ex. $E(X) = 1$, X non-negative $\rightarrow$

$\boxed{P(X \geq 100) \leq \frac{1}{100}}$ The inequality is

<u>sharp</u>, meaning that the upper bound

$\frac{E(X)}{t}$ on $P(X \geq t)$ is attainable, (*)

---

ex. $E(X) = 1$, X-nonnegative $\rightarrow$

put probability 0.99 on $X = 0$ and

0.01 $\quad$ 100

(*) but most of the time (i.e., for most distributions) it's a crude upper bound.

(30 May 19)

Can apply Markov inequality to the

r.v. $Y = [X - E(X)]^2$ to get

Chebyshev Inequality X r.v. with $V(X)$ existing $\rightarrow$ for every $t \geq 0$,

$$P\left[\,|X - E(X)| \geq t\,\right] \leq \frac{V(X)}{t^2}$$

(attributed to Pafnuty Chebyshev (1821 - 1894), also a Russian mathematician, one of whose Ph.D. students was Markov)

Ex. $E(X) = \mu$, $V(X) = \sigma^2$

Chebyshev says $P\left[\,\left|\frac{X-\mu}{\sigma}\right| \geq 3\,\right] \leq \frac{1}{3^2} = \frac{1}{9}$,

so no more than $\frac{1}{9} = 11\%$ of the probability in any distribution with finite variance can be more than 3 SDs away from the mean (recall for Normal dist. this prob. is 0.3%)

This upper bound is also sharp, but for most distributions it's (also) crude (as with the Markov bound).

Back to $\bar{X}_n$

$X_i \overset{IID}{\sim}$ some dist. with mean $E(X_i) = \mu$ $(i = 1, \dots, n)$ and variance $V(X_i) = \sigma^2 < \infty$

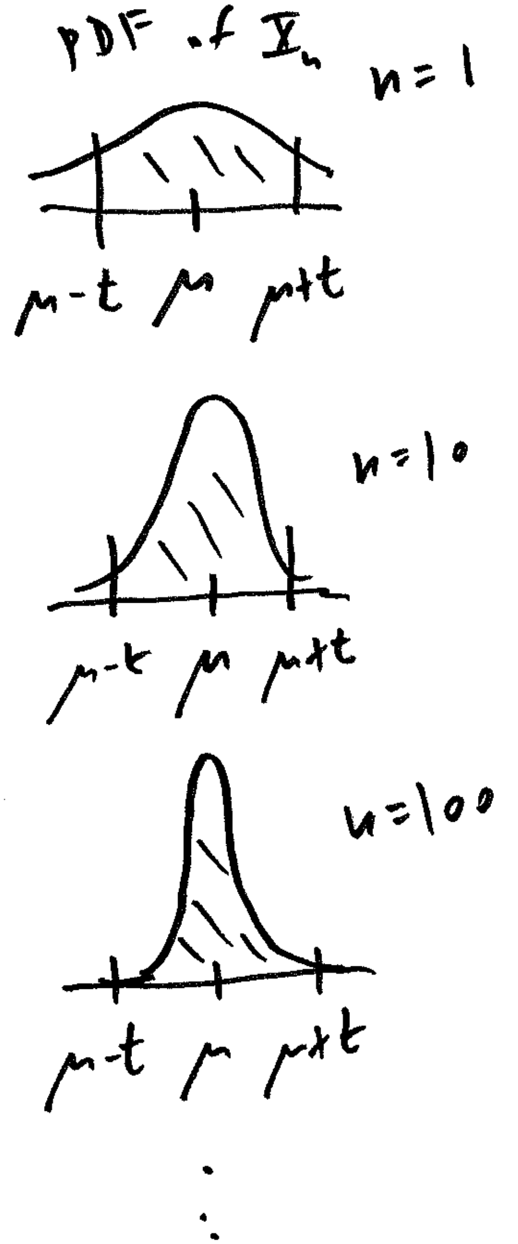
We have already shown that if $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

then $E(\bar{X}_n) = \mu$ for all $n = 1, 2, \dots$

and $V(\bar{X}_n) = \frac{\sigma^2}{n}$.

