Lecture 20 4/30/15 Math 241

Review

$$M_X(t) = E(g(X)) = E(e^{tX}) = \sum_{x \in supp} e^{tx} p(x) \quad \text{discrete}$$
$$= \int_{x \in supp} e^{tx} f(x)\, dx \quad \text{cont}$$

Properties

① $p(x) \Leftrightarrow M_X(t)$ and $f(x) \Leftrightarrow M_X(t)$ unique

② If $X_1, \ldots, X_n$ ind and mgf's exist

$$M_{\sum_{i=1}^n X_i}(t) = \prod_{i=1}^n M_{X_i}(t)$$

③ $M_X^{(n)}(0) = E(X^n)$ Big $!

④ $Y = aX + c \Rightarrow M_Y(t) = e^{tc} M_X(at)$

$X \sim Bern(p) \Rightarrow M_X(t) = 1 - p + pe^t$
$X \sim Bin(n, p) \Rightarrow M_X(t) = (1 - p + pe^t)^n$
$X \sim Exp(\lambda) \Rightarrow M_X(t) = \frac{\lambda}{\lambda - t}$ for $\lambda - t > 0$
$Z \sim N(0,1) \Rightarrow M_Z(t) = e^{t^2/2}$

Setup $X_1, \ldots, X_n \overset{iid}{\sim}$ something w/ mean μ, SE σ

Interested in $C_n = \dfrac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ } the standardized $\bar{X}$

C_n has mean 0, SE 1
no idea how it's distr.

Recall $Z_i := \frac{X_i - \mu}{\sigma}$ then..

$C_n = \dots = \frac{Z_1}{\sqrt{n}} + \frac{Z_2}{\sqrt{n}} + \dots + \frac{Z_n}{\sqrt{n}}$ through algebra tricks

Recall $E(Z_i) = 0$ and $SE(Z_i) = 1$ due to std.

$\Rightarrow Var(Z_i) = 1$

$\Rightarrow E(Z_i^2) = 1$ since $Var(X) = E(X^2) - \cancel{\mu^2}^{0}$

Remember

Also recall

prop. 3 by prop. 1 (and ident. distr)

$$M_{C_n}(t) = M_{\frac{Z_1}{\sqrt{n}} + \dots + \frac{Z_n}{\sqrt{n}}}(t) = M_{\frac{Z_1}{\sqrt{n}}}(t)\, M_{\frac{Z_2}{\sqrt{n}}}(t) \cdots M_{\frac{Z_n}{\sqrt{n}}}(t) = \left(M_{\frac{Z}{\sqrt{n}}}(t)\right)^n = \left(M_Z\left(\tfrac{t}{\sqrt{n}}\right)\right)^n$$

by prop. 4

We don't know $M_Z(t)$ since we don't know the distr of Z since we " the distr of X!

but we do know:

$$M_Z(t) := E[e^{tZ}] = 1 + \frac{t}{1!}E(Z) + \frac{t^2}{2!}E(Z^2) + \frac{t^3}{3!}E(Z^3) + \frac{t^4}{4!}E(Z^4) + \dots$$

by Taylor series expansion and linearity of the expectation operator

$$\Rightarrow M_Z\left(\tfrac{t}{\sqrt{n}}\right) = 1 + \frac{t}{\sqrt{n}}E(Z) + \frac{t^2}{n\,2!}E(Z^2) + \frac{t^3}{n^{3/2}\,3!}E[Z^3] + \frac{t^4}{n^2\,4!}E(Z^4) + \dots$$

But we know since Z is stdzd, $E(Z) = 0$, $E(Z^2) = 1$

know nothing

$$\Rightarrow M_Z\left(\tfrac{t}{\sqrt{n}}\right) = 1 + \frac{t^2/2}{n} + \underbrace{\frac{t^3}{n^{3/2}\,3!}E(Z^3) + \frac{t^4}{n^2\,4!}E(Z^4) + \dots}$$

let tail :=

$\Rightarrow M_Z\left(\frac{t}{\sqrt{n}}\right) = 1 + \frac{t^2/2}{n} + \text{tail}$

We say that $\text{tail} \in o\left(\frac{1}{n}\right)$ ← pronounced "little-o of one over n"

if $\lim_{n\to\infty} \frac{\text{tail}}{\frac{1}{n}} = 0$ meaning tail goes ~~to~~ 0 "faster" than $\frac{1}{n}$. For instance,

$$\frac{1}{n^2} \in o\left(\frac{1}{n}\right) \quad \text{since } \frac{1}{n^2} \to 0 \text{ faster than } \frac{1}{n} \to 0$$

$$\text{since } \lim_{n\to\infty} \frac{\frac{1}{n^2}}{\frac{1}{n}} = \lim_{n\to\infty} \frac{n}{n^2} = \lim \frac{1}{n} = 0$$

even $\frac{1}{n^{1.001}} \in o\left(\frac{1}{n}\right)$

$$\text{since } \lim \frac{\frac{1}{n^{1.001}}}{\frac{1}{n}} = \lim \frac{1}{n^{.001}} = 0$$

