Lecture 19 Dec 2, 2014

for r.v.'s we care about... i.e. where mgf exists

Recall mgf: $M_X(t) := E[e^{tX}] = 1 + tE[X] + \frac{t^2 E[X^2]}{2!} + \frac{t^3 E[X^3]}{3!} + \ldots = \sum_{i=0}^{\infty} \frac{t^i E[X^i]}{i!}$

① $M_X(t) = M_Y(t) \Leftrightarrow X \stackrel{d}{=} Y$ fingerprint.. just like CDF

② $M_{X+Y}(t) = M_X(t) M_Y(t)$ if X and Y are indep.

③ $M_X'(0) = E[X]$, $M_X''(0) = E[X^2]$, ..., $M_X^{(k)}(0) = E[X^k]$

Recall also $Z \sim N(0,1) \Leftrightarrow M_Z(t) = e^{t^2/2}$

Problem setup: $X_1, \ldots, X_n \stackrel{iid}{\sim}$ some PMF (or PDF) if discrete (or if cont)

indep. identically distr.

The PMF (or PDF) is unknown. But we know $E[X]$ and $E[X^2]$ exist (equiv., $\mu \in \mathbb{R}$, $\sigma^2 \in \mathbb{R}$)

$\bar{X}$ is thing of interest $\bar{X} := \frac{X_1 + \ldots + X_n}{n}$

$E[\bar{X}] = \mu$, $Var[\bar{X}] = \frac{\sigma^2}{n} \Rightarrow SE[\bar{X}] = \frac{\sigma}{\sqrt{n}}$

(I) $\Rightarrow C_n := \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ has $E(C_n) = 0$ and $SE(C_n) = 1$ always!

equivalently:

$T_n := X_1 + \ldots + X_n$ has $E(T_n) = n\mu$, $Var[T_n] = n\sigma^2 \Rightarrow SE(T_n) = \sqrt{n} \cdot \sigma$

(II) $\Rightarrow C_n := \frac{T_n - n\mu}{\sqrt{n}\,\sigma}$

since $X_1, \ldots, X_n \stackrel{iid}{\sim} f_1 \Rightarrow Z_1, \ldots, Z_n \stackrel{iid}{\sim} f_2$ as well

equivalently

$\bar{Z} = \frac{Z_1 + \ldots + Z_n}{n}$ where $Z_1 := \frac{X_1 - \mu}{\sigma}$, $Z_2 := \frac{X_2 - \mu}{\sigma}$, ..., $Z_n = \frac{X_n - \mu}{\sigma}$

(III) $\Rightarrow C_n := \sqrt{n}\, \bar{Z} = \frac{Z_1}{\sqrt{n}} + \frac{Z_2}{\sqrt{n}} + \ldots + \frac{Z_n}{\sqrt{n}}$ and $\frac{Z_1}{\sqrt{n}}, \ldots, \frac{Z_n}{\sqrt{n}} \stackrel{iid}{\sim} f_3$

Three equivalent (via algebra) ways to define C_n

because iid r.v.s easier to work with

We care about C_n as $n \to \infty$. what is $C_n \sim$ as? Use 3rd def (discuss me).

$Y = aX \ni M_Y(t) = E[e^{taX}] = M_X(at)$

let $M_Z(t)$ be mgf for Z

$e^x = \sum_{i=0}^{\infty} \frac{x^i}{i!}$ $\leftarrow$ famous!

$\Rightarrow$ What is $M_{\frac{Z}{\sqrt{n}}}(t)$? $= M_Z\left(\frac{t}{\sqrt{n}}\right) = E\left[e^{\frac{t}{\sqrt{n}}Z}\right] = E\left[1 + \frac{t}{\sqrt{n}}Z + \frac{t^2}{2n}Z^2 + \frac{t^3 Z^3}{3!\, n^{3/2}} + \dots\right]$

scale

What is $M_{C_n}(t) = M_{\frac{Z_1}{\sqrt{n}}}(t) \cdot M_{\frac{Z_2}{\sqrt{n}}}(t) \cdot \ldots \cdot M_{\frac{Z_n}{\sqrt{n}}}(t)$ by ind of $\frac{Z_i}{\sqrt{n}}$

$= M_{Z_1}\left(\frac{t}{\sqrt{n}}\right) \cdot M_{Z_2}\left(\frac{t}{\sqrt{n}}\right) \cdot \ldots \cdot M_{Z_n}\left(\frac{t}{\sqrt{n}}\right)$

$= \left(M_Z\left(\frac{t}{\sqrt{n}}\right)\right)^n$ by iden. distr.