Chebyshev then gives $P\left(|\bar{X}_n - \mu| \geq t\right) \leq \frac{\sigma^2}{n t^2}$ for all $t > 0$

This can be rewritten $P\left(|\bar{X}_n - \mu| < t\right) \geq 1 - \frac{\sigma^2}{n t^2}$



This suggests a way to quantify how close a r.v. like $\bar{X}_n$ is to a constant like μ:

$\boxed{\text{Def.}}$ A sequence $Z_1, Z_2, \ldots$ of r.v. is said to converge in probability to a constant b if

for all $\epsilon > 0$, $\lim_{n \to \infty} P(|Z_n - b| < \epsilon) = 1$;

this is denoted $Z_n \xrightarrow{P} b$.

An immediate consequence of Chebyshev &

This definition is

305

(weak) Law of Large Numbers

$X_i \stackrel{IID}{\sim}$ a dist. with mean μ and variance $\sigma^2 < \infty$, $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ ($\bar{X}_n$ is consistent for μ)

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

This result has a long history: the Italian mathematician Gerolamo Cardano (1501–1576) asserted it without proof; Jacob Bernoulli (1655–1705) proved it for $(X_i \mid \theta) \stackrel{IID}{\sim}$ Bernoulli (θ) (it took him 20 years to find a correct proof, published posthumously in 1713; Bernoulli thought that this theorem proved the existence of God); Siméon Denis Poisson named it the Law of Large Numbers in 1837.

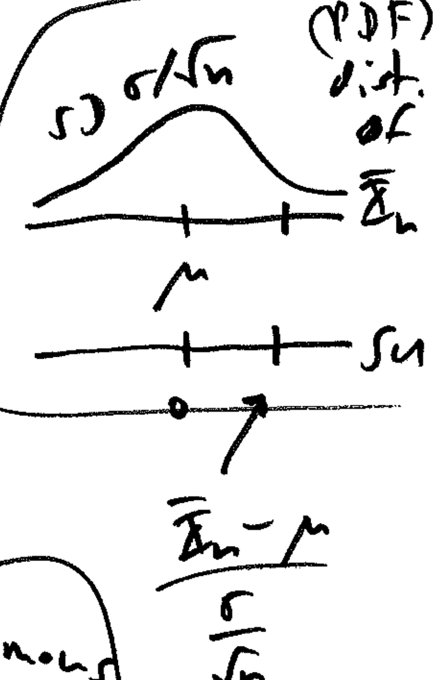
$\boxed{\text{Corollary}}$ If $Z_n \xrightarrow{P} b$ and $g(z)$ is continuous at $z = b$ then $g(Z_n) \xrightarrow{P} g(b)$

Central Limit Theorem (CLT)

Example $X_i \overset{IID}{\sim} N(\mu, \sigma^2)$, $\sigma < \infty$ $(i = 1, \dots, n)$

we know that $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$ has mean μ, variance $\frac{\sigma^2}{n}$ and is normally distributed, so that

$$\frac{\bar{X}_n - \mu}{\sigma / \sqrt{n}} \sim N(0, 1) \quad \text{for all } n = 1, 2, \dots$$



$\boxed{Q:}$ Does something like this work for other choices of $X_i \overset{IID}{\sim} \boxed{?}$?

$\boxed{A:}$ Yes: it's the most famous result in all of probability:

Central Limit (CLT) Theorem $X_i \overset{IID}{\sim}$ any dist. with mean μ and finite variance $0 < \sigma^2 < \infty$,

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \rightarrow \text{for large } n \quad \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \,\dot{\sim}\, N(0, 1)$$

Careful statement | **Def.** $X_1, X_2, \ldots$ a sequence of r.v.; let F_n be the CDF of X_n + if there exists a CDF F^* such that $\lim_{n \to \infty} F_n(x) = F^*(x)$ for all x at which $F^*(x)$ is continuous, then people say that $X_n \xrightarrow{D} F^*$ ("X_n converges in distribution to F^*")

CLT | $X_i \overset{IID}{\sim}$ (any) dist. with mean μ and variance $0 < \sigma^2 < \infty$, $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$

$$\rightarrow \sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right) \xrightarrow{D} N(0,1).$$

