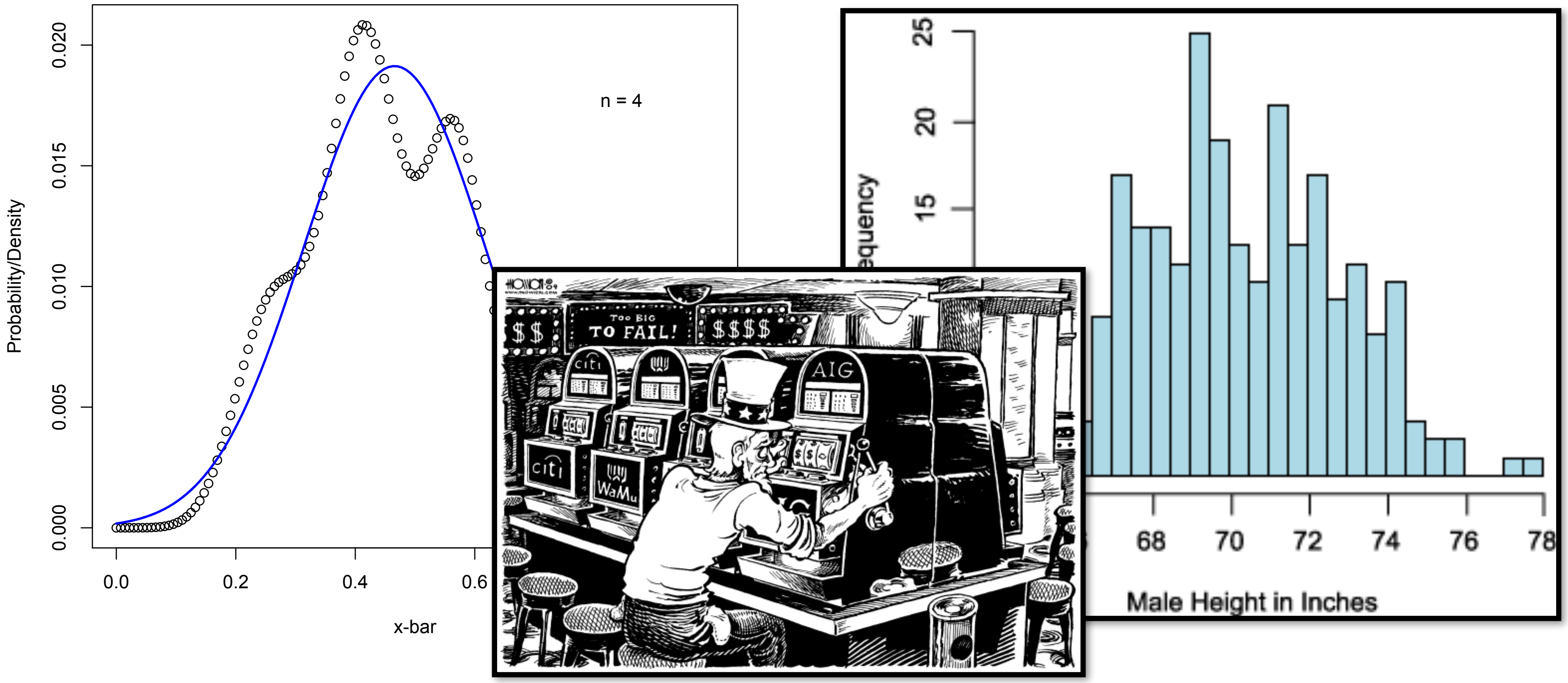
the law of large numbers & the CLT



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

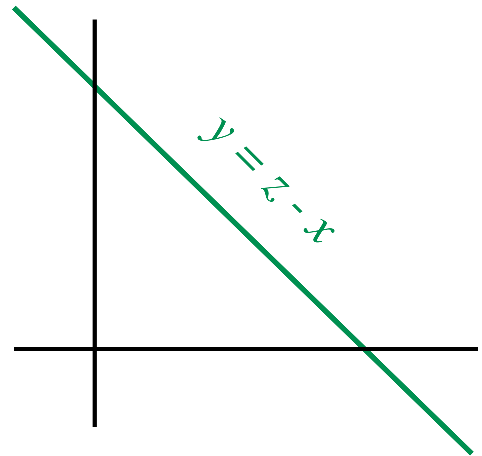
If X,Y are independent, what is the distribution of $Z = X + Y$?

Discrete case:

$$p_Z(z) = \Sigma_x \, p_X(x) \bullet p_Y(z-x)$$

Continuous case:

