning on the magnitudes of the Z_{nj} , thus rendering Z_{nj}^2 deterministic. The desired result now follows from application of the Lindeberg-Feller theorem. \square

We now move on to Chapter 3 in van der Vaart.

Theorem 11 (Delta Method, van der Vaart Theorem 3.1). *Let $\phi : D_\phi \subseteq \mathbb{R}^k \rightarrow \mathbb{R}^m$, differentiable at θ . Additionally, let T_n be random variables whose ranges lie in D_ϕ , and let $r_n \rightarrow \infty$. Then, given that $r_n(T_n - \theta) \xrightarrow{d} T$,*

- (i) $r_n(\phi(T_n) - \phi(\theta)) \xrightarrow{d} \phi'_\theta(T)$
- (ii) $r_n(\phi(T_n) - \phi(\theta)) - \phi'_\theta(r_n(T_n - \theta)) \xrightarrow{P} 0$

Proof. Given that $r_n(T_n - \theta) \xrightarrow{d} T$, it follows from Prohorov's Theorem that $r_n(T_n - \theta)$ is uniformly tight (UT). Differentiability implies that

$$\phi(\theta + h) - \phi(\theta) - \phi'_\theta(h) = o(\|h\|)$$

(from the definition of the derivative). Now consider $h = T_n - \theta$ and note that $T_n - \theta \xrightarrow{P} 0$ by UT and $r_n \rightarrow \infty$. By Lemma 2.12 in van der Vaart, it follows that

$$\phi(T_n) - \phi(\theta) - \phi'_\theta(T_n - \theta) = o_P(\|T_n - \theta\|).$$

Multiplying through by r_n , we have

$$r_n(\phi(T_n) - \phi(\theta) - \phi'_\theta(T_n - \theta)) = o_P(1),$$

thus proving (ii) above. Slutsky now implies that $r_n\phi'_\theta(T_n - \theta)$ and $r_n(\phi(T_n) - \phi(\theta))$ have the same weak limit. As a result, using the fact that ϕ'_θ is a linear operator and the Continuous Mapping Theorem, we have

$$r_n\phi'_\theta(T_n - \theta) = \phi'_\theta(r_n(T_n - \theta)) \xrightarrow{d} \phi'_\theta(T)$$

and so

$$r_n(\phi(T_n) - \phi(\theta)) \xrightarrow{d} \phi'_\theta(T).$$

\square

We now jump ahead to U -statistics.