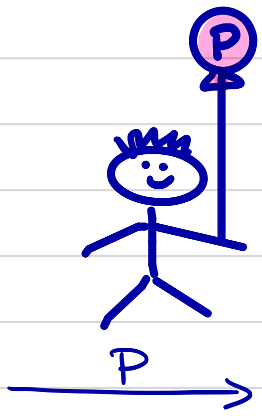
Where to next?

§ 5.5 Convergence Concepts



Definition 5.5.1 A sequence of random variables, $X_1, X_2, \ldots$, *converges in probability* to a random variable X if, for every $\epsilon > 0$,

$$\underbrace{\lim_{n\to\infty} P(|X_n - X| \geq \epsilon)}_{\text{probability}} = 0 \quad \text{or, equivalently,} \quad \lim_{n\to\infty} P(|X_n - X| < \epsilon) = 1.$$

start here

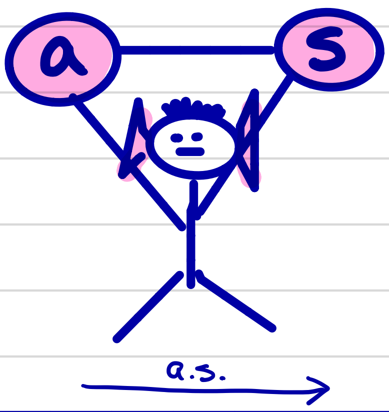
WLLN

$\bar{X}_n \xrightarrow{P} E[X]$

Theorem 5.5.2 (Weak Law of Large Numbers) *Let $X_1, X_2, \ldots$ be iid random variables with $\mathrm{E}X_i = \mu$ and $\mathrm{Var}\, X_i = \sigma^2 < \infty$. Define $\bar{X}_n = (1/n)\sum_{i=1}^n X_i$. Then, for every $\epsilon > 0$,*

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| < \epsilon) = 1;$$

that is, $\bar{X}_n$ converges in probability to μ.



Definition 5.5.6 A sequence of random variables, $X_1, X_2, \ldots$, *converges almost surely* to a random variable X if, for every $\epsilon > 0$,

$$P(\underbrace{\lim_{n\to\infty} |X_n - X| < \epsilon}_{\text{pointwise convergence}}) = 1.$$

start here

SLLN

$\bar{X}_n \xrightarrow{a.s.} E[X]$

Theorem 5.5.9 (Strong Law of Large Numbers) *Let $X_1, X_2, \ldots$ be iid random variables with $\mathrm{E}X_i = \mu$ and $\mathrm{Var}\, X_i = \sigma^2 < \infty$, and define $\bar{X}_n = (1/n)\sum_{i=1}^n X_i$. Then, for every $\epsilon > 0$,*

$$P(\lim_{n\to\infty} |\bar{X}_n - \mu| < \epsilon) = 1;$$

that is, $\bar{X}_n$ converges almost surely to μ.

Definition 5.5.10 A sequence of random variables, $X_1, X_2, \ldots$, *converges in distribution* to a random variable X if

convergence in Law

$$\lim_{n\to\infty} \underline{F_{X_n}(x)} = F_X(x)$$

start here (or mgf).

at all points x where $F_X(x)$ is continuous.

CLT

$f_{\bar{X}} \xrightarrow{\mathcal{L}} N(\mu, \sigma^2/n)$

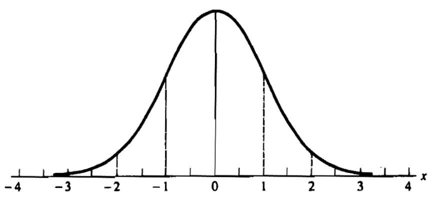


Figure 3.3.1. *Standard normal density*

Theorem 5.5.14 (Central Limit Theorem) *Let $X_1, X_2, \ldots$ be a sequence of iid random variables whose mgfs exist in a neighborhood of 0 (that is, $M_{X_i}(t)$ exists for $|t| < h$, for some positive h). Let $\mathrm{E}X_i = \mu$ and $\mathrm{Var}\, X_i = \sigma^2 > 0$. (Both μ and σ^2 are finite since the mgf exists.) Define $\bar{X}_n = (1/n)\sum_{i=1}^n X_i$. Let $G_n(x)$ denote the cdf of $\sqrt{n}(\bar{X}_n - \mu)/\sigma$. Then, for any x, $-\infty < x < \infty$,*

$$\lim_{n\to\infty} G_n(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy;$$

that is, $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ has a limiting standard normal distribution.

$$\therefore \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \xrightarrow{\mathcal{L}} N(0,1)$$

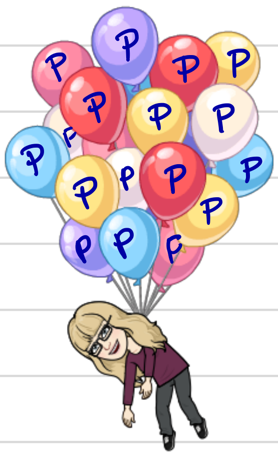
proof: homework-2

Theorem 5.5.15 (Stronger form of the Central Limit Theorem) *Let $X_1, X_2, \ldots$ be a sequence of iid random variables with $\mathrm{E}X_i = \mu$ and $0 < \mathrm{Var}\, X_i = \sigma^2 < \infty$. Define $\bar{X}_n = (1/n)\sum_{i=1}^n X_i$. Let $G_n(x)$ denote the cdf of $\sqrt{n}(\bar{X}_n - \mu)/\sigma$. Then, for any x, $-\infty < x < \infty$,*

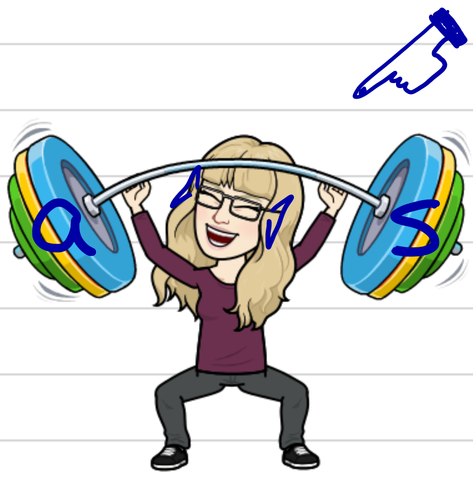
$$\lim_{n\to\infty} G_n(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy;$$

that is, $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ has a limiting standard normal distribution.

