Lecture 3

Extra remarks on Lindeberg method: non-linear functions, non-iid X_i.
(See previous notes for some details.)

Thm: (Lindeberg CLT) $(X_{i,n})_{1\le i\le n}$ independent, $\mathbb{E}X_{i,n}=0$, $\mathbb{E}X_{i,n}^2<\infty$

$$V_n^2 := \operatorname{Var}\sum_{i=1}^n X_{i,n} = \sum_{i=1}^n \mathbb{E}X_{i,n}^2 \qquad (\text{think: } V_n = \Theta(\sqrt{n}) \text{ usually})$$

Suppose Lindeberg condition holds: $\forall \varepsilon>0$,

$$\frac{1}{V_n^2}\sum_{i=1}^n \mathbb{E}X_{i,n}^2 \underbrace{\mathbb{1}\{|X_{i,n}|>\varepsilon V_n\}} \longrightarrow 0.$$

without this, just average variance.
with it, "contribution to avg. variance of large deviations".

~~Let~~ Then, $\hat{S}_n := \frac{1}{V_n}\sum_{i=1}^n X_{i,n} \Longrightarrow \mathcal{N}(0,1)$

$\hookrightarrow$ (if $\mathbb{E}X_{i,n}^2 \equiv 1$, $V_n=\sqrt{n}$)

Ex: $Y_{i,n}$ iid with $\mathbb{E}Y_{i,n}=0$, $\mathbb{E}Y_{i,n}^2=1$, $X_{i,n}=c_{i,n}Y_{i,n}$

$V_n^2 = \sum_{i=1}^n c_{i,n}^2$

$$(\text{LHS}) = \frac{1}{\sum_i c_{i,n}^2}\sum_i c_{i,n}^2\, \mathbb{E}Y_{i,n}^2 \mathbb{1}\left\{|Y_{i,n}|>\varepsilon\frac{\sqrt{\sum_j c_{j,n}^2}}{|c_{i,n}|}\right\}$$

$$\le \cancel{\frac{1}{\sum_i c_{i,n}^2}}\cancel{\left(\sum_i c_{i,n}^2\right)} \mathbb{E}Y_{1,1}^2 \mathbb{1}\left\{|Y_{1,1}|>\varepsilon\frac{\sqrt{\sum_j c_{j,n}^2}}{\max_i |c_{i,n}|}\right\}$$

$\rightsquigarrow$ enough to have $\frac{\|c_n\|_2}{\|c_n\|_\infty} \xrightarrow[n\to\infty]{} \infty$ (i.e., no one term contributes too much)

Ex: $X_{i,n} \overset{\text{ind}}{\sim} \operatorname{Ber}(p_n)$, $\overline{X_{i,n}} := X_{i,n}-\mathbb{E}X_{i,n} = X_{i,n}-p_n$

When does $\frac{1}{V_n^2}\sum_i \overline{X_{i,n}} \Longrightarrow \mathcal{N}(0,1)$?

Intuition: so long as p_n not too small.

$$V_n^2 = n \cdot \mathrm{Var}\, X_{1,n} = n \cdot p_n (1-p_n).$$

$$(\mathrm{LHS}) = \frac{1}{n p_n (1-p_n)} \sum_{i=1}^{n} \mathbb{E}\,(X_{i,n}-p_n)^2 \,\mathbb{1}\left\{|X_{i,n}-p_n| > \varepsilon\sqrt{n p_n (1-p_n)}\right\}$$

Say $p_n \in [0, \frac{1}{2}]$ (otherwise, consider $1 - X_{i,n}$). Two cases:

(1) $n p_n \longrightarrow \infty$: then $(\mathrm{LHS}) = 0$ for n suff. large.

(2) $n p_n$ bounded : then, for ε suff. small,

$$(\mathrm{LHS}) = \frac{1}{\cancel{n p_n (1-p_n)}} \cdot \cancel{n} \cdot (1-p_n)^{\cancel{2}} \cdot \cancel{p_n} = 1 - p_n \longrightarrow 1.$$

Thm: If $n p_n \longrightarrow \infty$, then $\frac{1}{V_n^2} \sum_i X_{i,n} \Longrightarrow \mathcal{N}(0,1)$.

"even very rare events obey CLT".

Q: What about $p_n = \frac{\lambda}{n}$?

Rk: $\mathrm{Law}(S_n) = \mathrm{Bin}(n, \frac{\lambda}{n})$.

$$S_n := \sum_{i=1}^{n} X_{i,n} \longrightarrow \begin{cases} \mathbb{E} S_n = n \cdot \frac{\lambda}{n} = \lambda \\ \mathrm{Var}\, S_n = \cancel{n} \cdot \frac{\lambda}{\cancel{n}} \left(1 - \frac{\lambda}{n}\right) \longrightarrow \lambda \end{cases}$$

Def: $P \sim \mathrm{Pois}(\lambda)$ if $P \in \{0, 1, 2, \ldots\}$, $\mathbb{P}[P = k] = \frac{\lambda^k}{k!} \exp(-\lambda)$.