What is $M_C(t) = \lim_{n\to\infty} M_{C_n}(t) = \lim_{n\to\infty} \left(M_Z\left(\frac{t}{\sqrt{n}}\right)\right)^n$?

1901–1902

"tail" on tour

$$\lim_{n\to\infty}\left(1 + \frac{t}{\sqrt{n}}\cancel{E[Z]}^{0} + \frac{t^2}{2n}\cancel{E[Z^2]}^{1} + \underbrace{\overbrace{\frac{t^3}{6n^{3/2}}E[Z^3] + \frac{t^4}{24n^2}E[Z^4] + \dots}^{e(n)}}_{e(n) \in o\left(\frac{1}{n}\right)}\right)^n$$

$\leftarrow$ asymptotic notation

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

if $\lim_{n\to\infty} \frac{e(n)}{\frac{1}{n}} = 0$ true?

$\Rightarrow \lim_{n\to\infty} n\, e(n) = 0 \Rightarrow \lim_{n\to\infty} n\left(\frac{c_1}{n^{3/2}} + \frac{c_2}{n^2} + \dots\right)$

$\Rightarrow \lim_{n\to\infty}\left(\frac{c_1}{n^{1/2}} + \frac{c_2}{n} + \dots\right)$

$0 + 0 + \dots = 0 \checkmark$

It means this term is going to 0 faster than $\frac{1}{n}$ is going to 0

Hardcore asymptotics... in the above, the $o\left(\frac{1}{n}\right)$ won't matter...

big hole How pick? $E[Z^3], E[Z^4], \dots$

$$\Rightarrow \lim_{n\to\infty}\left(1 + \frac{t^2/2}{n} + o\left(\frac{1}{n}\right)\right)^n = \lim_{n\to\infty}\left(1 + \frac{t^2/2}{n}\right)^n = e^{t^2/2} \Rightarrow \boxed{C \sim N(0,1)\,!!!}$$

$\Rightarrow C_n \xrightarrow{d} N(0,1)$

as ~~n→∞~~, C_n becomes more and more distributed as $N(0,1)$

all 3 defs:

$\Rightarrow \underbrace{\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{d} N(0,1)}_{\text{most useful}}$

$\sqrt{n}\,\bar{Z} \to N(0,1)$

Now... the most useless def.

$\underbrace{\frac{T_n - n\mu}{\sqrt{n}\sigma} \xrightarrow{d} N(0,1)}_{\text{second most useful}}$

So for "big n," and we don't know exactly how big (see HW #9 11 am),

① $\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$ and ② $\frac{T_n - n\mu}{\sqrt{n}\sigma} \sim N(0,1)$

For the purposes of this class, these are true if I say "big n."

$\Rightarrow$ shift and scales: $SE^2 = Var$... notation horror!

③ $\bar{X} \sim N\left(\mu, \left(\frac{\sigma}{\sqrt{n}}\right)^2\right)$ and ④ $T_n \sim N\left(n\mu, (\sqrt{n}\sigma)^2\right)$

These 4 r.v.'s at "big n" are of paramount importance!

We now are done with probability and are moving to Statistics.

SKIP to p5 ....

~~[illegible]~~

$$T_n \sim \text{Binom}(n,p), \quad E[T_n]=np, \quad Var[T_n]=np(1-p), \quad SE[T_n]=\sqrt{np(1-p)}$$

$$\bar{X} := \frac{X_1+\ldots+X_n}{n} = \frac{\sum \mathbb{1}_{X_i=1}}{n}$$

← Bernoulli! ← one zero otherwise

$E[\bar{X}]=p$, $Var[\bar{X}]=\frac{p(1-p)}{n}$, $SE[\bar{X}]=\sqrt{\frac{p(1-p)}{n}}$

In this special case, the average is called a "proportion".

