powerful math tricks for dealing with distributions

we won't do much with it, but mentioned/used in book, so a very brief introduction:

the k^{th} moment of r.v. X is $E[X^k]$; M.G.F. is $M(t) = E[e^{tX}]$

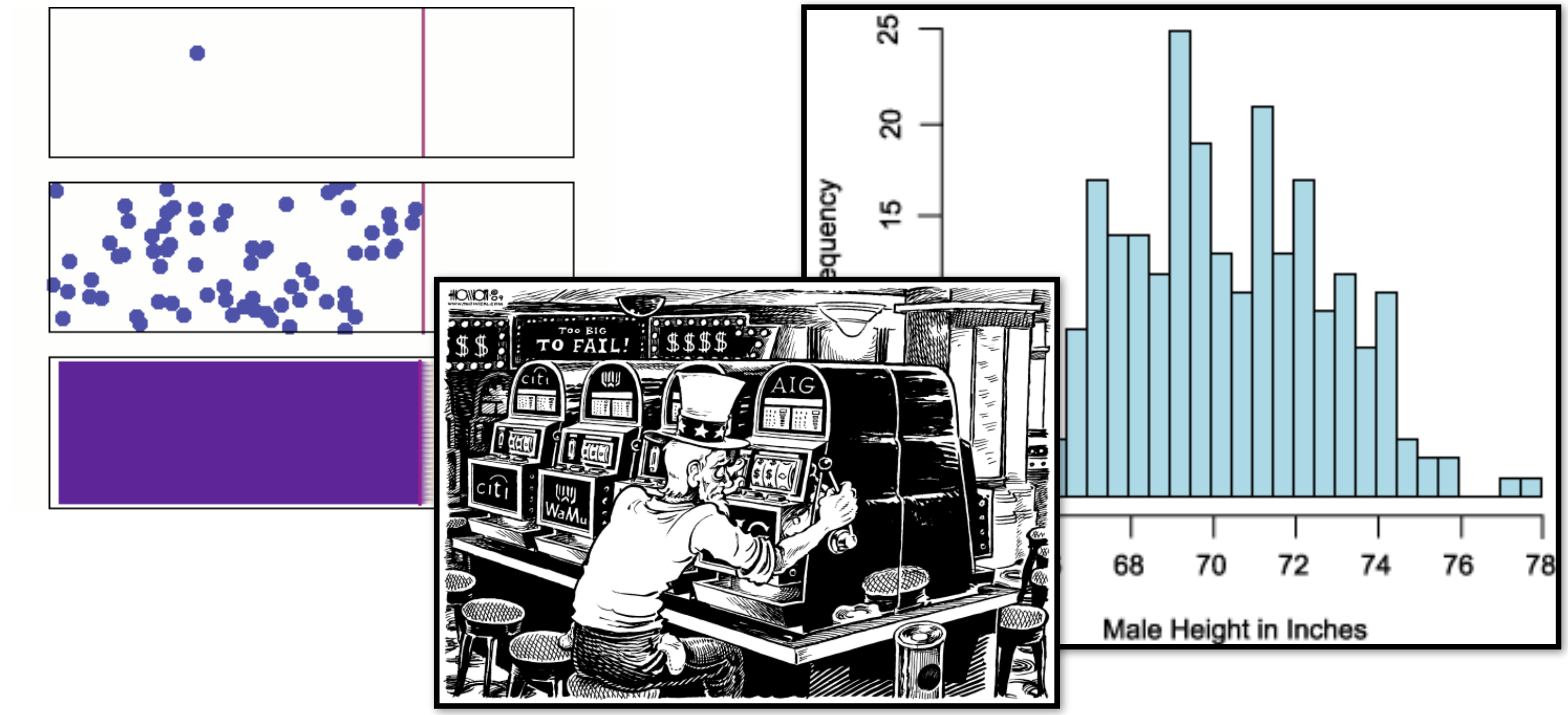
$$e^{tX} = X^0 \frac{t^0}{0!} + X^1 \frac{t^1}{1!} + X^2 \frac{t^2}{2!} + X^3 \frac{t^3}{3!} + \cdots$$