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x) \bullet f_Y(z-x) \, dx$$

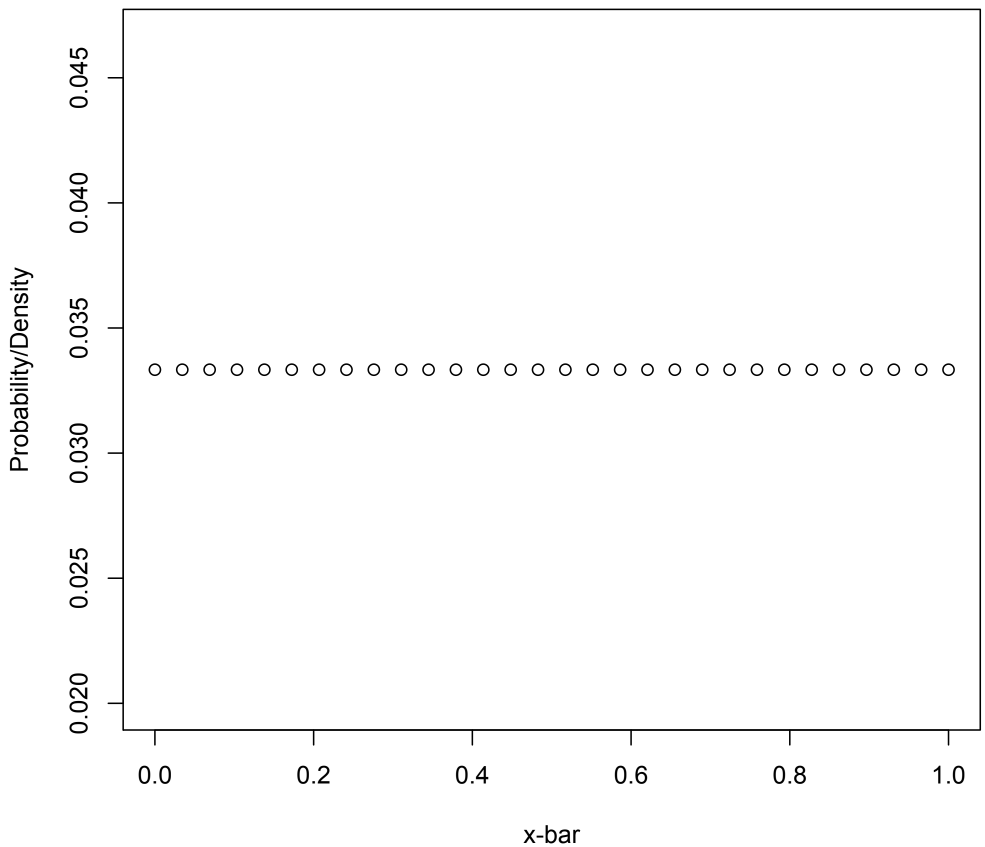


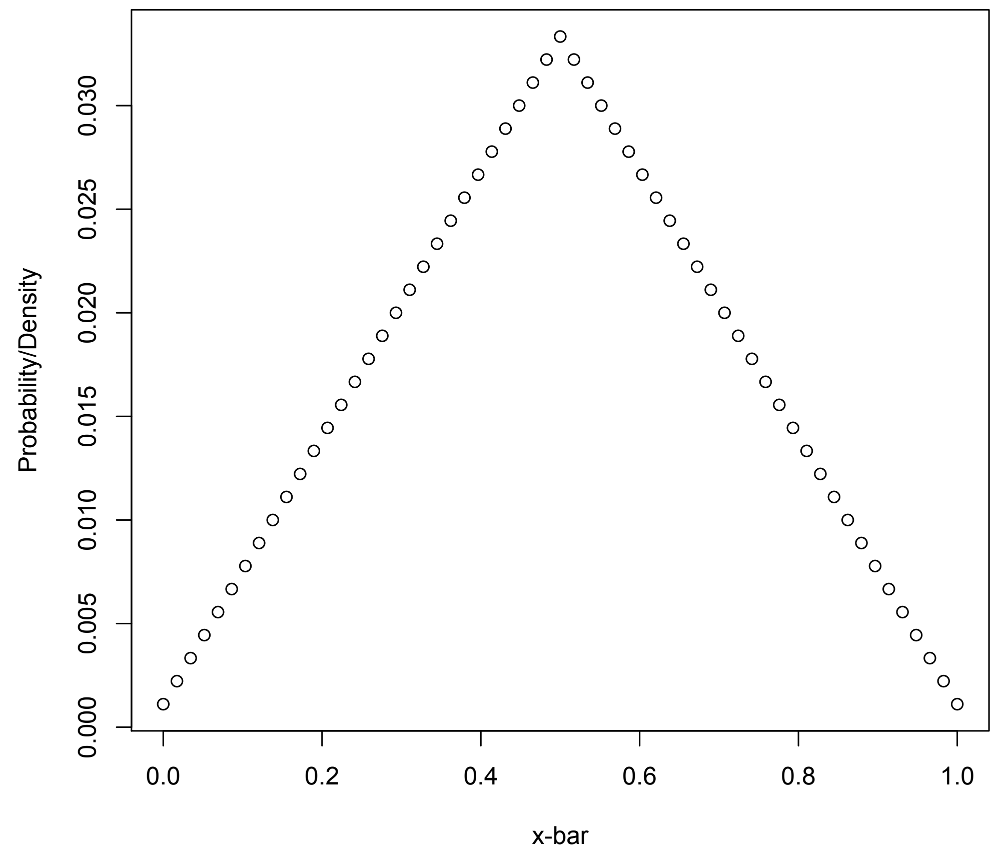
E.g. what is the p.d.f. of the sum of 2 normal RV's?

$W = X + Y + Z$? Similar, but double sums/integrals

$V = W + X + Y + Z$? Similar, but triple sums/integrals

If X and Y are *uniform*, then Z = X + Y is *not*; it's *triangular* (like dice):





Intuition: X + Y ≈ 0 or ≈ 1 is rare, but many ways to get X + Y ≈ 0.5

aka transforms; b&t 229

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 $E[X^k]$; $\boxed{\text{M.G.F. is } M(t) = E[e^{tX}]}$

$$
\begin{aligned}
e^{tX} &= X^0\frac{t^0}{0!} + X^1\frac{t^1}{1!} + X^2\frac{t^2}{2!} + X^3\frac{t^3}{3!} + \cdots \\
M(t) = E[e^{tX}] &= E[X^0]\frac{t^0}{0!} + E[X^1]\frac{t^1}{1!} + E[X^2]\frac{t^2}{2!} + E[X^3]\frac{t^3}{3!} + \cdots \\
\frac{d}{dt}M(t) &= 0 + E[X^1] + E[X^2]\frac{t^1}{1!} + E[X^3]\frac{t^2}{2!} + \cdots \\
\frac{d^2}{dt^2}M(t) &= 0 + 0 + E[X^2] + E[X^3]\frac{t^1}{1!} + \cdots
\end{aligned}
$$

$$
\boxed{\left.\frac{d}{dt}M(t)\right|_{t=0} = E[X]} \quad \boxed{\left.\frac{d^2}{dt^2}M(t)\right|_{t=0} = E[X^2]} \quad \ldots \quad \boxed{\left.\frac{d^k}{dt^k}M(t)\right|_{t=0} = E[X^k]} \quad \ldots
$$

An example:

MGF of normal(μ,σ^2) is $\exp(\mu t+\sigma^2 t^2/2)$

Two key properties:

1. MGF of *sum* independent r.v.s is *product* of MGFs:

$$M_{X+Y}(t) = E[e^{t(X+Y)}] = E[e^{tX} e^{tY}] = E[e^{tX}]\, E[e^{tY}] = M_X(t)\, M_Y(t)$$

2. Invertibility: MGF uniquely determines the distribution.

e.g.: $M_X(t) = \exp(at+bt^2)$, with $b>0$, then $X \sim \text{Normal}(a,2b)$

Important example: *sum of independent normals is normal:*

$$X\sim\text{Normal}(\mu_1,\sigma_1^2) \quad Y\sim\text{Normal}(\mu_2,\sigma_2^2)$$

$$M_{X+Y}(t) = \exp(\mu_1 t + \sigma_1^2 t^2/2) \bullet \exp(\mu_2 t + \sigma_2^2 t^2/2)$$

$$= \exp[(\mu_1+\mu_2)t + (\sigma_1^2+\sigma_2^2)t^2/2]$$

So X+Y has mean $(\mu_1+\mu_2)$, variance $(\sigma_1^2+\sigma_2^2)$ (duh) *and is normal!*

(*way* easier than slide 2 way!)

