5/4/22

Central Limit Theorem

Identically distributed.

$X_1, X_2, \cdots$: I.I.D. $\qquad \mu = \mathbb{E}X_1, \quad \sigma^2 = Var(X_1) < \infty$

Indep. $\qquad$ Martingale

$S_n = X_1 + \cdots + X_n$

$$\frac{S_n - \mathbb{E}S_n}{\sqrt{Var(S_n)}} \Rightarrow N(0,1) \quad \text{as} \quad n \to \infty$$

$\mathbb{E}S_n = \mathbb{E}[X_1 + \cdots + X_n] = \mathbb{E}X_1 + \cdots + \mathbb{E}X_n = n \cdot \mu$

$Var(S_n) = Var(X_1 + \cdots + X_n) = Var(X_1) + \cdots + Var(X_n)$ (Indep.)

$= n \cdot \sigma^2$

$$\frac{(X_1 + X_2 + \cdots + X_n) - n \cdot \mu}{\sqrt{n} \cdot \sigma = \sqrt{n \cdot \sigma^2}} = \frac{\left(\frac{X_1 + \cdots + X_n}{n}\right) - \mu}{\sigma / \sqrt{n}} \Rightarrow N(0,1)$$

convergence in dist.

$$\lim_{n \to \infty} P\left(\frac{\left(\frac{X_1 + \cdots + X_n}{n}\right) - \mu}{\sigma / \sqrt{n}} \le x \right) = \Phi(x)$$

$P(|\frac{S_n}{n} - \mu| \ge \varepsilon) \to 0$

for all x.

Def $\quad X_n \Rightarrow X$ in distribution

iff $\quad P(X_n \le x) \to P(X \le x)$ for all x at which

$F_X(x)$ is continuous.

$\left\{ \begin{array}{l} X_n \to X \text{ in prob.} \\ X_n \to X \text{ almost surely.} \end{array} \right.$

Idea of Proof

Fact $\quad Z_n$: RV $\quad M_{Z_n}(t) \to M_Z(t)$ for $-\varepsilon < t < \varepsilon$.

Implies $\quad Z_n \Rightarrow Z$ in distribution.

$$\underset{\sim}{Y_n} = \frac{X_n - \mu}{\sigma} \qquad \frac{S_n - \mu \cdot n}{\sigma \cdot \sqrt{n}} = \underbrace{\frac{Y_1 + \cdots + Y_n}{\sqrt{n}}}_{= Z_n}$$

WANT $\Rightarrow Z \sim N(0,1)$

$$M_{Z_n}(t) = \mathbb{E}\left[e^{t \cdot Z_n}\right]$$

$$= \mathbb{E}\left[e^{t \cdot \frac{Y_1 + \cdots + Y_n}{\sqrt{n}}}\right] = \mathbb{E}\left[e^{\frac{t}{\sqrt{n}} \cdot Y_1} \cdot e^{\frac{t}{\sqrt{n}} Y_2} \cdots e^{\frac{t}{\sqrt{n}} \cdot Y_n}\right]$$

$$= \underset{\sim}{\mathbb{E}\left[e^{\frac{t}{\sqrt{n}} Y_1}\right]} \cdot \mathbb{E}\left[e^{\frac{t}{\sqrt{n}} Y_2}\right] \cdots \mathbb{E}\left[e^{\frac{t}{\sqrt{n}} Y_n}\right]$$

$$= \left(M_{Y_1}\left(\frac{t}{\sqrt{n}}\right)\right)^n = \left(1 + \underbrace{\mathbb{E} Y_1}_{0} \cdot \frac{t}{\sqrt{n}} + \frac{1}{2} \underbrace{\mathbb{E} Y_1^2}_{1} \cdot \left(\frac{t}{\sqrt{n}}\right)^2 + \cdots\right)^n$$

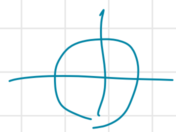
$\mathrm{Var}(Y_1) = 1$

$$f(x) = f(0) + f'(0) \cdot x + \frac{1}{2} f''(0) \cdot x^2 + \cdots$$

$$= \left(1 + \frac{1}{2} \cdot \frac{t^2}{n} + \cdots\right)^n \longrightarrow e^{-\frac{1}{2} t^2}$$

$$\left(1 + \frac{a}{n} + \cdots\right)^n \longrightarrow e^a$$

$$\phi_X(t) := \mathbb{E}\left[e^{itX}\right] = M_X(it)$$



$$\phi_Z(t) = e^{-\frac{t^2}{2}}$$

$$M_X(t) = \mathbb{E}\left[\underbrace{e^{tX}}_{\text{Could be too large}}\right] \qquad \phi(t) = \mathbb{E}\left[e^{itX}\right] \qquad \left|e^{itX}\right| \le 1$$

Example 15. An astronomer is interested in measuring the distance, in light-years, from his observatory to a distant star. Although the astronomer has a measuring technique, he knows that because of changing atmospheric conditions and normal error, each time a measurement is made, it will not yield the exact distance, but merely an estimate. As a result, the astronomer plans to make a series of measurements and then use the average value of these measurements as his estimated value of the actual distance. If the astronomer believes that the values of the measurements are independent and identically distributed random variables having a common mean d (the actual distance) and a common variance of 4 (light-years), how many measurements need he make to be reasonably sure that his estimated distance is accurate to within $\pm .5$ light-year?

95%

$\sigma = 2$.

$X_1, X_2, \cdots, X_n$: IID. $\quad S_n = \frac{X_1 + \cdots + X_n}{n}$

$\mathbb{E}X_1 = d, \quad Var(X_1) = \sigma^2 = 4.$

Find n s.t. $P\left(\left| \frac{X_1 + \cdots + X_n}{n} - d \right| \leq 0.5 \right) \geq 0.95$

$\sigma/\sqrt{n} \qquad \sigma/\sqrt{n}$

$$\mathbb{P}\left(|Z| \leq \frac{0.5}{\sigma/\sqrt{n}}\right) = \mathbb{P}\left(|Z| \leq \frac{\sqrt{n}}{4}\right) = 2 \cdot \Phi\left(\frac{\sqrt{n}}{4}\right) - 1 \geq 0.95$$

$$\Rightarrow \Phi\left(\frac{\sqrt{n}}{4}\right) \geq \frac{1 + 0.95}{2}.$$

Example 16. The number of students who enroll in a psychology course is a Poisson random variable with mean 100. The professor in charge of the course has decided that if the number enrolling is 120 or more, he will teach the course in two separate sections, whereas if fewer than 120 students enroll, he will teach all of the students together in a single section. What is the probability that the professor will have to teach two sections?