$$M(t) = E[e^{tX}] = E[X^0] \frac{t^0}{1!} + E[X^1] \frac{t^1}{1!} + E[X^2] \frac{t^2}{2!} + E[X^3] \frac{t^3}{3!} +$$

$$\frac{d^2}{dt^2} M(t) = 0 + 0 + E[X^2] + E[X^3] \cdot \frac{t^1}{1!} + \cdots$$

$$\boxed{\left.\frac{d^2}{dt^2} M(t)\right|_0 = E[X^2]} \qquad \boxed{\left.\frac{d^k}{dt^k} M(t)\right|_0 = E[X^k]}$$

the law of large numbers & the CLT



$$\Pr\left(\lim_{n\to\infty}\left(\frac{X_1+\cdots+X_n}{n}\right)=\mu\right)=1$$

i.i.d. (independent, identically distributed) random vars

$X_1, X_2, X_3, \ldots$

X_i has $\mu = E[X_i] < \infty$ and $\sigma^2 = Var[X_i]$

Consider the *empirical mean*: $$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

The Weak Law of Large Numbers:

For any $\varepsilon > 0$, as $n \to \infty$

$$\Pr(|\overline{X} - \mu| > \epsilon) \longrightarrow 0.$$

For any ε > 0, as n → ∞

$$\Pr(|\overline{X} - \mu| > \epsilon) \longrightarrow 0.$$

Proof: (*assume* $\sigma^2 < \infty$)

$$E[\overline{X}] = E[\frac{X_1 + \cdots + X_n}{n}] = \mu$$

$$\mathrm{Var}[\overline{X}] = \mathrm{Var}[\frac{X_1 + \cdots + X_n}{n}] = \frac{\sigma^2}{n}$$

By Chebyshev inequality,

$$\Pr(|\overline{X} - \mu| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2} \longrightarrow 0$$

