STATS310A - Lecture 15

Persi Diaconis
Scribed by Michael Howes

11/11/21

Contents

1 Final comments on Stein's method

1.1 Dependency

In the Poisson case we used dependency graphs. There is a version where we only require that the disconnected X_i are "not too dependent" rather than requiring that disconnected X_i are independent. Consider our familiar set up $X_i \in \{0,1\}$, $\mathbb{P}(X_i = 1) = p_i$, $\mathbb{P}(X_i = 1, X_j = 1) = p_{ij}$, $W = \sum_{i \in I} X_i$, $\lambda = \sum_{i \in I} p_i = \mathbb{E}[W]$ and $\mathbb{P}_W(A) = \mathbb{P}(W \in A)$. We assume that for each $i \in I$ we have a subset $N_i \subseteq I$. Next define

$$\begin{aligned} b_1 &= \sum_{i \in I} \sum_{j \in N_i \setminus \{i\}} p_{ij}, \\ b_2 &= \sum_{i \in I} \sum_{j \in N_i} p_i p_j, \\ b_3 &= \sum_{i \in I} \mathbb{E}(|X_i - p_i| | X_j \in N_i^c). \end{aligned}$$

The quantity b_3 measures how independent X_i is of N_i^c.

Theorem 1. *With notation as above*

$$\|P_W - \mathcal{P}_\lambda\|_{TV} \leq 2(b_1 + b_2 + b_3),$$

where $\mathcal{P}_\lambda$ is the Poisson distribution with parameter λ. Also,

$$\left|\mathbb{P}(W = 0) - e^{-\lambda}\right| \leq (b_1 + b_2 + b_3)\frac{1 - e^{-\lambda}}{\lambda}.$$

This result can be found in "Poisson Approximation and the Chen-Stein Method" by Arratia, Goldstein and Gord. The article "A short survey of Stein's method" by Sourav Chatterjee was presented at ICM 2014 and is recommended reading.

1.2 Stein's method and normal approximation

Although we will not show it here, Stein's method can be used to prove the central limit theorem. See "Normal approximation by Stein's method" by Chen, Goldstein and Shao. To do this we again use a characteristic operator. In particular we need the theorem

Theorem 2. *A random variable Z is $\mathcal{N}(0,1)$ if and only if for all bounded C_1 functions $f : \mathbb{R} \to \mathbb{R}$,*

$$\mathbb{E}[Zf(Z) - f'(Z)] = 0.$$

In the interest of displaying a variety of a probabilistic techniques we will not use Stein's method to prove the CLT in this class.

2 Normal approximation and the CLT

2.1 Normal heuristic

Let X_i, $i \in I$ has mean 0 and variance σ^2. Suppose that the X_i are not too wild and not too dependent. Define $S_n = \sum_{i \in I} X_i$, then

$$\mathbb{P}\left(\frac{S_n}{\sqrt{\mathrm{Var}(S_n)}} \leq x\right) \approx \Phi(x),$$

where Φ is the cummulative distribution function of the standard normal distribution.

Example 1. Suppose we have a finite graph (such as an $n \times n$ grid). On each vertex i place independent uniform random variables $U_i \in [0,1]$. For each vertex define $X_i = 1$ if X_i is a local maximum and $X_i = 0$ else. Then $\mathbb{P}(X_i = 1) = 1/(d_i + 1)$ where d_i is the number of neighbours of X_i. Let $W = \sum_{i \in I} X_i$, then under some assumptions on the graph,

$$\mathbb{P}\left(\frac{W - \sum_{i \in I} \frac{1}{d_i+1}}{\sqrt{\sum_{i \in I} \frac{1}{d_i}\left(1 - \frac{1}{d_i}\right)}} \leq x\right) \approx \Phi(x).$$

Example 2. Pick $i \in [n] = \{1, 2, \ldots, n\}$ uniformly at random and let W be the number of 1's in the binary expansion of i. Then

$$W = \sum_{j=1}^{\log_2(n+1)} X_j,$$

where

$$X_j = \begin{cases} 1 & \text{if the } j^{th} \text{ bit of } i \text{ equals } 1, \\ 0 & \text{else.} \end{cases}$$