The CLT also has a long history: it was

first demonstrated for $X_i \overset{IID}{\sim}$ Bernoulli(θ) by the French/British mathematician Abraham de Moivre (1667 - 1754) in 1733; almost forgotten until revived by the French mathematician Pierre-Simon de Laplace (1749 - 1827) in 1812; almost forgotten again until 1901, when the Russian mathematician Aleksandr Lyapunov gave a more general proof; even more general proof provided by J W Lindeberg (Finnish mathematician (1876 - 1932)) and independently by Paul Lévy (French mathematician (1886-1971)) in the early 1920s. CLT name due to Hungarian-American mathematician (1887-1985) George Pólya in 1920

Example: Contaminated water supply:

X = arsenic concentration

Y = lead concentration

(same units) (both ≥ 0)

Interest focuses on $R = \frac{Y}{X+Y}$

(proportion of contamination due to lead)

$E(R) = E\left(\frac{Y}{X+Y}\right)$ difficult to calculate.

Simulation approach: Randomly sample n pairs (X_i, Y_i) from the joint PDF of (X, Y), calculate $R_i = \frac{Y_i}{X_i + Y_i}$ and

$$\boxed{\bar{R}_n = \frac{1}{n} \sum_{i=1}^{n} R_i} \leftarrow \text{good } \underline{\text{Monte Carlo}}$$

(simulation) estimate of $E(R)$.

Q: How big does n need to be to achieve a desired accuracy target?

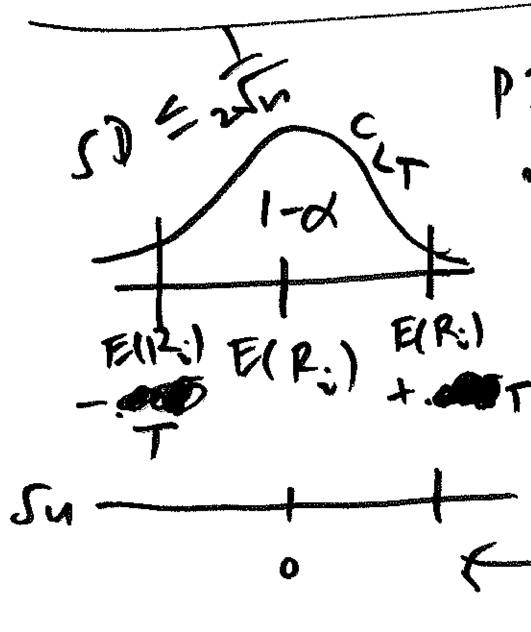
By definition

$$|R_i| = \left| \frac{I_i}{\Sigma_i + I_i} \right| \le 1 \text{; can show that}$$

as a result $V(R_i) \le \frac{1}{4}$.

CLT says that dist. of $\bar{R}_n$ will be close to Normal for large n, with mean $E(R_i)$ and variance $\frac{V(R_i)}{n} \le \frac{1}{4n}$

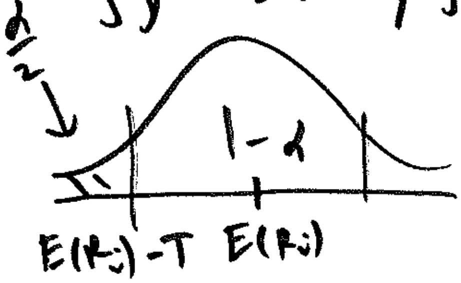
Suppose we want $\bar{R}_n$ to differ from $E(R_i)$ by no more than some tolerance T with probability at least $(1-\alpha)$ ...



PDF of $\bar{R}_n$, n large

$\leftarrow \frac{T}{SD}$

$SD \le \frac{1}{2\sqrt{n}}$, so $\frac{1}{SD} \ge 2\sqrt{n}$ and



$$\frac{-T}{SD} \le 2T\sqrt{n}$$

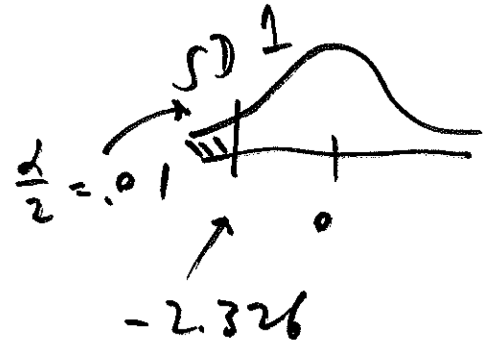
$$\Phi^{-1}\left(\frac{\alpha}{2}\right) = \frac{[E(R_i)-T]-E(R_i)}{SD} = \frac{-T}{SD} \le 2T\sqrt{n}$$

from which $n \ge \left[\dfrac{\Phi^{-1}\left(\frac{\alpha}{2}\right)}{2T}\right]^2$