Consider i.i.d. (independent, identically distributed) R.V.s

$X_1, X_2, X_3, \ldots$

Suppose X_i has $\mu = E[X_i] < \infty$ and $\sigma^2 = Var[X_i] < \infty$.
What are the mean & variance of their sum?

$$E[\sum_{i=1}^n X_i] = n\mu \text{ and } Var[\sum_{i=1}^n X_i] = n\sigma^2$$

So limit as $n \to \infty$ *does not exist* (except in the degenerate case where $\mu = 0$; note that if $\mu = 0$, the *center* of the data stays fixed, but if $\sigma^2 > 0$, then the variance is unbounded, i.e., its *spread* grows with n).

Consider i.i.d. (independent, identically distributed) R.V.s

$X_1, X_2, X_3, \ldots$

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

Note on notation: in general $X_n \neq \overline{X}_n$

What about the *sample mean*, as $n \to \infty$: $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

$$\mathrm{E}\left[\,\overline{X}_n\right] = \mathrm{E}\left[\frac{X_1 + \cdots + X_n}{n}\right] = \mu$$

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

So, limits *do* exist; mean is independent of *n*, variance shrinks.

Continuing: iid RVs $X_1, X_2, X_3, \ldots;$ μ = E[X_i]; σ^2 = Var[X_i]; $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

$$\mathrm{E}\left[\overline{X}_n\right] = \mathrm{E}\left[\frac{X_1 + \cdots + X_n}{n}\right] = \mu \quad \mathrm{Var}\left[\overline{X}_n\right] = \mathrm{Var}\left[\frac{X_1 + \cdots + X_n}{n}\right] = \frac{\sigma^2}{n}$$

Expectation is an important guarantee.

BUT: observed values may be far from expected values.

E.g., if X_i ~ Bernouli(½), the E[X_i]= ½, but X_i is *NEVER* ½.

Is it also possible that sample mean of X_i's will be far from ½?

Always? Usually? Sometimes? Never?

b&t 5.2

For any ε > 0, as n → ∞

$$\Pr\left(\left|\overline{X}_n - \mu\right| > \epsilon\right) \to 0$$

Proof: (*assume* σ² < ∞)

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

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

By Chebyshev inequality,

$$\Pr\left(\left|\overline{X}_n - \mu\right| > \epsilon\right) \le \frac{\sigma^2}{n\epsilon^2} \xrightarrow{n \to \infty} 0$$

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

$X_1, X_2, X_3, \ldots$

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

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 a finite number of n satisfying the weak law condition $|\overline{X}_n - \mu| \geq \epsilon$ (almost surely, i.e., with probability 1)

Supports the intuition of probability as long term frequency

Weak Law:

$$\lim_{n\to\infty} \Pr\left(\left|\frac{X_1+\cdots+X_n}{n} - \mu\right| > \epsilon\right) = 0$$

Strong Law:

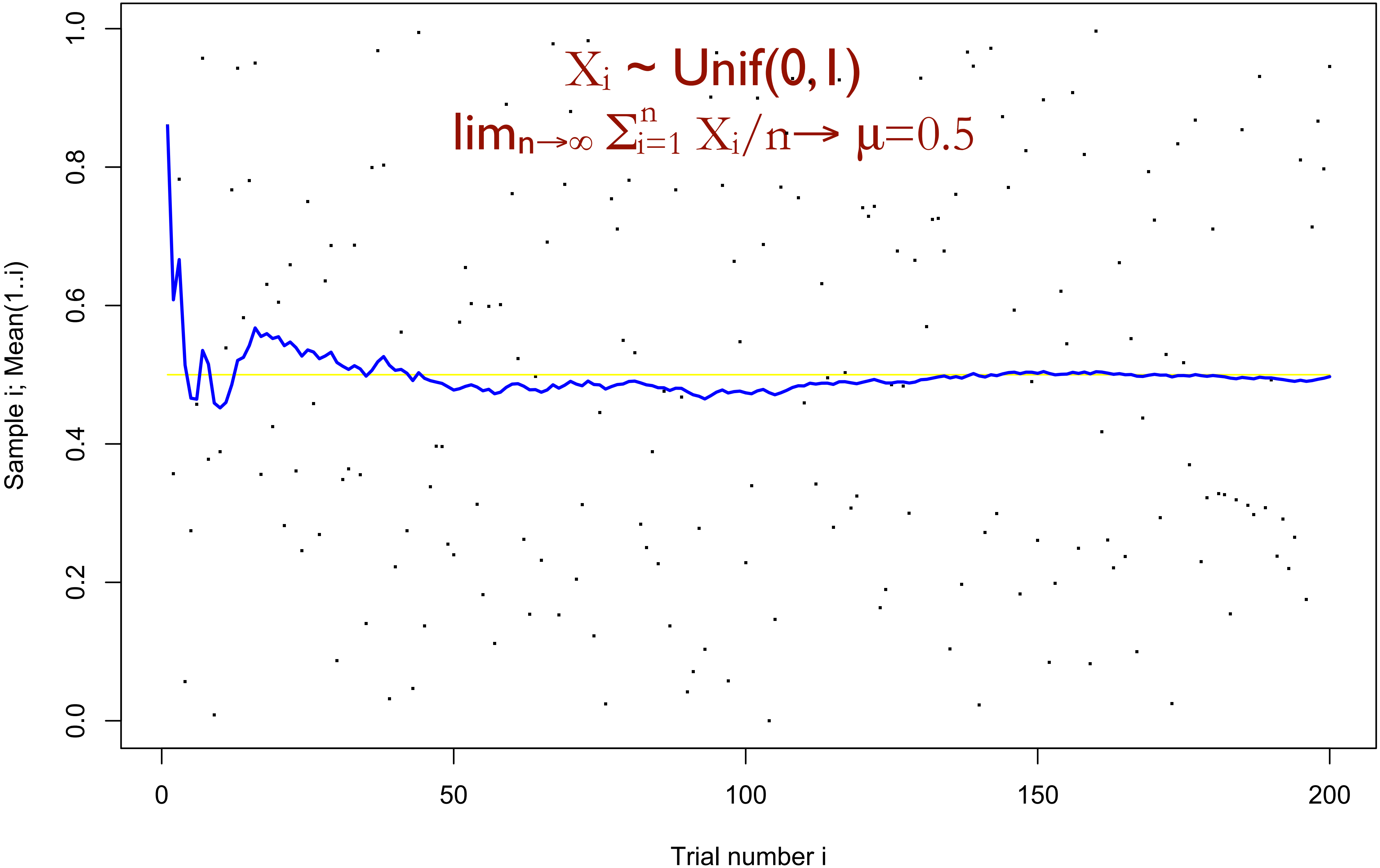
$$\Pr\left(\lim_{n\to\infty}\left(\frac{X_1+\cdots+X_n}{n}\right) = \mu\right) = 1$$

How do they differ? Imagine an infinite 2-D table, whose rows are indp infinite sample sequences X_i. Pick ε. Imagine cell m,n lights up if average of 1st n samples in row m is > ε away from μ.