So is it true $\text{tail} \in o\left(\frac{1}{n}\right)$?

$$\text{tail} = \frac{c_1}{n^{3/2}} + \frac{c_2}{n^2} + \frac{c_3}{n^{5/2}} + \dots$$

$$\lim \frac{\frac{c_1}{n^{3/2}} + \frac{c_2}{n^2} + \frac{c_3}{n^{5/2}} + \dots}{\frac{1}{n}} = \lim c_1 \frac{1}{n^{1/2}} + c_2 \frac{1}{n} + c_3 \frac{1}{n^{3/2}} + \dots = 0$$

$$\Rightarrow M_Z\left(\frac{t}{\sqrt{n}}\right) = 1 + \frac{t^2/2}{n} + o\left(\frac{1}{n}\right)$$

$$\Rightarrow M_{C_n}(t) = \left(M_Z\left(\frac{t}{\sqrt{n}}\right)\right)^n = \left(1 + \frac{t^2/2}{n} + o\left(\frac{1}{n}\right)\right)^n$$

Setup $X_1, \dots, X_n \overset{iid}{\sim}$ something w/ μ, σ finite. What if n gets large?

$$\lim_{n\to\infty} C_n = \lim_{n\to\infty} \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \Rightarrow \lim_{n\to\infty} M_{C_n}(t) = \lim_{n\to\infty} \left(1 + \frac{t^2/2}{n} + o\left(\frac{1}{n}\right)\right)^n$$

Recall...

$$e := \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n \quad \text{and} \quad \Rightarrow \quad e^a := \lim\left(1+\frac{a}{n}\right)^n \quad \text{for } a \in \mathbb{R}$$
(a constant)

What about?

$$\lim_{n\to\infty}\left(1+\frac{1}{n}+\frac{1}{n^2}\right)^n$$ does the $\frac{1}{n^2}$ matter? ... No...

$$\lim_{n\to\infty}\left(1+\frac{1}{n}+\frac{1}{n^{1.001}}\right)^n \quad \text{''} \quad \text{''}$$
something bigger than 1?

It has been proven that

Thm $$\lim_{n\to\infty}\left(1+\frac{a}{n}+o\left(\frac{1}{n}\right)\right)^n = e^a$$ But may do so slowly... see HW!!

which involves serious math!

By this

$$\Rightarrow \lim_{n\to\infty} M_{Z_n}(t) = \lim_{n\to\infty}\left(1+\frac{t^2/2}{n}+o\left(\frac{1}{n}\right)\right)^n = e^{t^2/2} = M_Z(t)$$
the MGF for $N(0,1)$!!!

$$\Rightarrow \quad \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1) \quad \text{as } n\to\infty$$

CENTRAL LIMIT THM

the crown jewel of Math 241 !!

Now some corollaries...

If n is large...

$$\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} = Z \sim N(0,1) \quad \Rightarrow \quad \bar{X} = \frac{\sigma}{\sqrt{n}} Z + \mu \sim N\left(\mu, \left(\frac{\sigma}{\sqrt{n}}\right)^2\right)$$

$\bar{X}$ itself is approx normally distr!

total / sum r.v

And... $\bar{X} = \frac{X_1+\dots+X_n}{n} = \frac{T_n}{n} = \frac{\sigma}{\sqrt{n}}Z+\mu$

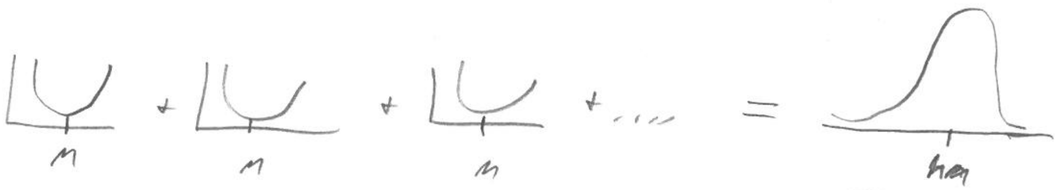
$$\Rightarrow T_n = \sigma\sqrt{n}\,Z + n\mu \sim N\left(n\mu, (\sigma\sqrt{n})^2\right)$$

the total r.v. is also approx Norm dist

What does this mean? $X \sim$ Bathtub (see HW)

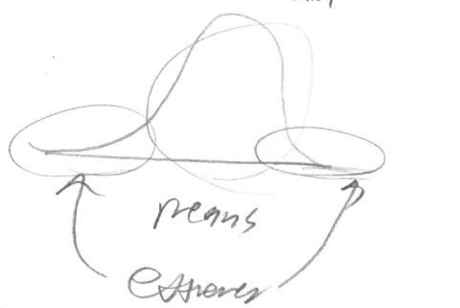
totally not Normal!

But



AMAZING!

What is it about $\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$



Normal dist. is the normal balance between means and extremes in the long run...
means happen a lot, extremes don't happen too often...

Many things affect the rate of convergence ... but we're not going to talk about it!

Example $X_1,\dots,X_{30} \overset{iid}{\sim} Geom(\frac{1}{2})$ wait for coin flips.