proof uses characteristic ft'n $E[e^{itX}]$ $(i^2 = -1)$



$\xrightarrow{P}$

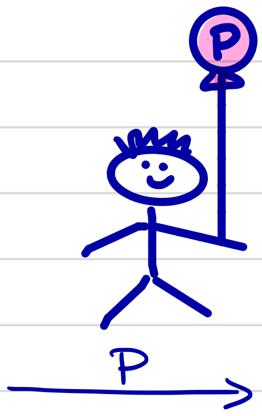


$\xrightarrow{a.s.}$

$\xrightarrow{\mathcal{L}}$

§ 5.5 Convergence Concepts

§ 5.5.1 Convergence in Probability $\xrightarrow{P}$



Definition 5.5.1 A sequence of random variables, $X_1, X_2, \ldots$, *converges in probability* to a random variable X if, for every $\epsilon > 0$,

$$\lim_{n\to\infty} \underbrace{P(|X_n - X| \geq \epsilon)}_{\text{probability}} = 0 \quad \text{or, equivalently,} \quad \lim_{n\to\infty} P(|X_n - X| < \epsilon) = 1.$$

start here

Dist'n of X_n changes as $n \to \infty$

Example $\bar{X}_1 = X_1$

$\bar{X}_2 = (X_1+X_2)/2$

$\vdots$

$\bar{X}_n = \Sigma X_i/n$

$\vdots$

$\bar{X}_n$ μ $\to$ WLLN

μ = pop'n mean

$\lim_{n\to\infty} (P(X_n - X) \geq \varepsilon) = 0$

Most famous result of this type:

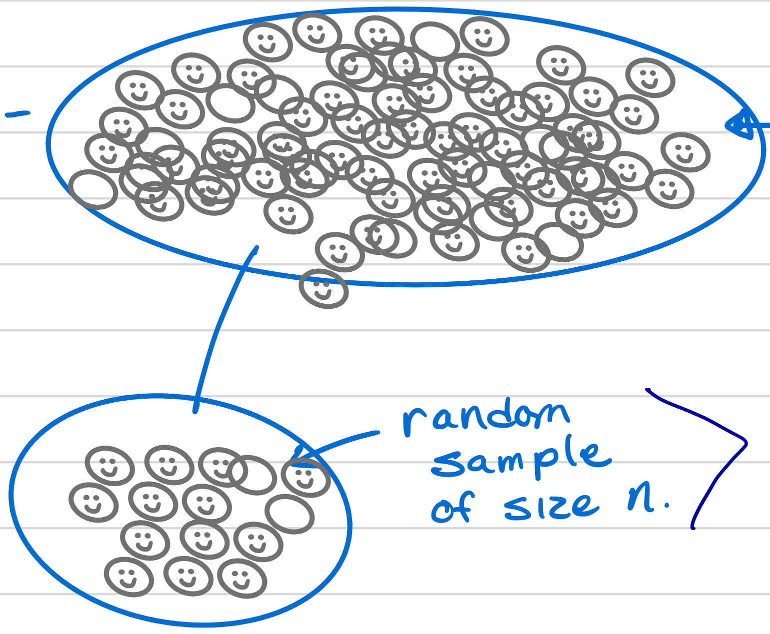
Theorem 5.5.2 (Weak Law of Large Numbers) *Let $X_1, X_2, \ldots$ be iid random variables with $\mathrm{E}X_i = \mu$ and $\mathrm{Var}\, X_i = \sigma^2 < \infty$. Define $\bar{X}_n = (1/n)\sum_{i=1}^n X_i$. Then, for every $\epsilon > 0$,*

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| < \epsilon) = 1;$$

that is, $\bar{X}_n$ converges in probability to μ.

$$\bar{X}_n \xrightarrow{P} \mu$$

sample mean $\xrightarrow{P}$ pop'n mean



population with dist'n $f(x)$ with mean μ

random sample of size n.

mean $\bar{X}_n$

WLLN 'quite elegantly states'

Sample mean $\bar{X}_n$ $\xrightarrow{\text{approaches}}$ Pop'n mean μ as $n \to \infty$.

Proof of WLLN (use Cheby's Inequality)

Theorem 3.6.1 (Chebychev's Inequality) *Let X be a random variable and let $g(x)$ be a nonnegative function. Then, for any $r > 0$,*

$$P\left(g(X) \geq r\right) \leq \frac{\mathrm{E}g(X)}{r}.$$

Start with Probability.

$$P(|\bar{X}_n - \mu| \geq \varepsilon) = P((\bar{X}_n-\mu)^2 \geq \varepsilon^2)$$
$$\leq E[(\bar{X}_n-\mu)^2] / \varepsilon^2$$
$$= \frac{Var(\bar{X})}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}$$

Theorem 5.2.6 *Let $X_1, \ldots, X_n$ be a random sample from a population with mean μ and variance $\sigma^2 < \infty$. Then*

a. $\mathrm{E}\bar{X} = \mu$,

b. $\mathrm{Var}\,\bar{X} = \frac{\sigma^2}{n}$,

c. $\mathrm{E}S^2 = \sigma^2$.

$$\therefore P(|\bar{X}_n-\mu| < \varepsilon) = [1 - P(|\bar{X}_n-\mu| \geq \varepsilon)]$$
$$\geq 1 - \sigma^2/n\varepsilon^2$$

Now take limits

$$\lim_{n\to\infty} P(|\bar{X}_n-\mu| < \varepsilon) \geq \lim_{n\to\infty}[1 - \sigma^2/n\varepsilon^2] = 1$$

since Probability can't be > 1

$$\lim_{n\to\infty} P(|\bar{X}_n-\mu| < \varepsilon) = 1 \quad // \text{ Q.E.D.}$$

Property summarized by WLLN is consistency (more later).