For instance set $T = 0.005$ ($\frac{1}{2}$ of 1%)

and $\alpha = .02$ to get

$$n \ge \left[\frac{-2.326}{2(.005)}\right]^2 \doteq 54{,}119$$



simulation replications needed

Case study: Escalators in the London Underground

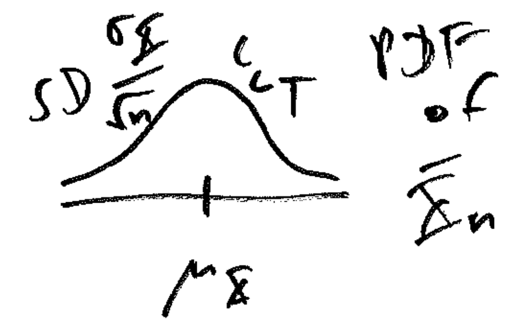
the Delta method

The CLT says that if $X_i \overset{IID}{\sim}$ any dist. with finite mean μ_X and finite variance σ_X^2, then the distribution of

$$\frac{\bar{X}_n - \mu_X}{\sigma_X / \sqrt{n}}$$

for large n is approximately standard normal, where $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$.

This is equivalent to saying that



$$\bar{X}_n \,\dot{\sim}\, N\left(\mu_X, \frac{\sigma_X^2}{n}\right)$$

---

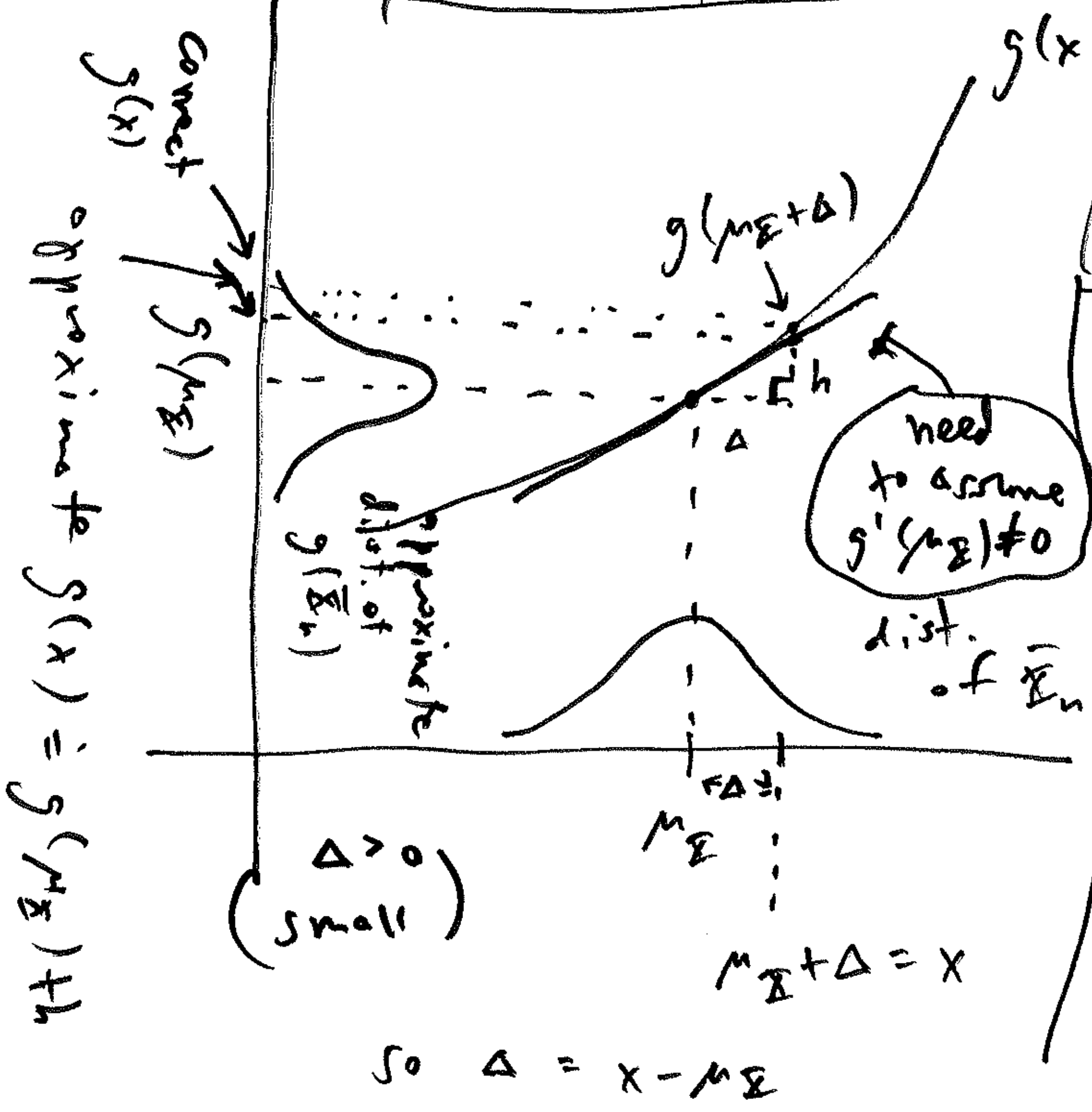
question: If $g(x)$ is a sufficiently "nice" function, is there a comparable result for $g(\bar{X}_n)$?