i.i.d. (independent, identically distributed) random vars

$X_1, X_2, X_3, \ldots$

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

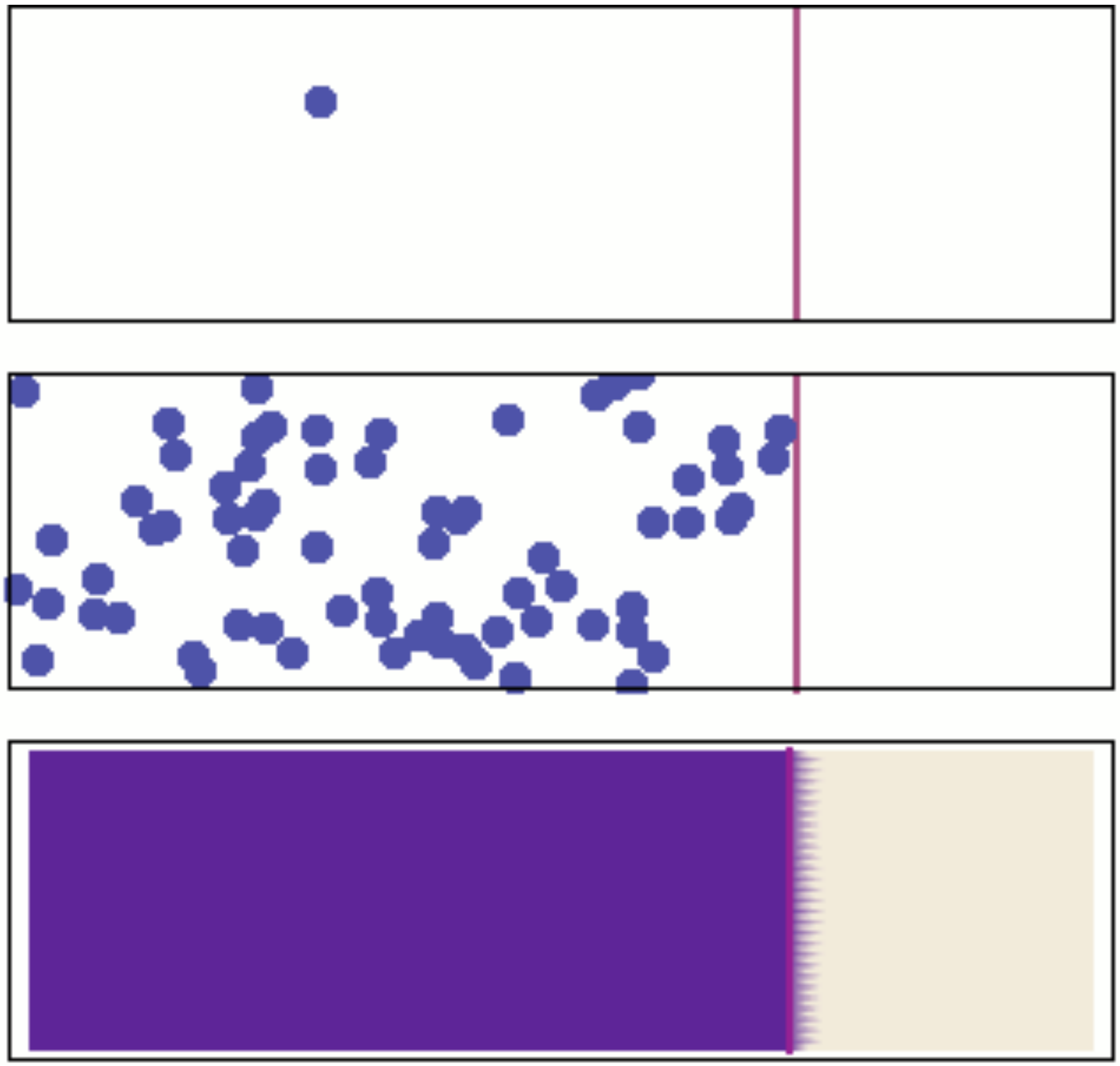
X_i has $\mu = E[X_i] < \infty$

$$\Pr\left(\lim_{n\to\infty}\left(\frac{X_1+\cdots+X_n}{n}\right)=\mu\right)=1$$

Strong Law ⇒ Weak Law (but not vice versa)

Strong law implies that for any ε > 0, there are only finite number of n such that the weak law condition $|\overline{X}-\mu| \geq \epsilon$ is violated.

diffusion



http://en.wikipedia.org/wiki/Law_of_large_numbers

the law of large numbers

Note: $D_n = E[\ |\ \Sigma_{1 \leq i \leq n}(X_i - \mu)\ |\ ]$ *grows* with n, but $D_n/n \to 0$

Justifies the "frequency" interpretation of probability

"Regression toward the mean"

Gambler's fallacy: "I'm *due* for a win!"