Definition 10.1.1 A sequence of estimators $W_n = W_n(X_1, \ldots, X_n)$ is a *consistent sequence of estimators* of the parameter θ if, for every $\epsilon > 0$ and every $\theta \in \Theta$,

(10.1.1) $\lim_{n\to\infty} P_\theta(|W_n - \theta| < \epsilon) = 1.$

Informally, (10.1.1) says that as the sample size becomes infinite (and the sample information becomes better and better), the estimator will be arbitrarily close to the parameter with high probability, an eminently desirable property. Or, turning things around, we can say that the probability that a consistent sequence of estimators misses the true parameter is small. An equivalent statement to (10.1.1) is this: For every $\epsilon > 0$ and every $\theta \in \Theta$, a consistent sequence W_n will satisfy

(10.1.2) $\lim_{n\to\infty} P_\theta(|W_n - \theta| \geq \epsilon) = 0.$

Example: Consistency of S^2

Sequence of $X_1, X_2, \ldots$ iid random variables with $E[X_i] = \mu$ + $Var[X_i] = \sigma^2 < \infty$

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{(n-1)}$$

Prove $S_n^2 \xrightarrow{P} \sigma^2$ Using Cheby's (again).

$$P(|S_n^2 - \sigma^2| \geq \varepsilon) \leq \frac{E[(S_n^2-\sigma^2)^2]}{\varepsilon^2} = \frac{Var(S_n^2)}{\varepsilon^2}$$

$$\lim_{n\to\infty} P(|S_n^2-\sigma^2| \geq \varepsilon) = \lim_{n\to\infty}\frac{Var(S_n^2)}{\varepsilon^2}$$

If $\lim_{n\to\infty} Var(S_n^2) = 0$

Then $\lim_{n\to\infty} P(|S_n^2 - \sigma^2| \geq \varepsilon) = 0$ //

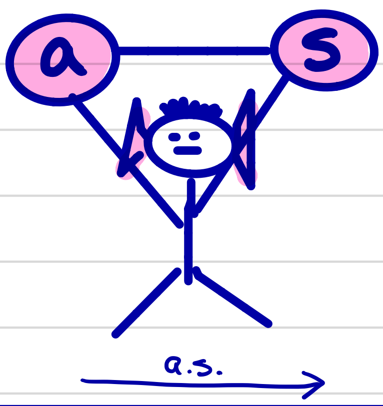
If $S_n^2 \xrightarrow{P} \sigma^2$ does $S_n \xrightarrow{P} \sigma$?

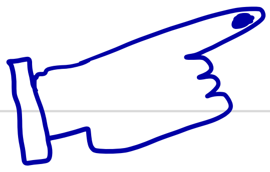
Theorem 5.5.4 *Suppose that $X_1, X_2, \ldots$ converges in probability to a random variable X and that h is a continuous function. Then $h(X_1), h(X_2), \ldots$ converges in probability to $h(X)$.*

Example (Consitency of S):

If $S_n^2 \xrightarrow{P} \sigma^2$ then $\sqrt{S_n^2} \xrightarrow{P} \sigma$ by Thm 5.5.4.

§5.5.2 Almost sure Convergence $\xrightarrow{a.s.}$





Definition 5.5.6 A sequence of random variables, $X_1, X_2, \ldots$, *converges almost surely* to a random variable X if, for every $\epsilon > 0$,

$$P(\lim_{n\to\infty} |X_n - X| < \epsilon) = 1.$$

pointwise convergence

start here

Similar to 'pointwise' convergence, but convergence need not occur on a set with probability zero.

a.k.a. (Hogg & Craig, 'convergence with probability 1).

a.k.a. may not converge for probability zero subspace But converges everywhere else.

Recall: **Definition 1.4.1** A *random variable* is a function from a sample space S into the real numbers.

∴ RV is real-valued ft'n defined on S

→ for $s \in S$ $X_n(s)$ and $X(s)$ are ft'ns defined on S

→ $X_n \xrightarrow{a.s.} X$ if $X_n(s) \xrightarrow{a.s.} X(s)$ $\forall s \in S$ except perhaps for $s \in N$ where $N \subset S$ and $P(N) = 0$.

→ i.e. the function $X_n(s)$ converges as $n \to \infty$ to $X(s)$

$\lim_{n\to\infty} |X_n(s) - X(s)| < \varepsilon$ with probability 1.

Example 5.5.7 (Almost sure convergence).

$$X_n \xrightarrow{a.s} X \quad P\left(\lim_{n\to\infty} |X_n - X| < \varepsilon\right) = 1$$

start with pointwise convergence

Let sample space S be interval $[0,1]$ $(0 \le s \le 1)$ with uniform prob. dist'n.

Define RVs $X_n(s) = s + s^n$
$X(s) = s$

Does $X_n(s) \xrightarrow{a.s.} X(s)$?

Consider $S = \{[0,1), 1\}$

Start with Pointwise Convergenc

$s \in [0,1)$, $0 \le s < 1$: for every $s \in [0,1)$ $\lim_{n\to\infty} s^n = 0$

$\therefore \lim_{n\to\infty} X_n(s) = X(s)$

$\lim_{n\to\infty} s^n + s = s + 0$

$s = 1$: for $s = 1$ $s^n = 1$ $\forall n$.
$X_n(1) = 1 + 1^n = 2$ $\forall n$.

$\lim_{n\to\infty} X_n(1) \not\to X(s)$

Now bring in Probability

any fixed ε, for any s, can choose n large enough for difference $< \varepsilon$.

$\lim_{n\to\infty} |(s+s^n) - s| < \varepsilon$ for $0 \le s < 1$; $P(0 \le s < 1) = \int_0^1 ds = 1$

$P(0 \le s < 1) = 1$

only $< \varepsilon$ if $\varepsilon > 1$, not true $\forall \varepsilon > 0$.

$\lim_{n\to\infty} |\underbrace{(1+1^n) - 1}_{1}| < \varepsilon$ not true for $s = 1$

$P(s=1) = \int_1^1 ds = 0$

Does $X_n(s) \xrightarrow{a.s.} X(s)$?

- Convergence occurs on set $[0,1)$ and $P(0 \le s < 1) = 1$.
- Pointwise convergence for sample space with probability 1.
- Pointwise convergence, except on probability zero subspace

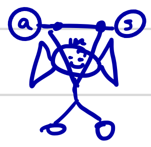
$$\therefore P\left(\lim_{n\to\infty} |X_n - X| < \varepsilon\right) = 1$$

Note similarity between def'ns of $\xrightarrow{P}$ and $\xrightarrow{a.s.}$ in 'looks' but very different statements.

$$\lim_{n\to\infty} P(|X_n - X| < \varepsilon) = 1$$

weak

$$P(\lim_{n\to\infty} |X_n - X| < \varepsilon) = 1$$



strong, pointwise

Example: $\xrightarrow{P}$ but not $\xrightarrow{a.s.}$

Describe a sequence which converges in prob $\xrightarrow{P}$ but not almost surely $\cancel{\xrightarrow{a.s.}}$

Let $S = [0,1]$ or $0 \le s \le 1$ with $U(0,1)$ pdf.