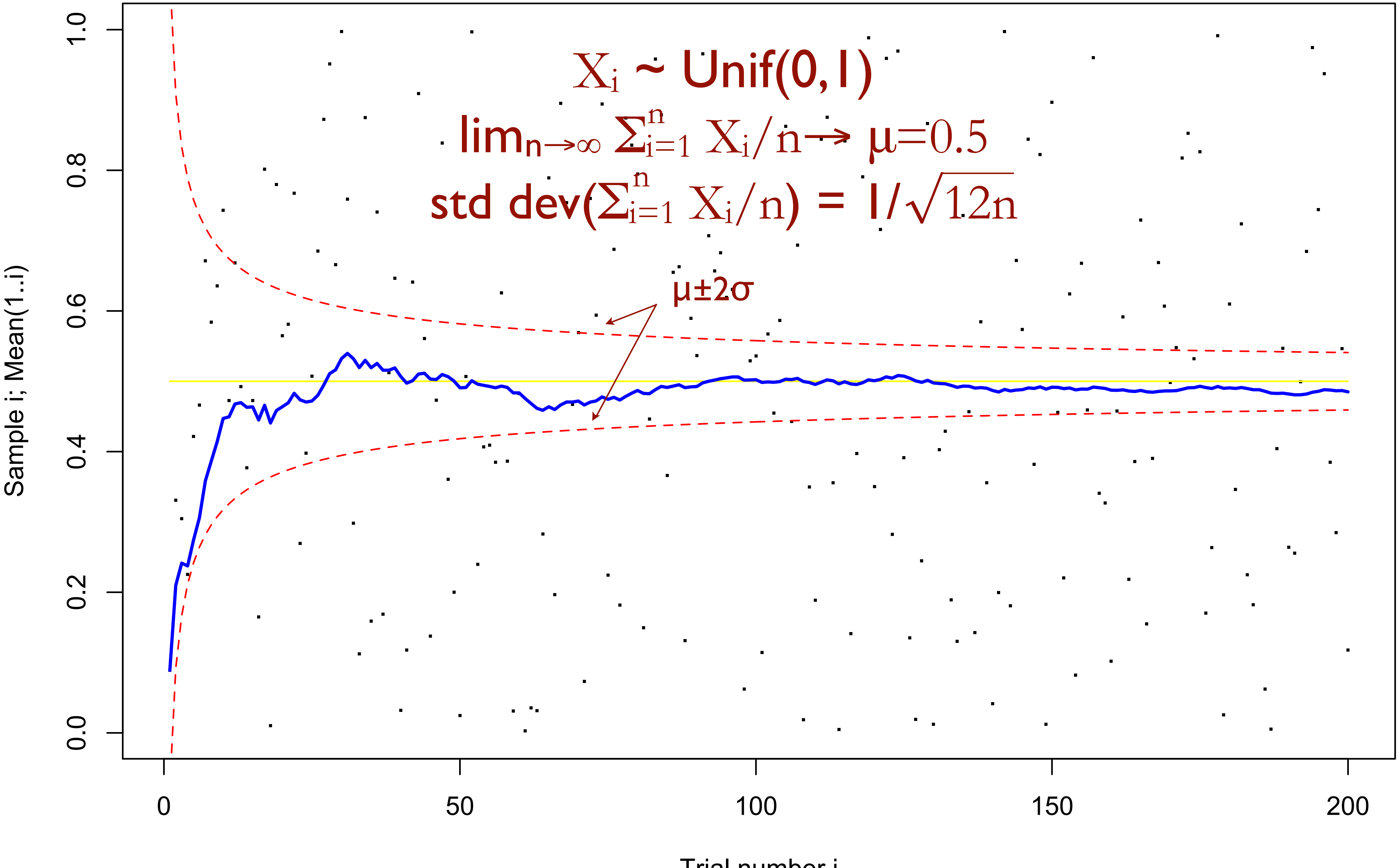
WLLN says fraction of lights in n^{th} column goes to zero as n →∞. It does not prohibit every row from having ∞ lights, so long as frequency declines.

SLLN also says only a vanishingly small fraction of rows can have ∞ lights.

