11/1

Convergence in distribution and Central Limit Theorem

Definition: Let $X_1, X_2, \ldots$ be a sequence of random variables with c.d.f. $F_1, F_2, \ldots$, and let X be a random variable with c.d.f F.

We say X_n converges in distribution to X if $\boxed{\lim_{n\to\infty} F_n(x) = F(x)}$ ✓ at every point at which F is continuous.

NOTATION: $X_n \xrightarrow{d} X$

Theorem: Let F_n be a sequence of c.d.f s with corresponding moment generating function M_n (i.e., X_n has m.g.f M_n)

Let F be the c.d.f with m.g.f. M.

If $M_n(t) \to M(t)$ for t in an open interval.

containg zero,
THEN $F_n(x) \to F(x)$ at all points of continuity of F.

Hence $X_n \xrightarrow{d} X$.

Aside: If X has mgf $M_X(t)$

and $\boxed{Y = a + bX}$

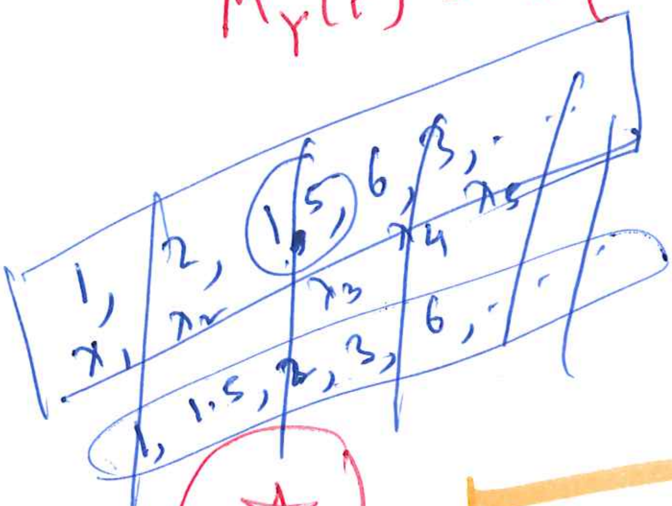
$$M_Y(t) \stackrel{def}{=} E\left(e^{tY}\right) = E\left(e^{t(a+bX)}\right)$$

$$= E\left(e^{ta}\, e^{tbX}\right)$$

$$\cancel{e^{ta}\, E\left(e^{t(bX)}\right)}$$

$$= e^{ta}\, E\left(e^{(tb)X}\right)$$

$$= e^{ta}\, M_X(bt)$$



$$\boxed{M_Y(t) = e^{ta}\, M_X(bt)}$$

Example: • Let $\lambda_1, \lambda_2, \ldots$ be an increasing (in tren sequence with $\lambda_n \to \infty$ as $n \to \infty$.

• Let $\{X_n\}$ be a sequence of Poisson random variable with mean (expectation) $= \lambda_n$.

- So, $E(X_n) = \lambda_n$
  $Var(X_n) = \lambda_n$ $\Big\}$

- $$Z_n = \frac{X_n - E(X_n)}{\sqrt{Var(X_n)}} = \frac{X_n - \lambda_n}{\sqrt{\lambda_n}}$$

NOT really a std normal distribution.

$n = 1, 2, 3, \ldots$

<u>Goal:</u> $Z_n \xrightarrow{d}$ Std. Normal

as $n \to \infty$

<u>To show:</u> $\boxed{\text{Plan}}$ We will find the m.g.f of Z_n : call them $M_{Z_n}(t)$

and then we show that

$M_{Z_n}(t) \to M(t)$ ← m.g.f of std. normal distribution.

$M_{X_n}(t) = e^{\lambda_n(e^t - 1)}$

$X_n \sim Poisson(\lambda_n)$

Using the previous result (4)

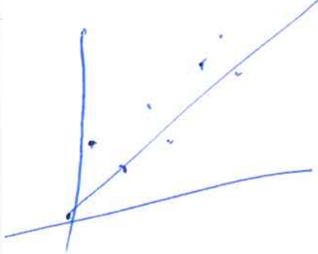
$$M_{Z_n}(t) = e^{-\sqrt{\lambda_n}\, t} \cdot M_{X_n}\left(\frac{t}{\sqrt{\lambda_n}}\right)$$

$$\boxed{M_{Z_n}(t) = e^{-t\sqrt{\lambda_n}} \exp\left(\lambda_n\left(e^{\frac{t}{\sqrt{\lambda_n}}} - 1\right)\right)}$$