Define sequence $X_1, X_2, \dots$ and $X(s) = 1$

$X_n(3/8)$

$X_1(s) = s + I_{[0,1]}(s)$ — $X_1(3/8) = 1\,3/8$

$X_2(s) = s + I_{[0,1/2]}(s)$ — $X_2(3/8) = 1\,3/8$

$X_3(s) = s + I_{[1/2,1]}(s)$ — $X_3(3/8) = 3/8$

$X_4(s) = s + I_{[0,1/3]}(s)$ — $X_4(3/8) = 3/8$

$X_5(s) = s + I_{[1/3,2/3]}(s)$ — $X_5(3/8) = 1\,3/8$

$X_6(s) = s + I_{[2/3,1]}(s)$ — $X_6(3/8) = 3/8$

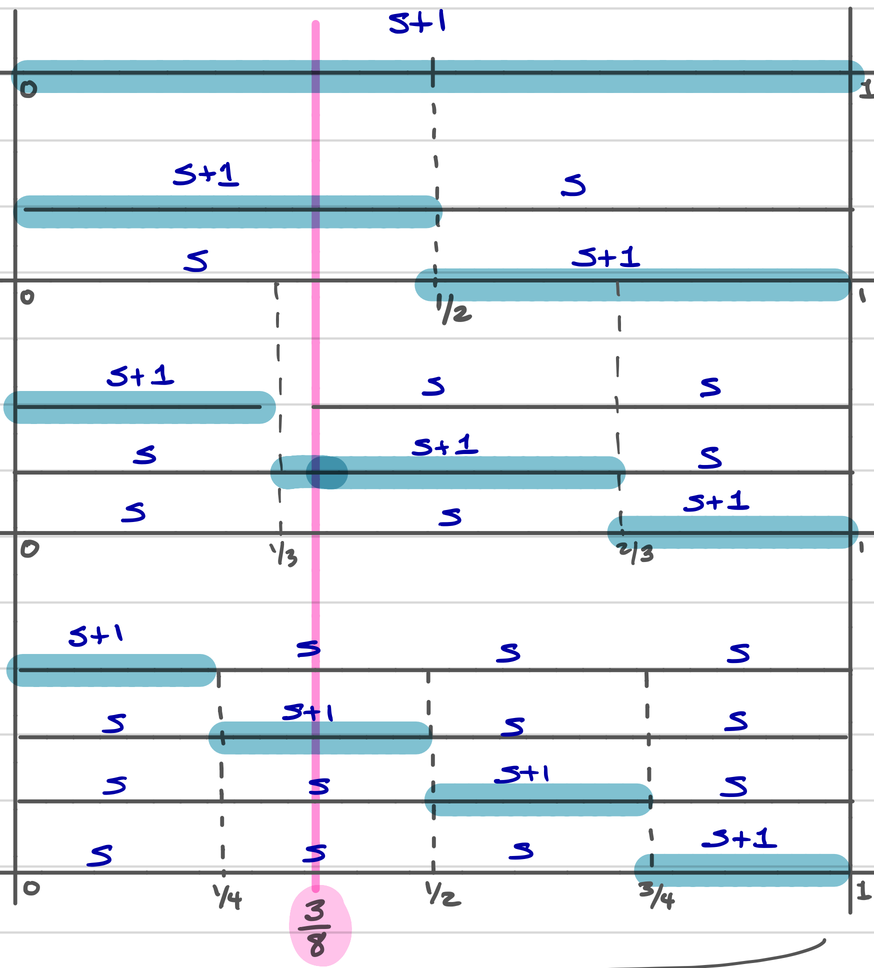
$X_7(s) = s + I_{[0,1/4]}(s)$ — $X_7(3/8) = 3/8$

$X_8(s) = s + I_{[1/4,1/2]}(s)$ — $X_8(3/8) = 1\,3/8$

$X_9(s) = s + I_{[1/2,3/4]}(s)$ — $X_9(3/8) = 3/8$

$X_{10}(s) = s + I_{[3/4,1]}(s)$ — $X_{10}(3/8) = 3/8$

⋮



as n increases interval width decreases

$\xrightarrow{P}$

Define W_n = width of interval for which $I_n(s) = 1$

$Pr(|X_n(s) - X(s)| \ge \varepsilon) = P(s \in W_n)$

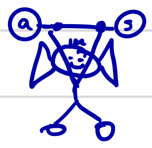
$\therefore \lim_{n\to\infty} P(|X_{(n)}(s) - X(s)| \ge \varepsilon) = \lim_{n\to\infty} P(s \in W_n) = 0$

$\rightarrow \lim_{n\to\infty} P(|X_n - X| \ge \varepsilon) = 0$

$\cancel{\xrightarrow{a.s.}}$

- Note that for $s = 3/8$, we keep returning to $X_i(3/8) = 1\,3/8$
- True for any $s \in [0,1]$
- There is no value $s \in S$ for which $X_n(s) \to s = X(s)$ as $n \to \infty$

i.e. For every $s \in S$, the sequence $X_1(s), X_2(s), \dots$ alternates infinitely often between s and $s+1$.

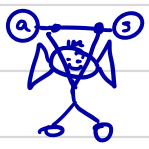


Convergence almost surely $\xrightarrow{a.s.}$ implies $\Rightarrow$ / does not imply $\not\Leftarrow$ $\xrightarrow{P}$ Convergence in prob.

<u>But</u>, if a sequence $\xrightarrow{P}$, it is possible to find a <u>subsequence</u> that $\xrightarrow{a.s.}$

(Homework: previous example find a subsequence that $\xrightarrow{a.s.}$)

weak   WLLN $\bar{X}_n \xrightarrow{P} \mu$ (constant)



strong, pointwise   SLLN $\bar{X}_n \xrightarrow{a.s.} \mu$

Theorem 5.5.9 (Strong Law of Large Numbers) *Let $X_1, X_2, \ldots$ be iid random variables with $\mathrm{E}X_i = \mu$ and $\mathrm{Var}\, X_i = \sigma^2 < \infty$, and define $\bar{X}_n = (1/n)\sum_{i=1}^n X_i$. Then, for every $\epsilon > 0$,*

$$P(\lim_{n\to\infty} |\bar{X}_n - \mu| < \epsilon) = 1;$$

that is, $\bar{X}_n$ converges almost surely to μ.