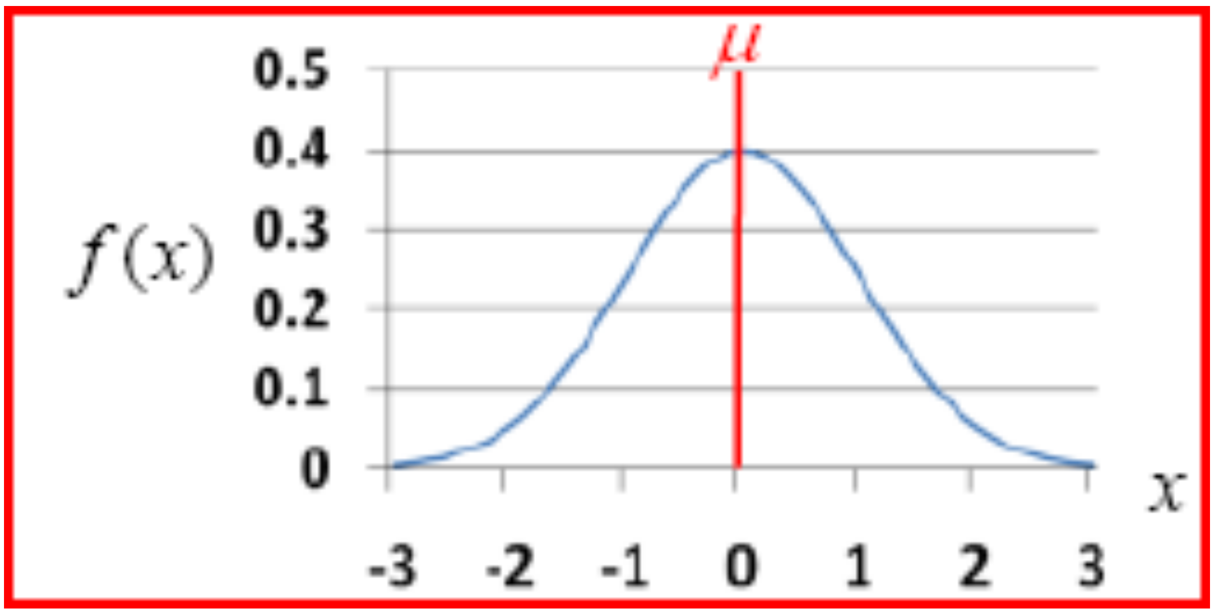
"Result will usually be close to the mean"

Many web demos, e.g.
http://stat-www.berkeley.edu/~stark/Java/Html/lln.htm

- X is a normal random variable $X \sim N(\mu,\sigma^2)$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

$$E[X] = \mu \qquad \mathrm{Var}[X] = \sigma^2$$



X is a normal random variable $X \sim N(\mu,\sigma^2)$

$Z \sim N(0,1)$ "standard (or unit) normal"

Use $\Phi(z)$ to denote CDF, i.e.

$$\Phi(z) = \Pr(Z \le z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx$$

no closed form ☹

the central limit theorem (CLT)