Then,

$$\mathbb{P}\left(\frac{W - \frac{\log_2(n)}{2}}{\sqrt{n/4}}\right) \to \Phi(x).$$

2.2 Lindeberg's condition

We will use Lindeberg's form of the central limit theorem and the proof will use the idea of coupling. First some notation

Definition 1. A *triangular array of random variables* is a collection of random variables

$$\begin{array}{lllll} X_{1,1}, & & & & \\ X_{2,1}, & X_{2,2} & & & \\ \vdots & \vdots & \ddots & & \\ X_{n,1} & X_{n,2} & \dots & X_{n,k_n} & \\ \vdots & \vdots & & \vdots & \ddots \end{array}$$

We will say that a triangular array has *independent rows* if for every n, the random variables $X_{n,1}, \dots, X_{n,k_n}$ are independent.

The idea of triangular arrays is to generalize the notion of i.i.d. sequences where now we allow for the distribution of X_i to depend on n.

Suppose we have a triangular array with independent rows such that $\mathbb{E}[X_{i,n}] = 0$ and $\sigma_{n,i}^2 = \mathrm{Var}(X_{n,i}) < \infty$. Then define $s_n^2 = \sum_{i=1}^{k_n} \sigma_{n,i}^2$. With this notation we have

Theorem 3 (Linderberg)**.** *Suppose that for all $\varepsilon > 0$ we have*

$$\frac{1}{s_n^2} \sum_{i=1}^{k_n} \int_{\{|X_{i,n}| > \varepsilon s_n\}} |X_{i,n}|^2 d\mathbb{P} \overset{n\to\infty}{\longrightarrow} 0. \tag{1}$$

Then

$$\mathbb{P}\left(\frac{S_n}{s_n} \le x\right) \to \Phi(x).$$

We will call equation (1) *Linderberg's condition.*

Example 3. Suppose we have an i.i.d. sequence $(X_i)_{i=1}^\infty$ with mean 0 and variance σ^2. Then

$$\mathbb{P}\left(\frac{S_n}{\sigma\sqrt{n}} \le x\right) \to \Phi(x).$$

To see why this holds consider the triangular array

$$\begin{array}{llll} X_1 & & & \\ X_1 & X_2 & & \\ X_1 & X_2 & X_3 & \\ \vdots & & & \ddots \end{array}$$

which clearly has independent rows. Note that $s_n^2 = n\sigma^2$ and so we can check Linderberg's condition (1)

$$\begin{aligned} \frac{1}{s_n^2} \sum_{i=1}^{n} \int_{\{|X_i| > \varepsilon s_n\}} |X_i|^2 d\mathbb{P} &= \frac{1}{n\sigma^2} \sum_{i=1}^{n} \int_{\{|X_i| > \varepsilon \sqrt{n}\sigma\}} |X_i|^2 d\mathbb{P} \\ &= \frac{1}{\sigma^2} \int_{\{|X_1| > \varepsilon \sqrt{n}\sigma\}} |X_1|^2 d\mathbb{P} \\ &\to 0, \end{aligned}$$

by the DCT.

Example 4. We can apply this to our ESP card guessing example. Consider cards labelled $1, \dots, n$. The cards are guessed one at a time and each time complete feedback is given. Define

$$X_{n,i} = \begin{cases} 1 & \text{if guess } i \text{ is correct,} \\ 0 & \text{else.} \end{cases}$$