Answer: Yes, via a Taylor-series-based approach called the Delta method

$\bar{X}_n$ should be close to μ_X for large n (that's the (weak) Law of Large Numbers); this suggests making a two-term Taylor expansion of $g(\bar{X}_n)$ around the point

$$x = \mu_X : \quad g(\bar{X}_n) \doteq g(\mu_X) + g'(\mu_X)(\bar{X}_n - \mu_X)$$

this is why it's called the Δ (Delta)-method



$$\frac{h}{\Delta} = g'(\mu_X)$$

so

$$g(x) \doteq g(\mu_X) + h$$

$$= g(\mu_X) + g'(\mu_X) \cdot \Delta$$

$$= g(\mu_X) + g'(\mu_X)(x - \mu_X)$$

$$g(\bar{X}_n) \doteq g(\mu_X) + g'(\mu_X)(\bar{X}_n - \mu_X) \text{ so}$$

constant

r.v.

$$E[g(\bar{X}_n)] \doteq E\left[g(\mu_X) + g'(\mu_X)(\bar{X}_n - \mu_X)\right]$$

$$= g(\mu_X) + g'(\mu_X)\left[\cancel{E(\bar{X}_n) - \mu_X}\right] \quad 0$$

$$\text{so } E[g(\bar{X}_n)] \doteq g(\mu_X) = g[E(\bar{X}_n)]$$

and

constant

$$V[g(\bar{X}_n)] \doteq V\left[g(\mu_X) + g'(\mu_X)(\bar{X}_n - \mu_X)\right]$$

r.v.

$$= [g'(\mu_X)]^2 \cdot V(\bar{X}_n - \mu_X)$$

$$\text{so } V[g(\bar{X}_n)] \doteq [g'(\mu_X)]^2 V(\bar{X}_n),$$

$$\text{i.e., } V[g(\bar{X}_n)] \doteq [g'(\mu_X)]^2 \frac{\sigma_X^2}{n}$$

There's one hidden assumption in this calculation: $g'(\mu_{\bar{X}}) \neq 0$.

This works for <u>any</u> r.v. with finite variance, not just $\bar{X}_n$:

V any r.v. with finite variance σ_V^2 (and therefore finite mean μ_V), $W = g(V)$

$\rightarrow$ $E(W) \doteq g(\mu_V)$ and

$V(W) \doteq [g'(\mu_V)]^2 \sigma_V^2$,

$\boxed{\Delta \text{ method}}$ part 1

provided $g'(v)$ is continuous and

$g'(\mu_V) \neq 0$

Moreover, if V is Normal then $W = g(V)$ is Normal also

$\boxed{\Delta \text{ method}}$ part 2

Example A bank typically has a single queue (line) at which customers arrive to transact banking business.