If $Z \sim N(0,1)$ (Std. Normal dbn)

$$M_Z(t) = e^{t^2/2}$$

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$



$$\boxed{\log M_{Z_n}(t)} = -t\sqrt{\lambda_n} + \lambda_n\left(e^{\frac{t}{\sqrt{\lambda_n}}} - 1\right)$$

$$= -\cancel{t\sqrt{\lambda_n}} + \lambda_n\left(\left(\cancel{1} + \cancel{\frac{t}{\sqrt{\lambda_n}}} + \frac{t^2}{2\lambda_n} + \frac{t^3}{6(\sqrt{\lambda_n})^3} + \frac{t^4}{24\lambda_n^2} + \cdots\right) - \cancel{1}\right)$$

$$= \boxed{\frac{t^2}{2} + \frac{t^3}{6\sqrt{\lambda_n}} + \frac{t^4}{24\lambda_n} + \cdots}$$

$$\lim_{n\to\infty} \log M_{Z_n}(t) = \frac{t^2}{2} \quad \left(\text{as } \lambda_n \to \infty \text{ when } n \to \infty\right)$$

$$\Rightarrow \log\left(\lim_{n\to\infty} M_{Z_n}(t)\right) = \frac{t^2}{2}$$

$$\Rightarrow \boxed{\lim_{n\to\infty} M_{Z_n}(t)} = e^{t^2/2} = \boxed{M_Z(t)}$$

m.g.f. of std. normal distribution.

$$\boxed{\frac{t^2}{2} + \frac{t^3}{6} \cdot O\left(\frac{1}{\sqrt{\lambda_n}}\right)}$$

By the previous theorem,

$$Z_n \xrightarrow{d} Z \text{ as } n \to \infty$$

* • $X_1, X_2, \ldots$ is a sequence of independent random variables with mean: μ (expectation) and variance: σ^2

• Construct $S_n = \sum_{i=1}^{n} X_i$

$$\boxed{E(S_n) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} \mu = n\mu.}$$

$$\boxed{Var(S_n) \underset{\uparrow}{=} \sum_{i=1}^{n} Var(X_i) = \sum_{i=1}^{n} \sigma^2 = n\sigma^2}$$

($\because X_i$'s are independent)

• Construct: $Z_n = \dfrac{S_n - E(S_n)}{\sqrt{Var(S_n)}}$

i.e., $\boxed{Z_n = \dfrac{S_n - n\mu}{\sqrt{n}\cdot\sigma}}$

Central Limit Theorem

Theorem: Let $X_1, X_2, \ldots$ be a sequence of independent random variables having mean 0 and variance σ^2 (All the X_i's are from same distribution)

Let $S_n = \sum_{i=1}^{n} X_i$

Then $\lim_{n\to\infty} P\left(\frac{S_n - 0}{\sqrt{n}\,\sigma} \le x\right) = \Phi(x)$

$\left(\text{c.d.f of } \frac{S_n - 0}{\sqrt{n}\,\sigma}\right)$

c.d.f of std. normal dbn

$(\text{for, } -\infty < x < \infty)$

Proof:

$$Z_n = \frac{S_n - 0}{\sqrt{n}\,\sigma} = \frac{\sum_{i=1}^{n} X_i}{\sqrt{n}\,\sigma}$$

(np)

$$M_{S_n}(t) = E\left(e^{tS_n}\right) = E\left(e^{t\sum_{i=1}^{n} X_i}\right)$$

$(\because X_i\text{'s are independent})$

$$= \prod_{i=1}^{n} E\left(e^{tX_i}\right)$$

$$= \prod_{i=1}^{n} M(t)$$

$$= (M(t))^n$$

$\because X_i$'s are from SAME distribution

$M(t) = E(e^{tX_i})$ for every i.

$$M_{Z_n}(t) = \left[M\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n$$ (by (*) on page 2)

$$M\left(\frac{t}{\sigma\sqrt{n}}\right) = M(0) + \frac{t}{\sigma\sqrt{n}} M'(0) + \frac{1}{2}\left(\frac{t}{\sigma\sqrt{n}}\right)^2 M''(0) + \frac{1}{3!}\left(\frac{t}{\sigma\sqrt{n}}\right)^3 M'''(0) + \cdots$$

Taylor expansion around 0

$\cancel{M'(0) = E(X_i)}$

given

$$M(0) = 1$$
$$M'(0) = E(X_i) = 0$$
$$M''(0) = E(X_i^2) = E((X_i - 0)^2) = Var(X_i) = \sigma^2$$

For every i

$$M\left(\frac{t}{\sigma\sqrt{n}}\right) = 1 + 0 + \frac{1}{2}\left(\frac{t}{\sigma\sqrt{n}}\right)^2 \cdot \sigma^2 + \frac{1}{3!}\left(\frac{t}{\sigma\sqrt{n}}\right)^3 M'''(0) + \cdots$$