i.i.d. (independent, identically distributed) random vars

$X_1, X_2, X_3, \ldots$

X_i has $\mu = E[X_i]$ and $\sigma^2 = Var[X_i]$

As n → I,

$$\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \longrightarrow N(0,1)$$

Restated: As n → ∞,

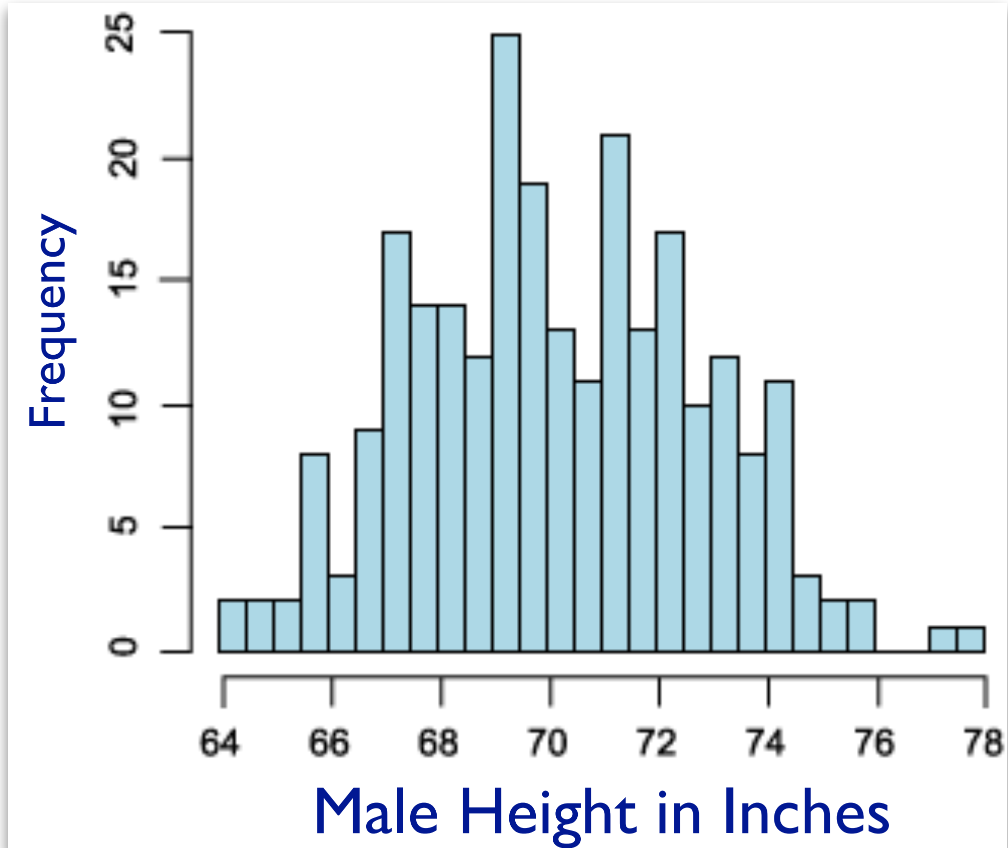
$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

CLT in the real world

CLT is the reason many things appear normally distributed
Many quantities = sums of (roughly) independent random vars

Exam scores: sums of individual problems
People's heights: sum of many genetic & environmental factors
Measurements: sums of various small instrument errors
...

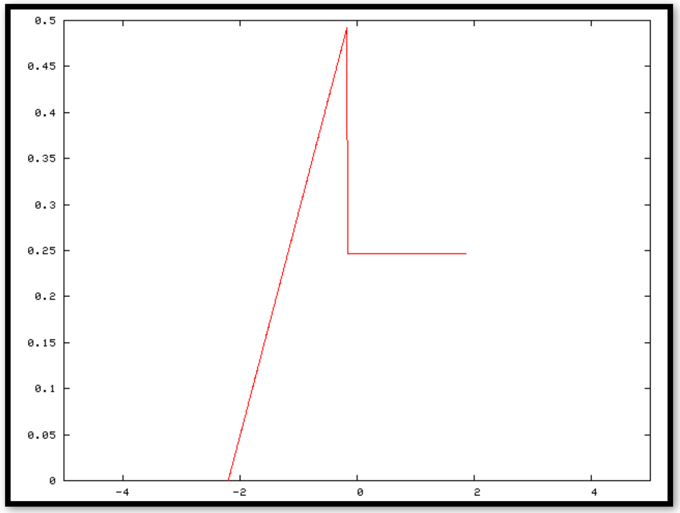
Human height is approximately normal.