Let X_i = time customer i waits from reaching the head of the queue until served. To be completely realistic, the dist. of X_i would vary by day of week and time of day, so pick a single time slot (e.g. Tue 10-10.15am) and observe the X_i from week to week only in that time slot; now the $\{X_i, i = 1, 2, \dots\}$ form a stationary stochastic process with fixed (non-time-varying), finite $E(X_i) = \mu_X > 0$

and fixed (non-time-varying) finite $V(\underline{X}_i) = \sigma^2_X$.

Gather data over many weeks and form $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ for large n.

complication: seasonal effects (ignored here)

The rate of service is defined to be $g(\mu_X) = \frac{1}{\mu_X}$, which would naturally be estimated by $\underline{\underline{g}}(\bar{X}_n) = \frac{1}{\bar{X}_n}$.

$E(\bar{X}_n) = \mu_X$

$V(\bar{X}_n) = \frac{\sigma^2_X}{n}$

$\boxed{g(x) = \frac{1}{x} = x^{-1}}$ $g'(x) = -\frac{1}{x^2}$

$\boxed{g'(\mu_X) = -\frac{1}{\mu^2_X}}$

$\bar{X}_n \,\dot\sim$ Normal by CLT

so Δ-method says $g(\bar{X}_n) = \frac{1}{\bar{X}_n} \,\dot\sim$ Normal also, with mean $g(\mu_X) = \frac{1}{\mu_X}$ and variance $\sigma^2_X / (n \mu^4_X)$

$[g'(\mu_X)]^2 = \frac{1}{\mu_X^4} \neq 0$

318

Specific calculation | Under some plausible assumptions, we've seen that $(X_i \mid \lambda) \overset{IID}{\sim} \text{Exponential}(\lambda)$

may be a reasonable model for waiting times.

$E(X_i) = \frac{1}{\lambda} = \mu_X$, $V(X_i) = \frac{1}{\lambda^2} = \sigma_X^2$

$(X_i \mid \lambda)$ has PDF

$f_{X_i}(x_i \mid \lambda) = \lambda e^{-\lambda x_i} I(x_i > 0)$

so $\frac{1}{\bar{X}_n}$ should (for large n) be approximately Normal with mean $\frac{1}{\frac{1}{\lambda}} = \lambda$

and SD $\frac{\sigma_X}{\mu_X^2 \sqrt{n}} = \frac{\frac{1}{\lambda}}{\left(\frac{1}{\lambda}\right)^2 \sqrt{n}} = \frac{\lambda}{\sqrt{n}}$.

Fancy version of Δ-method | $Y_1, Y_2, \dots$ sequence of (discrete or continuous) r.v.; F^* continuous CDF;

θ a real number; $a_1, a_2, \dots \uparrow \infty$ positive sequence

$g(\cdot)$ a real-valued function of a real variable such that $g'(\cdot)$ is continuous and $g'(\theta) \neq 0$; then if $a_n(Y_n - \theta) \xrightarrow{D} F^*$,

$$a_n \left[\frac{g(Y_n) - g(\theta)}{|g'(\theta)|} \right] \xrightarrow{D} F^*$$

also

Typical application: $X_1, X_2, \ldots$ IID

$$Y_n = \bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \,; \quad \theta = \mu_X \,; \quad a_n = \frac{\sqrt{n}}{\sigma_X} \,;$$

$F^* = \Phi$, the standard normal CDF.

In this context the theorem says that

$$\text{if } \frac{\bar{X}_n - \mu_X}{\sigma_X / \sqrt{n}} \mathrel{\dot\sim} N(0,1) \xrightarrow{\text{then}} \frac{g(\bar{X}_n) - g(\mu_X)}{|g'(\mu_X)|\, \sigma_X / \sqrt{n}}$$

is also $\mathrel{\dot\sim} N(0,1)$

(28 Aug 17)

~~(20 Aug 09)~~

A little bit more about the continuity correction

Tay-Sachs case study, revisited

X = # T-S babies in family of $n = 5$ children, both parents carriers so that

$P(\text{T-S baby}) = \frac{1}{4} = p$ $\quad$ $\boxed{X \sim \text{Binomial}(n, p)}$