§ 5.5.3 Convergence in Dist'n (a.k.a Convergence in Law)

Definition 5.5.10 A sequence of random variables, $X_1, X_2, \ldots$, *converges in distribution* to a random variable X if

convergence in Law

$$\lim_{n\to\infty} \underline{F_{X_n}(x)} = F_X(x)$$

start here (or mgf).

at all points x where $F_X(x)$ is continuous.

- "talk" of sequence $X_1, X_2, \cdots$ converging in dist'n (Law).
- It is really cdfs that converge, not R.V.s
- $\therefore$ $\xrightarrow{\mathcal{L}}$ different from $\xrightarrow{P}$ and $\xrightarrow{a.s.}$

Random Variable

Look at large sample behavior of $\bar{X}$, the sample mean, and its limiting dist'n

"one of the most startling theorems in statistics is the Central Limit Theorem"

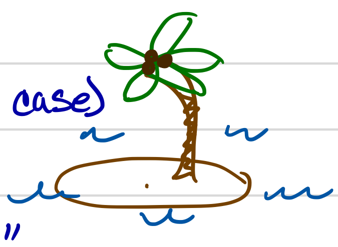
Theorem 5.5.14 (Central Limit Theorem) *Let $X_1, X_2, \ldots$ be a sequence of iid random variables whose mgfs exist in a neighborhood of 0 (that is, $M_{X_i}(t)$ exists for $|t| < h$, for some positive h). Let $\mathrm{E}X_i = \mu$ and $\mathrm{Var}\ X_i = \sigma^2 > 0$. (Both μ and σ^2 are finite since the mgf exists.) Define $\bar{X}_n = (1/n)\sum_{i=1}^n X_i$. Let $G_n(x)$ denote the cdf of $\sqrt{n}(\bar{X}_n - \mu)/\sigma$. Then, for any x, $-\infty < x < \infty$,*

$$\lim_{n\to\infty} G_n(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy;$$

that is, $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ has a limiting standard normal distribution.

$$\therefore \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \xrightarrow{\mathcal{L}} N(0,1)$$

- Starting with 'virtually no assumptions' (independence and finite variance) → $\bar{X} \sim$ normal.
- Normality comes from sums of small independent disturbances (finite variances)
- Limitations
  + Useful approximation, but no automatic way to know how good
  + goodness of approximation is ft'n of original dist'n (check case by case)
  + availability of computing power, importance of CLT lessened



"However despite its limitations, it is still a marvelous result."

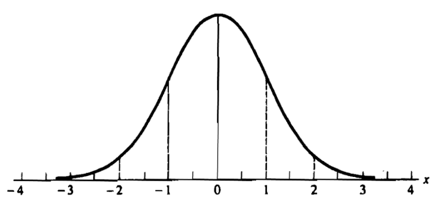


Figure 3.3.1. *Standard normal density*

You look marvelous!

Billy Crystal, SNL ~1985...

Theorem 2.3.12 (Convergence of mgfs) *Suppose $\{X_i,\ i = 1, 2, \ldots\}$ is a sequence of random variables, each with mgf $M_{X_i}(t)$. Furthermore, suppose that*

$$\lim_{i\to\infty} M_{X_i}(t) = M_X(t), \qquad \text{for all } t \text{ in a neighborhood of 0,}$$

and $M_X(t)$ is an mgf. Then there is a unique cdf F_X whose moments are determined by $M_X(t)$ and, for all x where $F_X(x)$ is continuous, we have

$$\lim_{i\to\infty} F_{X_i}(x) = F_X(x).$$

That is, convergence, for $|t| < h$, of mgfs to an mgf implies convergence of cdfs.

Tools needed for proof of CLT.

Definition 2.3.6 Let X be a random variable with cdf F_X. The *moment generating function (mgf) of* X (or F_X), denoted by $M_X(t)$, is

$$M_X(t) = \mathrm{E}\, e^{tX},$$

provided that the expectation exists for t in some neighborhood of 0. That is, there is an $h > 0$ such that, for all t in $-h < t < h$, $\mathrm{E}e^{tX}$ exists. If the expectation does not exist in a neighborhood of 0, we say that the moment generating function does not exist.

Theorem 2.3.7 *If X has mgf $M_X(t)$, then*

$$\mathrm{E}\, X^n = M_X^{(n)}(0),$$

where we define

$$M_X^{(n)}(0) = \frac{d^n}{dt^n} M_X(t)\Big|_{t=0}.$$

That is, the nth moment is equal to the nth derivative of $M_X(t)$ evaluated at $t = 0$.

Theorem 2.3.12 (Convergence of mgfs) *Suppose $\{X_i,\, i = 1, 2, \ldots\}$ is a sequence of random variables, each with mgf $M_{X_i}(t)$. Furthermore, suppose that*

$$\lim_{i\to\infty} M_{X_i}(t) = M_X(t), \quad \text{for all } t \text{ in a neighborhood of } 0,$$

and $M_X(t)$ is an mgf. Then there is a unique cdf F_X whose moments are determined by $M_X(t)$ and, for all x where $F_X(x)$ is continuous, we have

$$\lim_{i\to\infty} F_{X_i}(x) = F_X(x).$$

That is, convergence, for $|t| < h$, of mgfs to an mgf implies convergence of cdfs.

Lemma 2.3.14 *Let $a_1, a_2, \ldots$ be a sequence of numbers converging to a, that is, $\lim_{n\to\infty} a_n = a$. Then*

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

Theorem 2.3.15 *For any constants a and b, the mgf of the random variable $aX+b$ is given by*

$$M_{aX+b}(t) = e^{bt} M_X(at).$$

Theorem 4.6.7 (Generalization of Theorem 4.2.12) *Let $X_1, \ldots, X_n$ be mutually independent random variables with mgfs $M_{X_1}(t), \ldots, M_{X_n}(t)$. Let $Z = X_1 + \cdots + X_n$. Then the mgf of Z is*

$$M_Z(t) = M_{X_1}(t) \cdot \cdots \cdot M_{X_n}(t).$$

In particular, if $X_1, \ldots, X_n$ all have the same distribution with mgf $M_X(t)$, then

$$M_Z(t) = (M_X(t))^n.$$

See Appendix

Definition 5.5.20 If a function $g(x)$ has derivatives of order r, that is, $g^{(r)}(x) = \frac{d^r}{dx^r} g(x)$ exists, then for any constant a, the *Taylor polynomial of order r about a* is

$$T_r(x) = \sum_{i=0}^{r} \frac{g^{(i)}(a)}{i!} (x - a)^i.$$

Taylor's major theorem, which we will not prove here, is that the *remainder* from the approximation, $g(x) - T_r(x)$, always tends to 0 faster than the highest-order explicit term.