$\bar{X} = \frac{X_1+\dots+X_{30}}{30}$ What is the prob on avg I wait more than 2.5 flips?

$X \sim Geom(\frac{1}{2}) \Rightarrow \mu = \frac{1}{p} = 2$, $\sigma^2 = \frac{1-p}{p^2} = \frac{\frac{1}{2}}{\frac{1}{4}} = 2 \Rightarrow \sigma = \sqrt{2} \Rightarrow \frac{\sigma}{\sqrt{n}} = \frac{\sqrt{2}}{\sqrt{30}} = \sqrt{\frac{1}{15}} \approx .258$

n=30 is LARGE by CLT $\Rightarrow \bar{X} \sim N\left(2, (.258)^2\right)$ Now $P(\bar{X} > 2.5) = P\left(\frac{\bar{X}-2}{.258} > \frac{2.5-2}{.258}\right) = P(Z > 1.94) \approx P(Z > 2) = 2.5\%$

Another example...

$X_1, \dots, X_n \overset{iid}{\sim} Bern(0.3)$ $n = 50$ ← experiments

What is prob of less than 27% of the 50 being successful

$P(\bar{X} < 0.27)$ $X \sim Bern(0.3) \Rightarrow \mu = 0.3$ $\sigma^2 = 0.3(1-0.3) = 0.21 \Rightarrow \sigma = \sqrt{0.21} = 0.458$

$\Rightarrow \frac{\sigma}{\sqrt{n}} = \frac{\sigma}{\sqrt{50}} = 0.0648$

$\bar{X} = \frac{X_1 + \dots + X_n}{n} \sim N\left(p, \left(\sqrt{\frac{p(1-p)}{n}}\right)^2\right)$

μ $\frac{\sigma}{\sqrt{n}} = \frac{\sqrt{p(1-p)}}{\sqrt{n}}$

$\Rightarrow \bar{X} \sim N(0.3, 0.0648^2)$

$P(\bar{X} < 0.27) = P\left(\frac{\bar{X} - 0.3}{0.0648} < \frac{0.27 - 0.3}{0.0648}\right)$

$= P(Z < 0.92) \approx P(Z < 1) = \boxed{16\%}$

In general

$X_1, \dots, X_n \overset{iid}{\sim} Bern(p)$ not for any r.v.

→ the avg r.v.

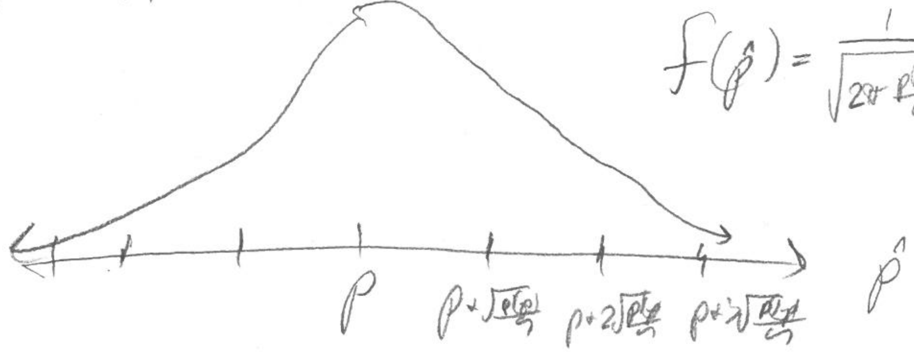
$\bar{X}$ has a special name here, we call it $\hat{p}$

which is the r.v. for the sample proportion

$\bar{X}$ has a special name here, we call it $\hat{p}$, the sample proportion
↳ the sample avg

$\hat{p} = \frac{\#1\text{'s}}{n}$ example: Who likes mushrooms?

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

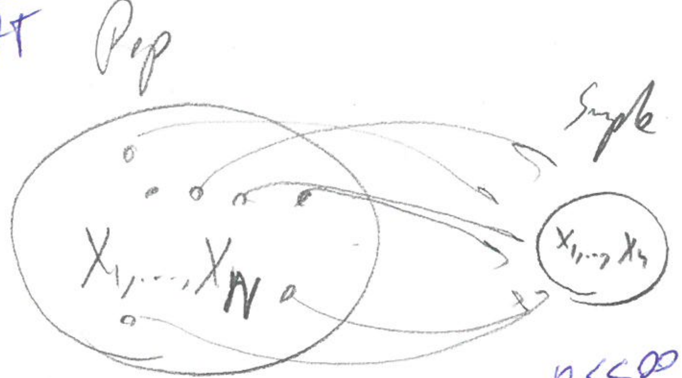


In a single experiment (a run of n iid samples)

$\hat{p}$ could fall anywhere here

↑ PROB

↓ STAT



$n << \infty$ waaay less than

$N = |Pop| \approx |\mathbb{N}|$

if n is big CLT kicks in

$\hat{p} \sim N$

$\hat{p}$ is a realization

$X_1,\dots,X_n \overset{iid}{\sim} Bern(p)$

p is the pop. prop.

(previously "parameter")

Is this sample representative?