sample mean → population mean

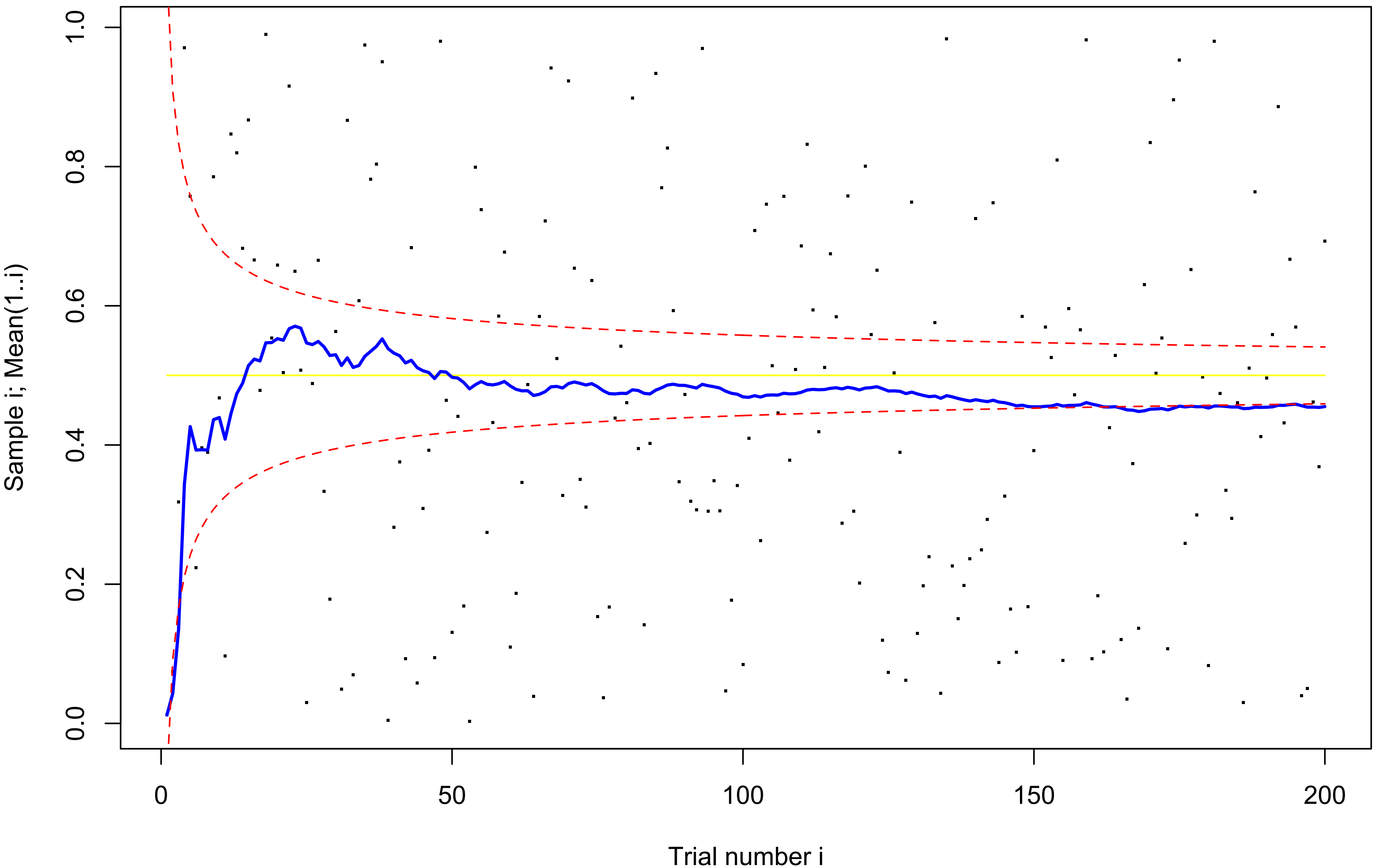


sample mean → population mean

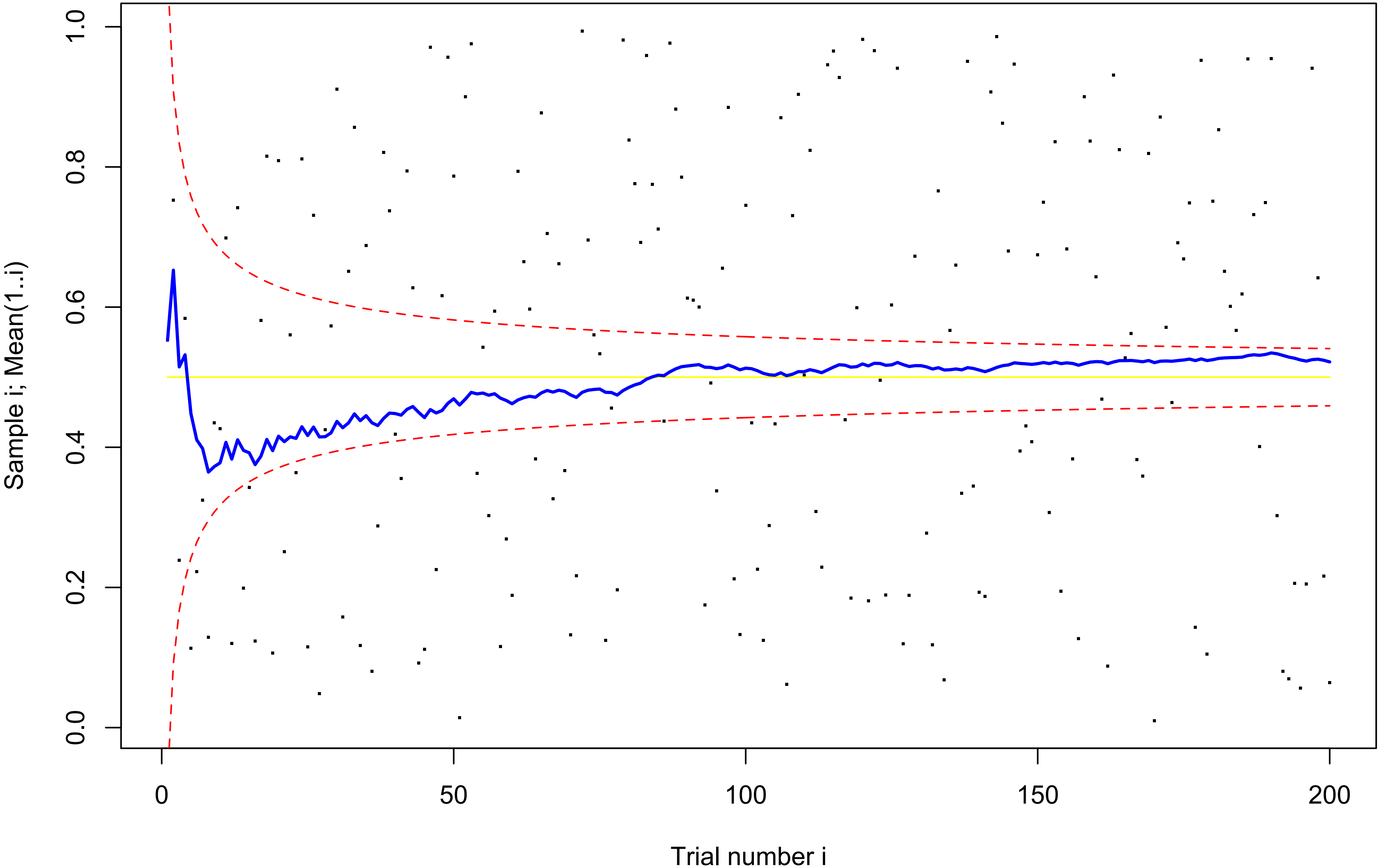


demo

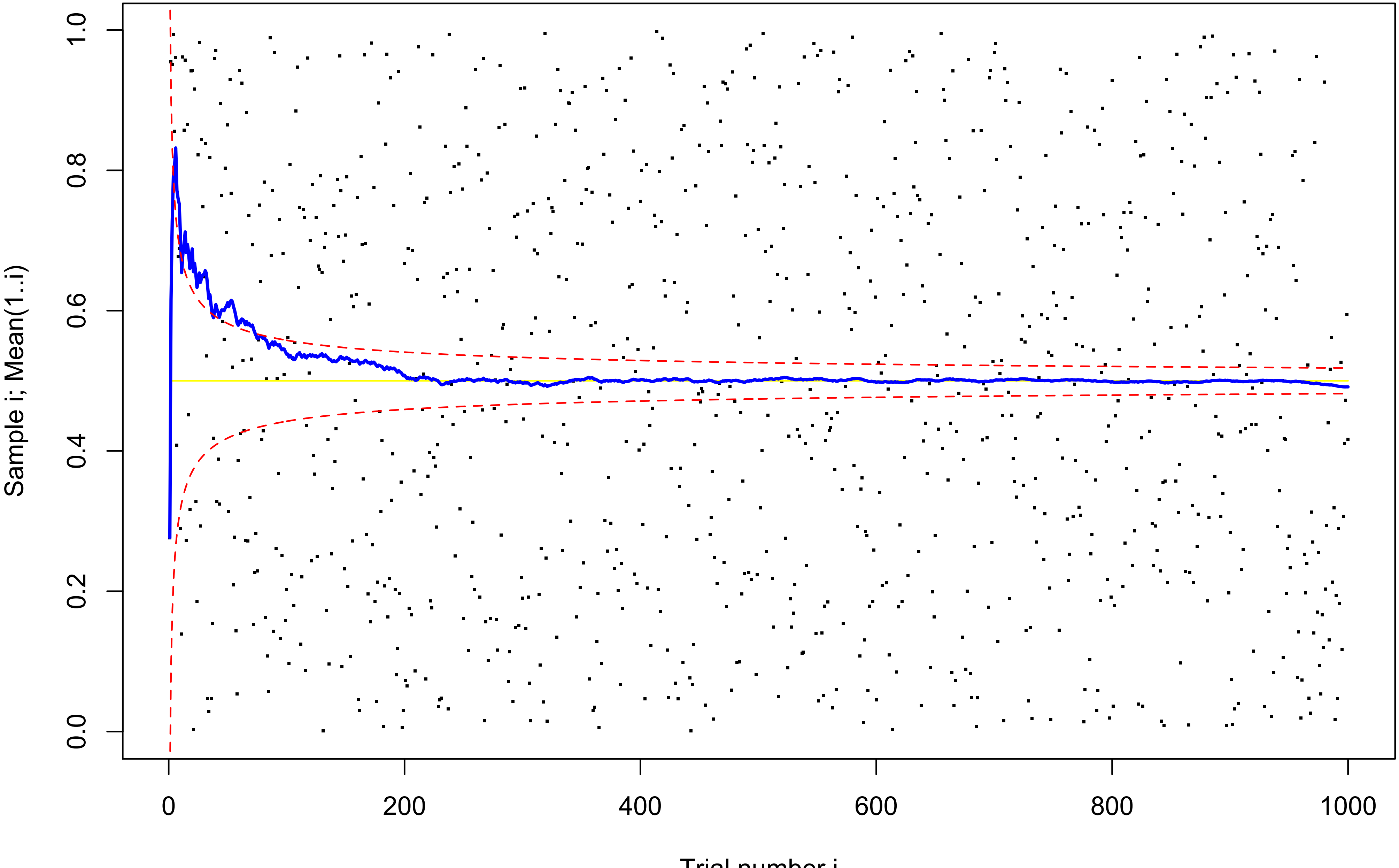
another example



another example



another example



Weak Law:

$$\lim_{n\to\infty} \Pr\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| > \epsilon\right) = 0$$

Strong Law:

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

How do they differ? Imagine an infinite 2-D table, whose rows are indp infinite sample sequences X_i. Pick ε. Imagine cell m,n lights up if average of 1st n samples in row m is > ε away from μ.