Theorem 5.5.21 (Taylor) *If $g^{(r)}(a) = \frac{d^r}{dx^r} g(x)\big|_{x=a}$ exists, then*

$$\lim_{x\to a} \frac{g(x) - T_r(x)}{(x-a)^r} = 0.$$

In general, we will not be concerned with the explicit form of the remainder. Since we are interested in approximations, we are just going to ignore the remainder. There are, however, many explicit forms, one useful one being

$$g(x) - T_r(x) = \int_a^x \frac{g^{(r+1)}(t)}{r!} (x - t)^r \, dt.$$

'Somewhat anticlimatic' proof of the CLT

→ Goal show as $n \to \infty$ $\bar{X} \sim N(\mu, \sigma^2)$, easier work with $\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \sim N(0,1)$ as $n \to \infty$

→ Approach: Show that for $|t|<h$, the mgf of $\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \xrightarrow{\text{converges}}$ to $e^{t^2/2}$ (mgf of $N(0,1)$)

Normal(μ, σ^2)	
pdf	$f(x\|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}$, $-\infty < x < \infty$, $-\infty < \mu < \infty$, $\sigma > 0$
mean and variance	$EX = \mu$, $\operatorname{Var} X = \sigma^2$ $\quad I_{(-\infty,\infty)}(x)$
mgf	$M_X(t) = e^{\mu t + \sigma^2 t^2/2}$
notes	Sometimes called the *Gaussian* distribution.

→ Assume $X_1, X_2, \ldots$ iid w/ $E[X_i] = \mu$ $Var[X_i] = \sigma^2$ with mgf $M_{X_i}(t)$ exists for $|t|<h$.

Define $Y_i = \frac{X_i - \mu}{\sigma}$ (mean 0, variance 1),

$M_Y(t) = $ mgf Y_i exists for $|t| < \sigma h$

Aside (minor for BIOS 6632)

by Thm 2.3.15

$M_{\frac{X_i-\mu}{\sigma}}(t) = M_{\frac{X_i}{\sigma} - \frac{\mu}{\sigma}}(t)$

$= e^{-\mu/\sigma t} M_X(t/\sigma)$ for $|t/\sigma| < h$

→ Express $\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}$ as ft'n of Y_i.

$$\Sigma Y_i = \Sigma\left(\frac{X_i - \mu}{\sigma}\right) = \frac{\Sigma X_i - n\mu}{\sigma} = \frac{\frac{\Sigma X_i - n\mu}{n}}{\sigma/n} = \frac{\bar{X}-\mu}{\sigma/n} = \frac{n(\bar{X}-\mu)}{\sigma}$$

$$\therefore \frac{\Sigma Y_i}{\sqrt{n}} = \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}$$

→ Remember Goal show $\lim_{n\to\infty} M_{\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}}(t) \longrightarrow e^{t^2/2}$

$$M_{\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}}(t) = M_{\frac{\Sigma Y_i}{\sqrt{n}}}(t) \underset{\text{Thm 2.3.15}}{=} M_{\Sigma Y_i}(t/\sqrt{n}) \underset{\text{Thm 4.6.7}}{=} \left(M_Y(t/\sqrt{n})\right)^n$$

Aside (deja vu)

$M_Y(t)$ exists $|t| < \sigma h$

$M_Y(t/\sqrt{n})$ exists $|t| < \sqrt{n}\sigma h$

by Thm 2.3.15

→ $M_{\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}}(t) = \left(M_Y(t/\sqrt{n})\right)^n$ Expand $M_Y(t/\sqrt{n})$ in a 2nd order ($r=2$) Taylor's Series about zero ($a=0$).

$$M_Y(t/\sqrt{n}) = \sum_{k=0}^{\infty} M_Y^{(k)}(0) \frac{(t/\sqrt{n} - 0)^k}{k!} \quad \text{where } M_Y^{(k)}(0) = \frac{d^k}{dt^k} M_Y(t)\Big|_{t=0}$$

$$M_Y(t/\sqrt{n}) = M_Y^{(0)}(0)\frac{(t/\sqrt{n})^0}{0!} + M_Y^{(1)}(0)\frac{(t/\sqrt{n})^1}{1!} + M_Y^{(2)}(0)\frac{(t/\sqrt{n})^2}{2!} + \sum_{k=3}^{\infty} M_Y^{(k)}(0)\frac{(t/\sqrt{n})^k}{k!}$$

$M_Y^{(0)}(0) = E[e^{tX}]\big|_{t=0} = 1$

$\left.\begin{array}{l} M_Y^{(1)}(0) = E[Y] = 0 \\ M_Y^{(2)}(0) = E[Y^2] = 1 \end{array}\right\}$ since Y has mean 0, var 1 by construction.

$\rightarrow M_{\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}}(t) = \left(M_Y(t/\sqrt{n})\right)^n$ Expand $M_Y(t/\sqrt{n})$ in a 2^{nd} order $(r=2)$ Taylor's Series about zero $(a=0)$.

$$M_Y(t/\sqrt{n}) = M_Y^{(0)}(0)\frac{(t/\sqrt{n})^0}{0!} + M_Y^{(1)}(0)\frac{(t/\sqrt{n})^1}{1!} + M_Y^{(2)}(0)\frac{(t/\sqrt{n})^2}{2!} + \sum_{k=3}^{\infty} M_Y^{(k)}(0)\frac{(t/\sqrt{n})^k}{k!}$$

$$M_Y(t/\sqrt{n}) = 1 + 0 + \frac{(t/\sqrt{n})^2}{2!} + R_Y(t/\sqrt{n})$$

$$R_Y(t/\sqrt{n}) = \sum_{k=3}^{\infty} M_Y^{(k)}(0)\frac{(t/\sqrt{n})^k}{k!}$$

$\rightarrow$ By Taylor's Thm $\lim_{n\to\infty} \frac{R_Y(t/\sqrt{n})}{(t/\sqrt{n})^2} = 0$

note: as $n\to\infty$ $t/\sqrt{n}\to 0$ $(r=2,\ a=0)$

$$\left[\ \lim_{x\to a}\frac{g(x)-T_2(x)}{(x-a)^2} = 0 \quad \left(\text{2nd order Taylor's } r=2, a=0\right)\right]$$