The # of samples becomes positive ... #! In this special notation:

$$\hat{P} := \frac{\sum_{i=1}^{n} \mathbb{1}_{X_i=1}}{n}$$ distribution of sample proportion

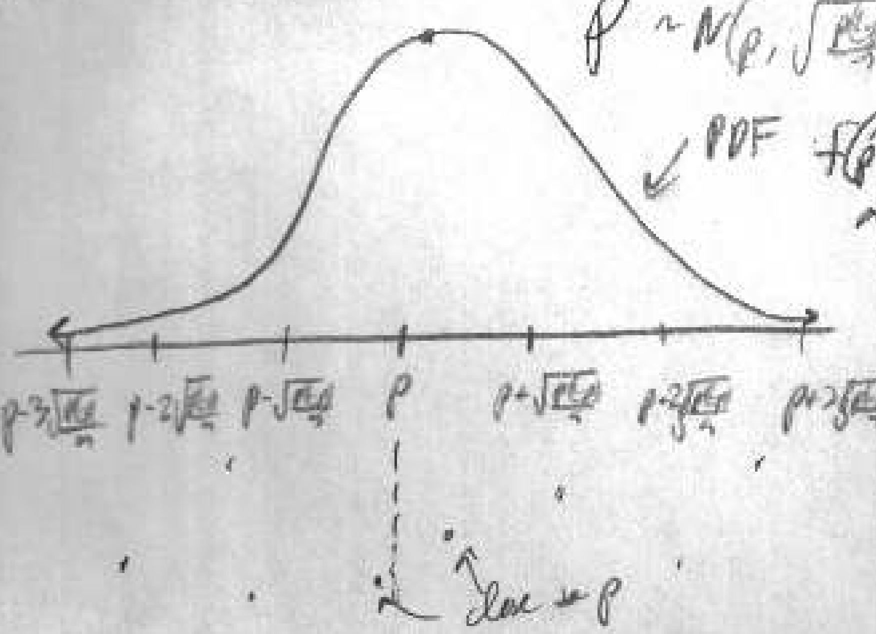
$$\hat{p} := \frac{\sum_{i=1}^{n} \mathbb{1}_{x_i=1}}{n}$$ a sample proportion

Recall $\bar{X} \sim N\left(\mu, \left(\frac{\sigma}{\sqrt{n}}\right)^2\right)$ for big n

Here, $\hat{P} \sim N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$

What does this look like?

$\hat{P} \sim N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$

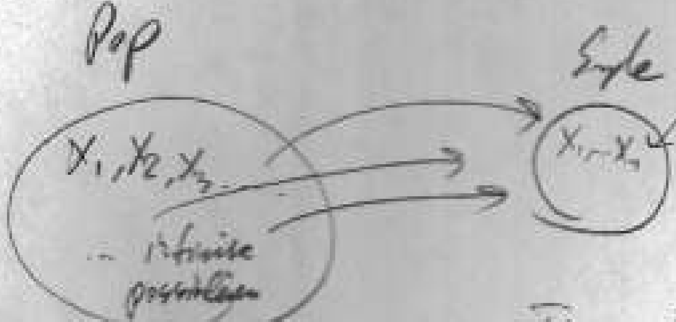


PDF $f(\hat{p}) = \frac{1}{\sqrt{2\pi \frac{p(1-p)}{n}}} e^{-\frac{1}{2\frac{p(1-p)}{n}}(\hat{p}-p)^2}$

free variable are the draws from $\hat{P}$, the realizations.

... from $\hat{p}$

Here's the deal. You "sample" n iid Bernoullis from the population and realize them...



size n. "If n is big", CLT kicks in and $\hat{p} \sim N$ as we drew

Thus $\hat{p}$ is a draw/realization from $\hat{P}$.

Could be anyplace! (SRS)

We want to infer p.

$|Pop| \approx |N|$ ("all ∞")

Almost there... need couple more tools. Review the norm. distr. again!

$Z \sim N(0,1)$
$$P(Z \in [-1,1]) = 68\%$$
$$P(Z \in [-2,2]) = 95\%$$
$$P(Z \in [-3,3]) = 99.7\%$$
review of the empirical rule for standard normal dist

If $X \sim N(\mu, \sigma)$
$$P(X > \mu+\sigma) = P(X-\mu > \sigma) = P\left(\frac{X-\mu}{\sigma} > 1\right) = P(Z > 1)$$
$$P(X < \mu-\sigma) = P(X-\mu < -\sigma) = P\left(\frac{X-\mu}{\sigma} < -1\right) = P(Z < -1)$$
$$\Rightarrow P(X \in [\mu-\sigma, \mu+\sigma]) = 1 - P(|Z| > 1) = 68\%$$
$$P(X \in [\mu-2\sigma, \mu+2\sigma]) = 1 - P(|Z| > 2) = 95\%$$
$$P(X \in [\mu-3\sigma, \mu+3\sigma]) = 1 - P(|Z| > 3) = 99.7\%$$
empirical rule for general normal dist.