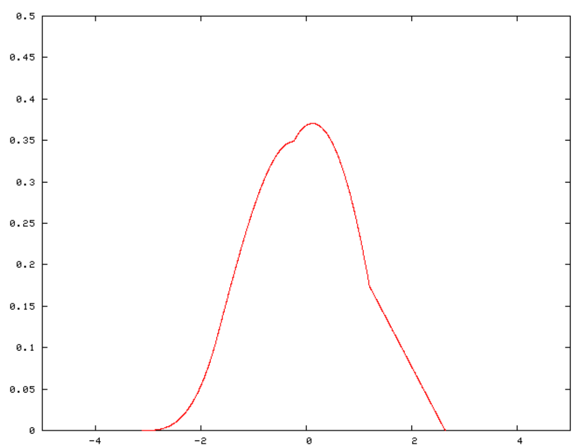


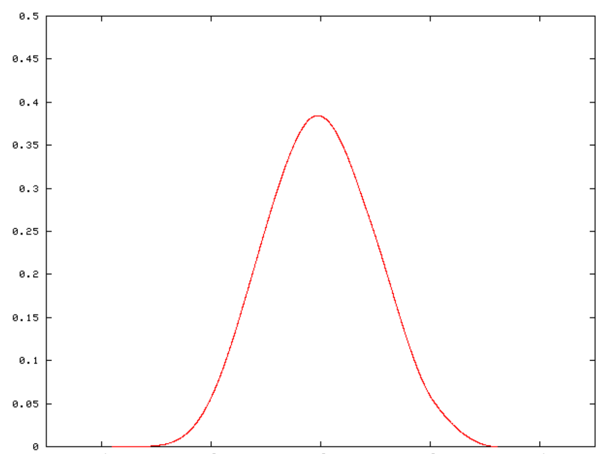
Why might that be true?

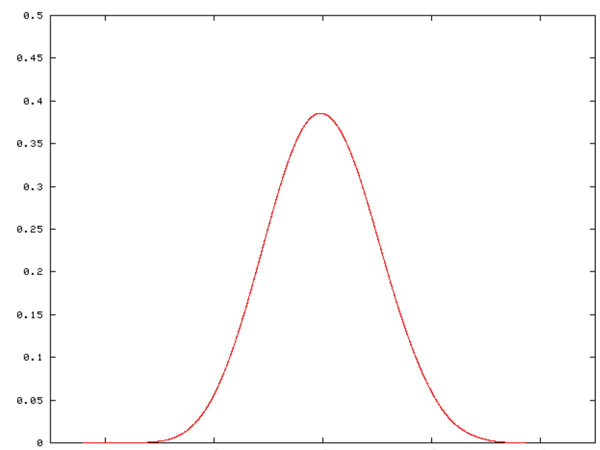
R.A. Fisher (1918) noted it would follow from CLT if height were the sum of many independent random effects, e.g. many genetic factors (plus some environmental ones like diet). I.e., suggested part of *mechanism* by looking at *shape* of the curve.

CLT convergence









Chernoff bound was CLT (for binomial) in disguise

- Suppose X ~ Bin(n,p)
- μ = E[X] = pn

- Chernoff bound:

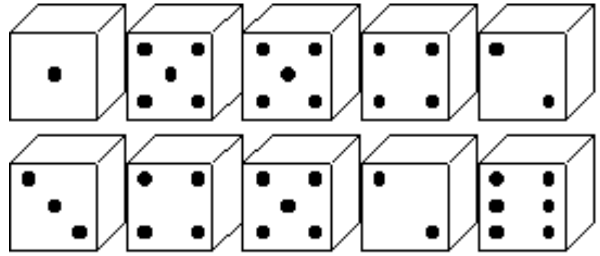
For any δ with $0 < \delta < 1$,

$$P(X > (1+\delta)\mu) \leq e^{-\frac{\delta^2 \mu}{2}}$$

$$P(X < (1-\delta)\mu) \leq e^{-\frac{\delta^2 \mu}{3}}$$

rolling more dice

- Roll 10 6-sided dice



- X = total value of all 10 dice
- Win if: X ≤ 25 or X ≥ 45
- Roll…

$$E[X] = 10E[X_i] = 10(3.5) = 35 \qquad \mathrm{Var}(X) = 10\,\mathrm{Var}(X_i) = 10\frac{35}{12} = \frac{350}{12}$$

$$1 - P(25.5 \le X \le 44.5) = 1 - P\left(\frac{25.5-35}{\sqrt{350/12}} \le \frac{X-35}{\sqrt{350/12}} \le \frac{44.5-35}{\sqrt{350/12}}\right)$$

$$\approx 1 - (2\Phi(1.76) - 1) \approx 2(1 - 0.9608) = 0.0784$$