$\rightarrow$ t fixed $\lim_{n\to\infty}\frac{R_Y(t/\sqrt{n})}{(1/\sqrt{n})^2} = \lim_{n\to\infty} n R_Y(t/\sqrt{n}) = t^2\cdot 0 = 0$

$\rightarrow$ Recall goal show $\lim_{n\to\infty} M_{\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}}(t) \longrightarrow e^{t^2/2}$

$$M_{\frac{\sqrt{n}(\bar{x}-\mu)}{\sigma}}(t) = \left(M_Y(t/\sqrt{n})\right)^n$$

For any fixed t $\lim_{n\to\infty} M_{\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}}(t) = \lim_{n\to\infty}\left[1 + \frac{(t/\sqrt{n})^2}{2!} + R_Y(t/\sqrt{n})\right]^n$

$$\rightarrow \lim_{n\to\infty}\left[1 + \frac{(t/\sqrt{n})^2}{2!} + R_Y(t/\sqrt{n})\right]^n = \lim_{n\to\infty}\left[1 + \frac{1}{n}\underbrace{\left[t^2/2 + nR_Y(t/\sqrt{n})\right]}_{a_n}\right]^n$$

$\lim_{n\to\infty} a_n = a = t^2/2$ Lemma 2.3.1

$\rightarrow$ By Lemma 2.3.1 $\lim_{n\to\infty}\left(1+\frac{a_n}{n}\right)^n = e^a$

$$\therefore \lim_{n\to\infty} M_{\frac{\sqrt{n}(\bar{x}-\mu)}{\sigma}}(t) = \lim_{n\to\infty}\left(M_Y(t/\sqrt{n})\right)^n = \lim_{n\to\infty}\left[1+\frac{1}{n}\left(t^2/2 + nR_Y(t/\sqrt{n})\right)\right]^n$$

$$= e^{t^2/2} = \text{mgf of } N(0,1) //$$

Q.E.D.

Theorem 5.5.15 (Stronger form of the Central Limit Theorem) *Let $X_1, X_2, \ldots$ be a sequence of iid random variables with* $\mathrm{E}X_i = \mu$ *and* $0 < \mathrm{Var}\, X_i = \sigma^2 < \infty$. *Define* $\bar{X}_n = (1/n)\sum_{i=1}^n X_i$. *Let* $G_n(x)$ *denote the cdf of* $\sqrt{n}(\bar{X}_n - \mu)/\sigma$. *Then, for any* x, $-\infty < x < \infty$,

$$\lim_{n\to\infty} G_n(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy;$$

that is, $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ *has a limiting standard normal distribution.*

- Note no requirement for mgf to exist for $|t| < \sigma h$
- Proof similar to previous proof
  - uses characteristic ft'n $E[e^{itX}]$, which always exists $i^2 = -1$
    Fourier transform vs. Laplace transform
  - Characteristic ft'n always exists.

+ CLT summary

  $X_1, X_2 \ldots$ iid with $E[X_i] = \mu$
  $\mathrm{Var}[X_i] = \sigma^2 < \infty$

  as $n \to \infty$

  $$\bar{X} \sim N(\mu, \sigma^2/n)$$

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

Example (Normal approximation to negative binomial)

$X_1, \dots, X_n$ iid neg bin (r, p). [# failures to r^{th} success.]

Negative binomial(r, p)

pmf	$P(X = x \mid r, p) = \binom{r+x-1}{x} p^r (1-p)^x; \quad x = 0, 1, \dots; \quad 0 \le p \le 1$
mean and variance	$\mathrm{E}X = \frac{r(1-p)}{p}, \quad \mathrm{Var}\,X = \frac{r(1-p)}{p^2}$
mgf	$M_X(t) = \left(\frac{p}{1-(1-p)e^t}\right)^r, \quad t < -\log(1-p)$
notes	An alternate form of the pmf is given by $P(Y = y \mid r, p) = \binom{y-1}{r-1} p^r (1-p)^{y-r}$, $y = r, r+1, \dots$. The random variable $Y = X + r$. The negative binomial can be derived as a gamma mixture of Poissons. (See Exercise 4.34.)

$E[X] = r(1-p)/p \qquad Var[X] = r(1-p)/p^2$

by CLT as $n \to \infty$ $\qquad \bar{X} \sim N\left[\frac{r(1-p)}{p}, \frac{r(1-p)}{p^2}\right]$

$$\frac{\bar{X} - r(1-p)/p}{\sqrt{\frac{r(1-p)}{p^2}}/\sqrt{n}} \sim N(0,1)$$

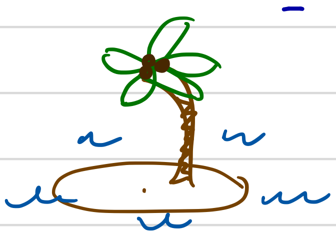
$$\frac{\sqrt{n}\,(\bar{X} - r(1-p)/p)}{\sqrt{r(1-p)/p^2}} \sim N(0,1)$$

- Assume $n=30$ trials with $r=10$, $p=1/2$
- Find $Pr(\bar{X} < 11) = P\left(\sum_{i=1}^{30} X_i \le 330\right)$

$$= \sum_{x=0}^{330} \binom{300+x-1}{x} \left(\tfrac{1}{2}\right)^{300} \left(\tfrac{1}{2}\right)^x$$

use mgf to show $\Sigma X_i \sim$ neg bin (nr, p)

$M_{\Sigma X_i}(t) = [M_X(t)]^n$



by hand (assuming live on small deserted island with no computer or internet) difficult calculation.

with R - not so tough...

$$\sum_{x=0}^{330} \binom{300+x-1}{x} \left(\tfrac{1}{2}\right)^{300} \left(\tfrac{1}{2}\right)^x$$

$= .8916$

by CLT $\quad P(\bar{X} \le 11) = P\left(\frac{\sqrt{30}(\bar{X}-10)}{\sqrt{20}} \le \frac{\sqrt{30}(11-10)}{\sqrt{20}}\right)$

$= P(Z \le 1.2247) = .8888$

Appendix: Taylor's Series Approximation

- Taylor's series = mathematical (vs. statistical) approximation to a function, using polynomials.