WLLN says fraction of lights in nth column goes to zero as n →∞. It does not prohibit every row from having ∞ lights, so long as frequency declines.

SLLN also says only a vanishingly small fraction of rows can have ∞ lights.

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

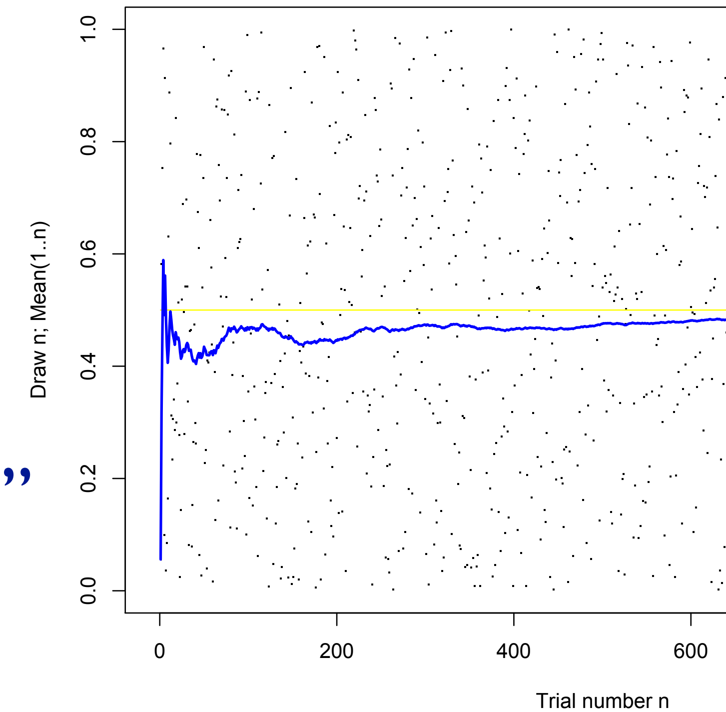
Justifies the "frequency" interpretation of probability

"Regression toward the mean"

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

"Swamps, but does not compensate"

"Result will usually be close to the mean"



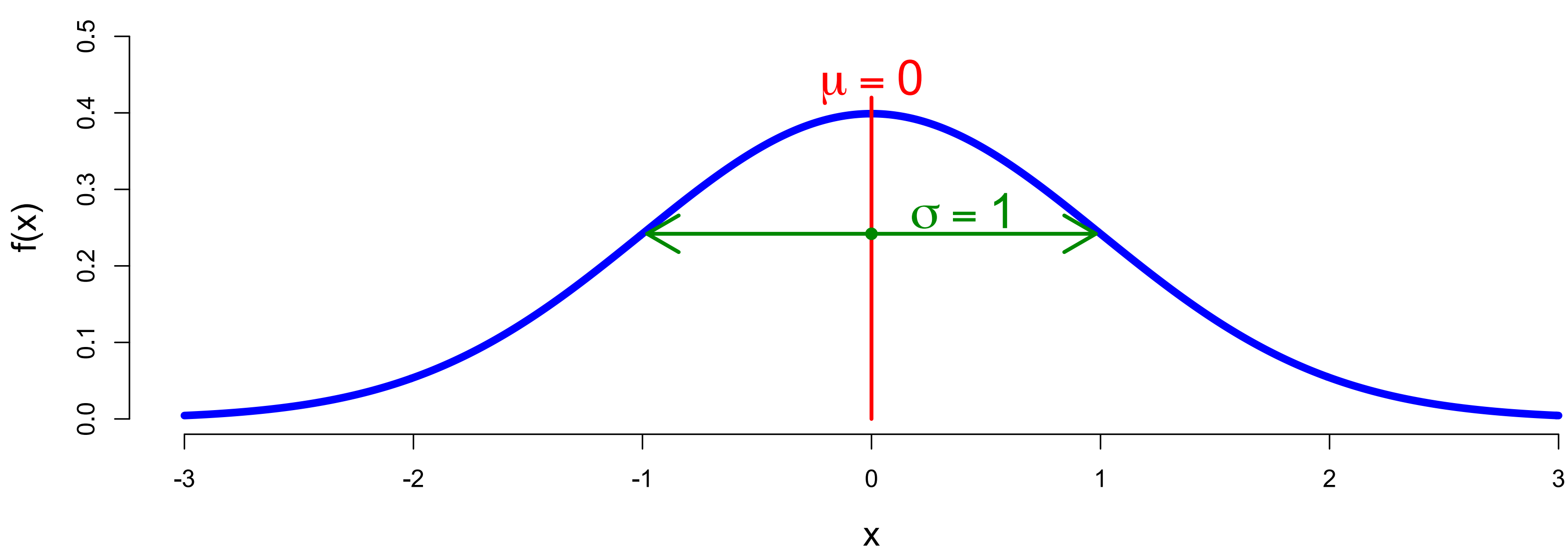
Many web demos, e.g.
http://stat-www.berkeley.edu/~stark/Java/Html/lln.htm