Then $\mathbb{P}(X_{n,i} = 1) = \frac{1}{n-i+1}$ since we are given complete feedback and will never guess a card that we have seen already (and we don't have ESP!). Define

$$Y_i = X_{n,i} - \frac{1}{n-i+1}.$$

We then have $\mathbb{E}[Y_i] = 0$ and $\mathrm{Var}(Y_i) = \mathrm{Var}(X_i) = \frac{1}{n-i+1}\left(1 - \frac{1}{n-i+1}\right)$. Thus

$$\begin{aligned} s_n^2 &= \sum_{i=1}^n \frac{1}{n}\left(1 - \frac{1}{n}\right) \\ &= \log(n) - \gamma + \frac{\pi^2}{6} + O(1/n) \\ &\sim \log(n), \end{aligned}$$

where γ is Euler's constant (Persi directs interested students to "Euler's constant: Euler's work and modern developments" by Lagarias for the AMS Bulliten). Thus we can check Linderberg's condition (1). Note that

$$\frac{1}{s_n^2}\sum_{i=1}^n \int_{\{|Y_{i,n}|>\varepsilon s_n\}} |X_{i,n}|^2 d\mathbb{P} \approx \frac{1}{\log(n)}\sum_{i=1}^n \int_{\{|X_{i,n}|>\varepsilon\sqrt{\log(n)}\}} |Y_{i,n}|^2 d\mathbb{P},$$

and the right hand side is 0 for sufficiently large n because $|Y_{i,n}|$ is bounded by 1. Thus

$$\mathbb{P}\left(\frac{S_n}{s_n} \le x\right) \to \Phi(x).$$

2.3 Weak* convergence

Definition 2. Let Q_n, Q be probability measures on $\mathbb{R}^k$ with the Borel subsets. We will say that Q_n *converges weak** to Q and write $Q_n \Rightarrow Q$ if for all bounded and continuous $g : \mathbb{R}^k \to \mathbb{R}$,

$$\int_{\mathbb{R}^k} g dQ_n \longrightarrow \int_{\mathbb{R}^k} g dQ.$$

Theorem 4 (Portmanteau)**.** *Let Q_n, Q be probability measures on $\mathbb{R}^k$, then the following are equivalent:*

(a) The sequence of measures Q_n converges weak* to Q.

(b) For all $f : \mathbb{R}^k \to \mathbb{R}$ which are continuous with compact support, $\int f dQ_n \to \int f dQ$.

(c) For all $f : \mathbb{R}^k \to \mathbb{R}$ which are smooth with compact support, $\int f dQ_n \to \int f dQ$.

(d) If F_n is the cdf of Q_n and F if the cdf of Q (so that $F_n(x) = Q_n(\{y \in \mathbb{R}^k : y_i \le x_i\})$ and $F(x) = Q(\{y \in \mathbb{R}^k : y_i \le x_i\})$), then $F_n(x) \to F(x)$ at all points $x \in \mathbb{R}^k$ where F is continuous.

Proof. We will start the proof today and finish it next week. We will also only prove the case when $k = 1$. The general case is essentially the same but one has to work with the norm $\|x\| = \sqrt{\sum_{i=1}^k x_i^2}$ instead of the absolute value.

Note that we trivially have (a) implies (b) and (b) implies (c). We will first prove (b) implies (a). Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function such that $|f(x)| \le c$ for all x. Given $\varepsilon > 0$ there exists N such that

$$Q(\{|x| > N\}) \le \frac{\varepsilon}{4C}.$$

Define

$$\theta_N(x) = \begin{cases} 1 & \text{if } |x| \le N, \\ 0 & \text{if } |x| \ge N+1, \\ N+1-|x| & \text{if } |x| \in (N, N+1). \end{cases}$$

The function θ_N is continuous and has compact support. Thus

$$\begin{aligned} \underline{\lim} Q_n(\{|x| \le N+1\}) &\ge \underline{\lim} \int \theta_N dQ_n \\ &= \int \theta_n dQ \\ &\ge 1 - \frac{\varepsilon}{4C}. \end{aligned}$$

Thus

$$\begin{aligned} \overline{\lim} Q_n(\{|x| > N+1\}) &= 1 - \underline{\lim} Q_n(\{|x| \le N+1\}) \\ &= \frac{\varepsilon}{4C}. \end{aligned}$$