Definition 5.5.20 If a function $g(x)$ has derivatives of order r, that is, $g^{(r)}(x) = \frac{d^r}{dx^r}g(x)$ exists, then for any constant a, the *Taylor polynomial of order r about a* is

$$T_r(x) = \sum_{i=0}^{r} \frac{g^{(i)}(a)}{i!}(x-a)^i.$$

$T_1(x) = g(a) + g'(a)(x-a)$

$T_2(x) = g(a) + g'(a)(x-a) + g''(a)(x-a)^2/2!$

Taylor's major theorem, which we will not prove here, is that the *remainder* from the approximation, $g(x) - T_r(x)$, always tends to 0 faster than the highest-order explicit term.

Theorem 5.5.21 (Taylor) *If $g^{(r)}(a) = \frac{d^r}{dx^r}g(x)\big|_{x=a}$ exists, then*

$$\lim_{x\to a}\frac{g(x) - T_r(x)}{(x-a)^r} = 0.$$

In general, we will not be concerned with the explicit form of the remainder. Since we are interested in approximations, we are just going to ignore the remainder. There are, however, many explicit forms, one useful one being

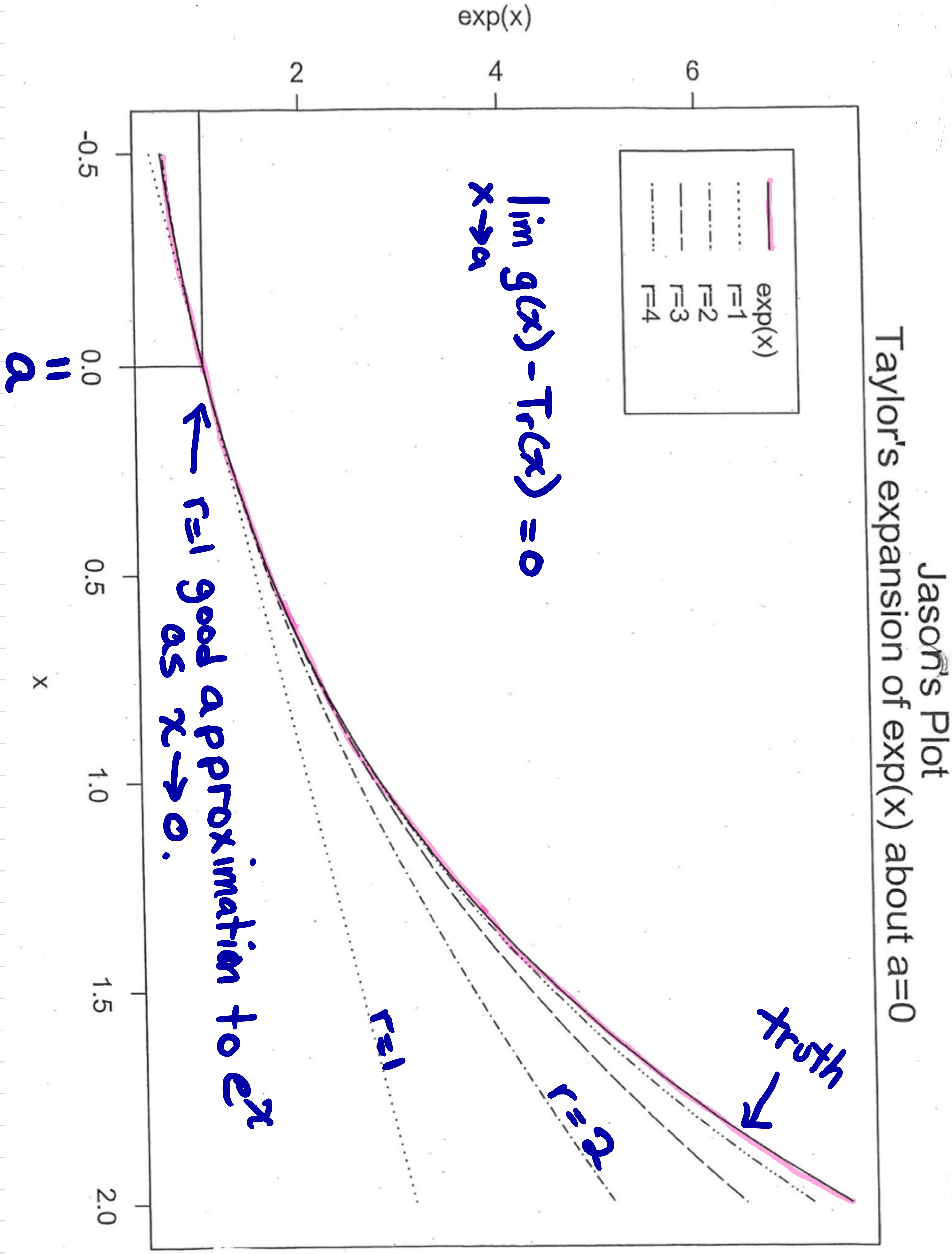
$$g(x) - T_r(x) = \int_a^x \frac{g^{(r+1)}(t)}{r!}(x-t)^r\,dt.$$

- $\lim_{x\to a} \frac{g(x) - T_r(x)}{(x-a)^r} = 0$
  - numerator going to zero faster than denominator as $x \to a$
  - numerator is remainder term (difference between function $g(x)$ & approximation)

- one form of remainder:

  $g(x) - T_r(x) = \int_a^x \frac{g^{(r+1)}(t)}{r!}(x-t)^r dt$

- For statistical applications usually use $r=1$ (1st order Taylor's approximation)
  CLT is exception (2nd order approx)


Jason's Plot
Taylor's expansion of exp(x) about a=0
exp(x)
r=1
r=2
r=3
r=4
2
4
6
-0.5
0.0
0.5
1.0
1.5
2.0
x
$\lim_{x \to a} g(x) - T_r(x) = 0$
= a
r=1 good approximation to e^x as $x \to 0$.
truth
r=1
r=2