So, $$M_{Z_n}(t) = \left(M\left(\frac{t}{\sigma\sqrt{n}}\right)\right)^n$$

$$= \left(1 + \frac{1}{2}\left(\frac{t}{\sigma\sqrt{n}}\right)^2 \sigma^2 + \frac{1}{3!}\left(\frac{t}{\sigma\sqrt{n}}\right)^3 m'''(0) + \cdots \right)^n$$

As $n \to \infty$

$$\lim_{n\to\infty} M_{Z_n}(t) = \lim_{n\to\infty}\left(1 + \frac{1}{2}\left(\frac{t}{\sigma\sqrt{n}}\right)^2\sigma^2 + \underbrace{\boxed{\frac{1}{3!}\left(\frac{t}{\sigma\sqrt{n}}\right)^3 m'''(0) + \cdots}}_{\to 0}\right)^n$$

$$= \lim_{n\to\infty}\left(1 + \frac{t^2/2}{n}\right)^n$$

$$\boxed{\lim_{n\to\infty}\left(1 + \frac{a_n}{n}\right)^n = e^a}$$

$$= e^{t^2/2}$$

$\parallel$

m.g.f for Std. normal distribution.

Hence $\lim_{n\to\infty} M_{Z_n}(t) = e^{t^2/2}$

. Centeral limit theorem proved!

16] Suppose that $X_1, X_2, \ldots, X_{20}$ are <u>independent</u> random variables with density function $f(x) = 2x, \quad 0 \le x \le 1$.

Let $S = X_1 + X_2 + \cdots + X_{20}$.

Use the central limit theorem to approximate $P(S \le 10)$.

Solution: $E(S) = E\left(\sum_{i=1}^{20} X_i\right) = 20\,E(X)$

(where $E(X)$ is the expectation of any X_i, $i = 1, 2, \ldots, 20$).

$$Var(S) = Var\left(\sum_{i=1}^{20} X_i\right) = \sum_{i=1}^{20} Var(X_i) = 20\,Var(X)$$

(X_i's are independent)

$$E(X) = \int_0^1 x \cdot f(x)\,dx = \int_0^1 x \cdot 2x\,dx = \frac{2}{3}$$

$$E(X^2) = \int_0^1 x^2 \cdot f(x)\,dx = \int_0^1 x^2 \cdot 2x\,dx = \frac{2}{4}$$

So, $Var(\bar{X}) = E(\bar{X}^2) - (E(\bar{X}))^2$

$$= \frac{2}{4} - \left(\frac{2}{3}\right)^2 = \frac{1}{18}$$

So, by $\underline{\underline{\text{CLT}}}$

$$\frac{S - \overbrace{\left(\frac{2}{3}\right)20}^{20E(\bar{X})}}{\underbrace{\sqrt{20}\sqrt{\frac{1}{18}}}_{\sqrt{20 Var(\bar{X})}}} = \frac{S - 13.33}{\sqrt{1.11}} \sim N(0,1)$$

So, $P(S \leq 10)$

$$= P\left(\frac{S - 13.33}{\sqrt{1.11}} \leq \frac{10 - 13.33}{\sqrt{1.11}}\right)$$

$$\simeq P(Z \leq -3.16) = \Phi(-3.16)$$

$$= \underline{\underline{0.0008}}$$

~~[14] Suppose that a measurement~~

18] Suppose that a company ships packages that are variable in weight with an average weight of 15 lb and standard deviation of 10.

Assuming that the packages come from a large number of different customers (so that it is reasonable to model their weights as independent random variables), find the probability that 100 packages will have a total weight exceeding 1700 lbs.

Solution: X_i = i^{th} package weight (in lbs)

(Total weight) $\rightarrow S = \sum_{i=1}^{100} X_i$

We want to find $P(S > 1700)$.

$$E(S) = \sum_{i=1}^{100} E(X_i) = \sum_{i=1}^{100} 15 = 15 * 100$$

$$= 1500$$

X_i's are independent.

$$Var(S) = \sum_{i=1}^{100} Var(X_i) = \sum_{i=1}^{100} 10^2 = 10^2 * 100$$

$$= 10000$$

Std. deviation $= 10$
So, variance $= 10^2$

By CLT:

$$\frac{S - E(S)}{\sqrt{Var(S)}} = \frac{S - 1500}{\sqrt{10000}} \sim N(0,1)$$

$$P(S > 1700) = 1 - P(S \le 1700)$$

$$= 1 - P\left(\frac{S-1500}{\sqrt{10000}} \le \frac{1700-1500}{\sqrt{10000}}\right)$$

$\sim N(0,1)$

$$= 1 - P(Z \le 2)$$

$$= 1 - 0.9772$$

$$= \underline{0.0228}$$