Now let $g = f\theta_{N+1}$. The function g is continuous with compact support. We have

$$\left|\int f dQ_n - \int f dQ\right| \le \left|\int f dQ_n - \int g dQ_n\right| + \left|\int g dQ_n - \int g dQ\right| + \left|\int g dQ - \int f dQ\right|.$$

The middle term goes to 0 since we have assumed (b). Furthermore, $f = g$ when $|x| \le N+1$. Thus

$$\begin{aligned} \overline{\lim}\left|\int f dQ_n - \int g dQ_n\right| &\le \overline{\lim} \int_{|x|>N+1} |f(x)| dQ_n \\ &\le C \overline{\lim} Q_n(\{|x| > N+1\}) \\ &\le \frac{\varepsilon}{4}. \end{aligned}$$

Likewise

$$\begin{aligned} \overline{\lim}\left|\int f dQ - \int g dQ\right| &= \left|\int f dQ - \int g dQ\right| \\ &\le \int_{|x|>N+1} |f(x)| dQ \\ &\le CQ(\{|x| > N+1\}) \\ &\le CQ(\{|x| > N\}) \\ &\le \frac{\varepsilon}{4}. \end{aligned}$$

Thus we have

$$\overline{\lim}\left|\int f dQ_n - \int f dQ\right| \le \frac{\varepsilon}{2}.$$

Letting $\varepsilon \to 0$ we see that $\int f dQ_n \to \int f dQ$ proving (b) implies (a).

We will now prove that (c) implies (b). Let f be continuous with compact support for $\eta > 0$ define

$$\rho_\eta(x) = Z(\eta) e^{-\frac{1}{1-\frac{x^2}{\eta^2}}} \delta_{[-\eta,\eta]}(x),$$

where $Z(\eta)$ is such that $\int_{-\eta}^{\eta} \rho_\eta(x)dx = 1$. One can prove by induction that ρ_η is smooth. The function ρ_η clearly has compact support.

Now let $\varepsilon > 0$ be given. Since f is continuous with compact support, f is uniformly continuous. Thus there exists $\eta = \eta(\varepsilon)$ such that for all $x.y \in \mathbb{R}$, $|x-y| \le \eta$ implies that $|f(x) - f(y)| \le \varepsilon$. Define

$$f^\varepsilon(x) = \int_{\mathbb{R}} f(y)\rho_\eta(x-y)dy = \int_{-\eta}^{\eta} f(x-y)\rho_\eta(y)dy.$$

The function f^ε is also smooth and f^ε has compact support since f and ρ_η both have compact support. Furthermore

$$\begin{aligned} |f(x) - f^\varepsilon(x)| &= \left| \int_{-\eta}^{\eta} f(x)\rho_\eta(y)dy - \int_{-\eta}^{\eta} f(x-y)\rho_\eta(y)dy \right| \\ &\le \int_{-\eta}^{\eta} |f(x) - f(x-y)|\rho_\eta(y)dy \\ &\le \int_{-\eta}^{\eta} \varepsilon\rho_\eta(y)dy \\ &= \varepsilon. \end{aligned}$$

Since Q_n and Q are both probability measures we thus have

$$\left| \int f dQ_n - \int f^\varepsilon dQ_n \right| \le \varepsilon \quad \text{and} \quad \left| \int f dQ - \int f^\varepsilon dQ \right| \le \varepsilon.$$

Thus, by a similiar argument to before we have

$$\begin{aligned} &\overline{\lim} \left| \int f dQ_n - \int f dQ \right| \\ &\le \overline{\lim} \left| \int f dQ_n - \int f^\varepsilon dQ_n \right| + \overline{\lim} \left| \int f^\varepsilon dQ_n - \int f^\varepsilon dQ \right| + \overline{\lim} \left| \int f^\varepsilon dQ - \int f dQ \right| \\ &\le \varepsilon + 0 + \varepsilon \\ &= 2\varepsilon. \end{aligned}$$

Letting $\varepsilon \to 0$ we see that $\int f dQ_n \to \int f dQ$. Thus (c) implies (b). $\square$

We will prove that (d) is also equivalent next lecture.