But also let $T_i = \begin{Bmatrix} 1 & \text{if child } i \text{ is T-S baby} \\ 0 & \text{else} \end{Bmatrix}$ $\quad i = 1, \dots, n = 5$

Then $(T_i) \overset{\text{IID}}{\sim} \text{Bernoulli}(p)$ $(i = 1, \dots, n)$ and $X = \sum_{i=1}^{n} T_i$

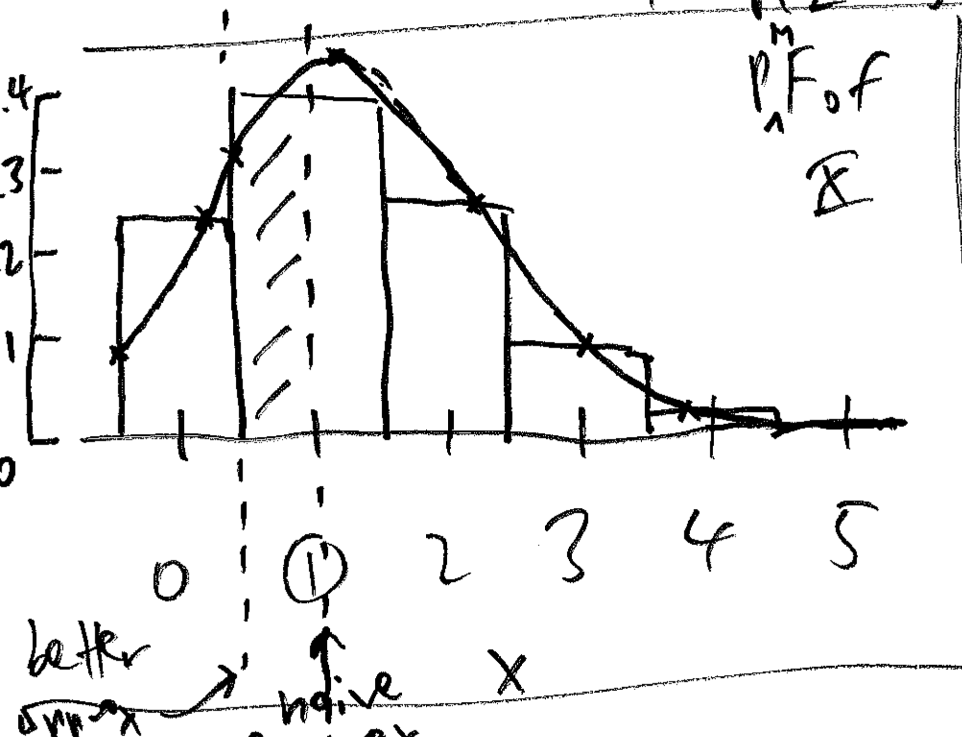
So by the CLT the dist. of X should be approximately Normal with mean

$\mu_X = E(X) = np = 1.25$ and SD

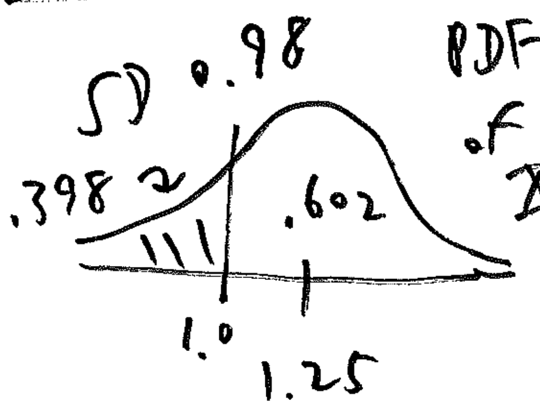
$$\sigma_{X} = \sqrt{V(X)} = \sqrt{np(1-p)} \doteq 0.98$$

On day 1 of this class we worked out that $P(\text{1 or more T-S babies}) = P(X \geq 1)$

$$1 - P(\text{no T-S babies}) = 1 - (1-p)^n \doteq 0.76$$
$$= 1 - P(X=0)$$



Naive Normal approximation, from CLT:



$$\frac{1.0 - 1.25}{0.98} \doteq -0.26$$

$$P(X \geq 1) \doteq 1 - P(X' < 1)$$

$$= 1 - 0.398$$

$$\doteq 0.602$$ (quite a bad approximation)