Recall

X is a normal random variable X ~ N(μ,σ²)

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

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



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

$X_1, X_2, X_3, \ldots$

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

As $n \to \infty$,

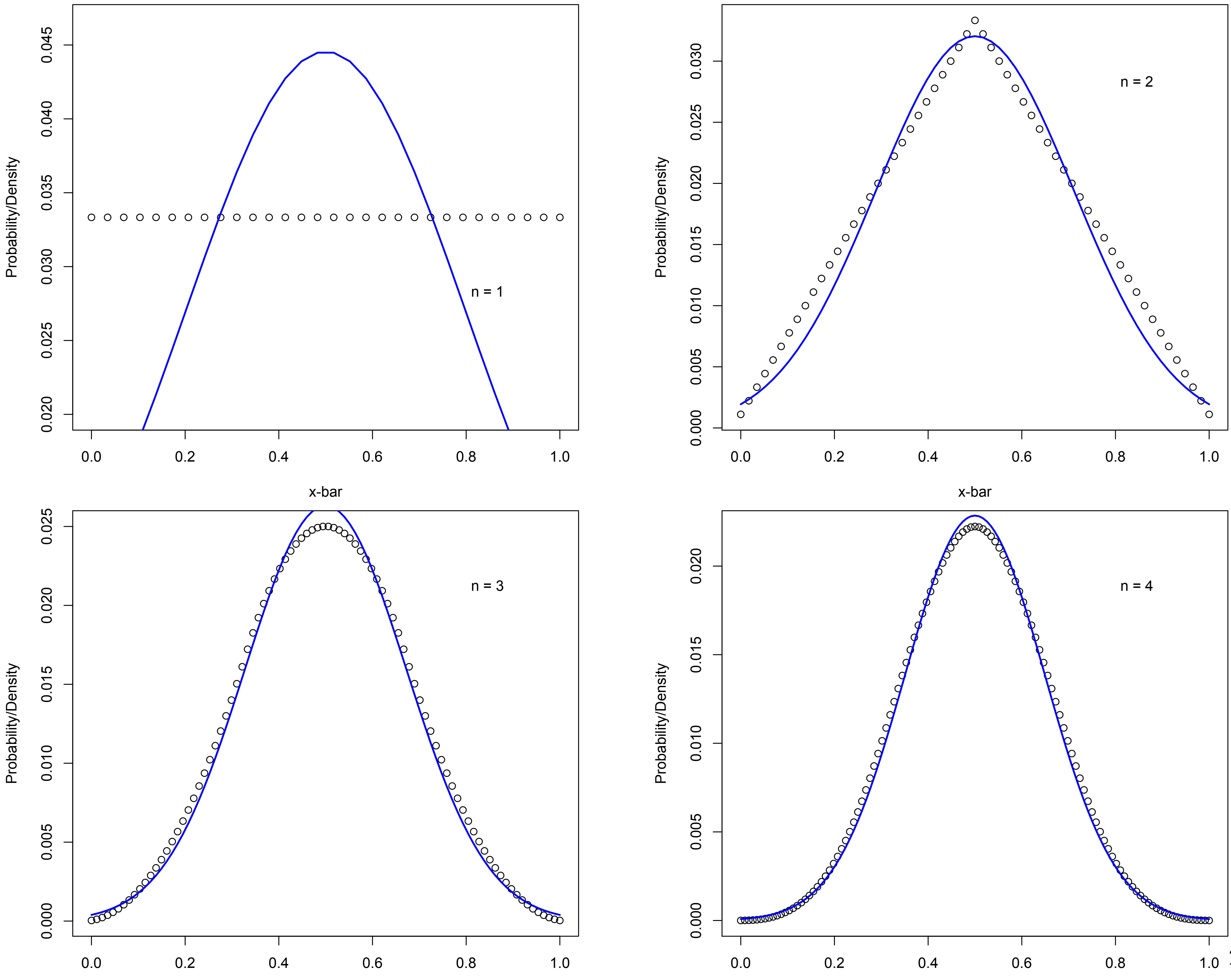
$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Restated: As $n \to \infty$,