Thm: (PLT) If $X_{i,n} \overset{\text{ind}}{\sim} \mathrm{Ber}\left(\frac{\lambda}{n}\right)$, then $S_n \Longrightarrow \mathrm{Pois}(\lambda)$.

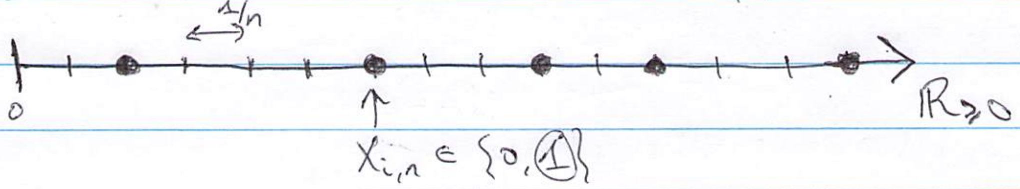
Rk: No normalization!

Pf: Enough to show $\mathbb{P}[S_n = k] \longrightarrow \mathbb{P}[P = k] = \frac{\lambda^k}{k!} \exp(-\lambda)$

$$\mathbb{P}[S_n = k] = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} = \underbrace{\cancel{\frac{n(n-1)\cdots(n-k+1)}{n \cdot n \cdots n}}}_{\to 1} \frac{\lambda^k}{k!} \underbrace{\cancel{\left(1 - \frac{\lambda}{n}\right)^{-k}}}_{\to 1} \left(1 - \frac{\lambda}{n}\right)^n \longrightarrow \frac{\lambda^k}{k!} \exp(-\lambda).$$

"Poisson Point Process".

Preview of PPP: Consider $X_{i,n} \overset{\text{ind}}{\sim} \text{Ber}\left(\frac{\lambda}{n}\right)$, $i = 1, 2, \ldots, n, \ldots$



For each n, defines μ_n prob. measure over subsets of $\mathbb{R}_{\geq 0}$.
Will see later: in some sense, "$\mu_n \longrightarrow \mu$", for $\mu = \text{PPP}(\lambda)$

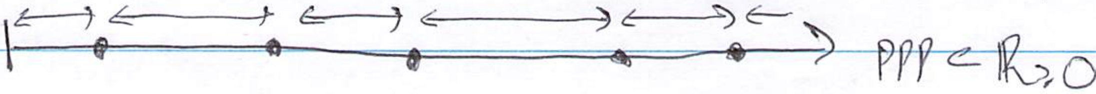
Previous analysis: $A \sim \text{PPP}(\lambda) \Rightarrow \text{Law}(|A \cap [0,1]|) = \text{Pois}(\lambda)$.
Similarly, $\text{Law}(|A \cap [x,y]|) = \text{Pois}(\lambda(y-x))$
$\text{Law}(|A \cap B|) = \text{Pois}(\lambda \cdot \text{Lebesgue measure of } B)$.

For now: consider spacings of points, e.g. first point:
$T_n := \min\{i : X_{i,n} = 1\}$, $\hat{T}_n := \frac{T_n}{n}$ (= position of leftmost point of PPP.)

Thm: $\hat{T}_n \Rightarrow \text{Exp}(\lambda)$ (**Def:** $T \sim \text{Exp}(\lambda)$ has density $\mathbb{1}\{x \geq 0\} \lambda e^{-\lambda x}$)

Pf:
$$\mathbb{P}[\hat{T}_n \leq t] = 1 - \mathbb{P}[\hat{T}_n > t] = 1 - \mathbb{P}[T_n > tn]$$
$$= 1 - \left(1 - \frac{\lambda}{n}\right)^{\lfloor tn \rfloor} \longrightarrow 1 - \exp(-\lambda t).$$

Exc: $T_n^{(1)} := T_n$, $T_n^{(k)} := (k^{\text{th}} \text{ smallest } i : X_{i,n} = 1) - ((k-1)^{\text{th}} \text{ smallest})$
$\hat{T}_n^{(k)} := \frac{1}{n} T_n^{(k)} \rightsquigarrow T_n^{(k)} \Rightarrow \text{Exp}(\lambda)$,
$(T_n^{(1)}, \ldots, T_n^{(k)}) \Rightarrow (k \text{ independent Exp}(\lambda)\text{'s})$.



Standard modeling choice for "waiting" or "arrival" times.