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

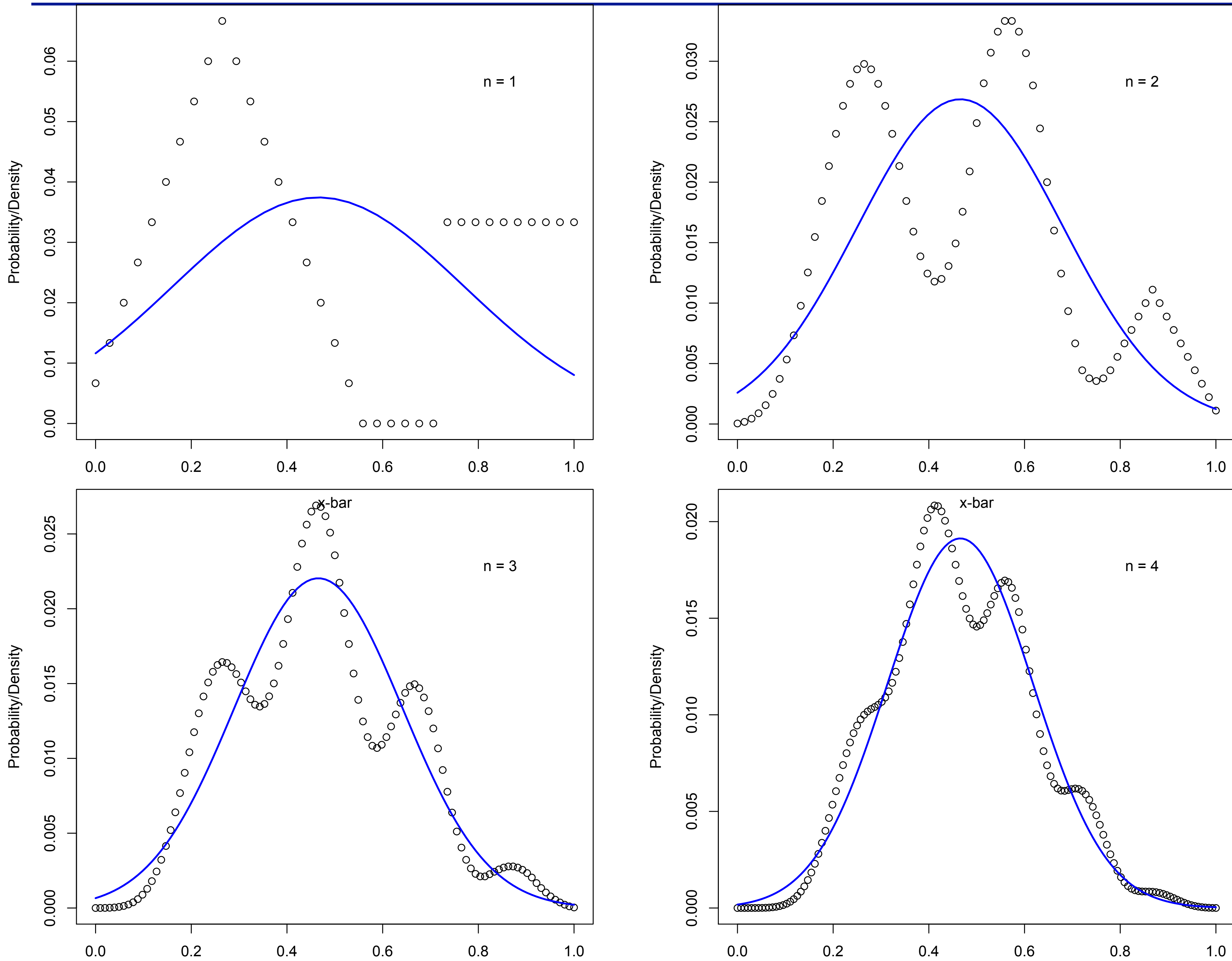
Note: on slide 5, showed sum of normals is exactly normal. Maybe not a surprise, given that sums of almost *anything* become approximately normal...

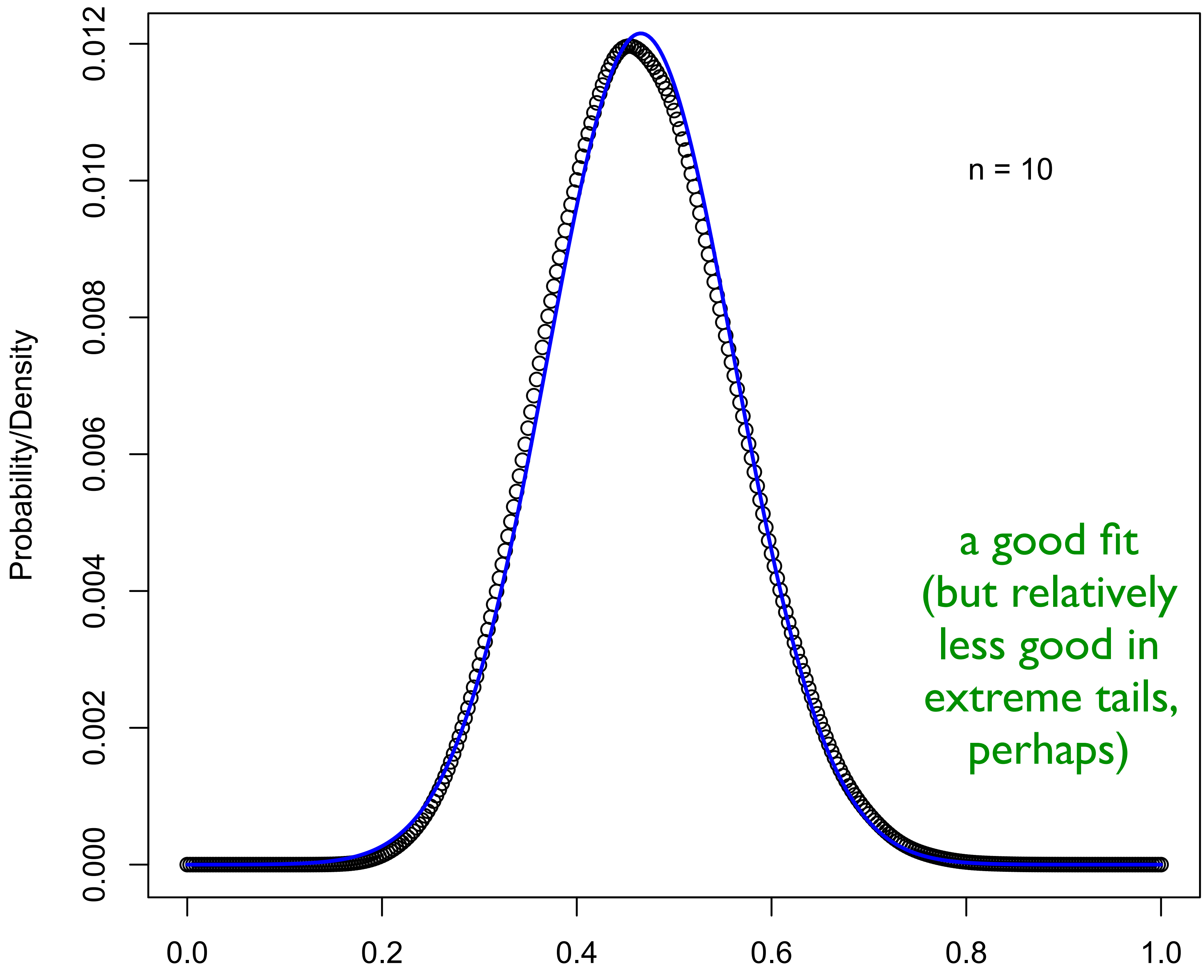
demo



n = 1
n = 2
n = 3
n = 4
Probability/Density
x-bar
0.0
0.2
0.4
0.6
0.8
1.0

CLT applies even to whacky distributions





Probability/Density
0.000
0.002
0.004
0.006
0.008
0.010
0.012
0.0
0.2
0.4
0.6
0.8
1.0
n = 10
a good fit
(but relatively
less good in
extreme tails,
perhaps)

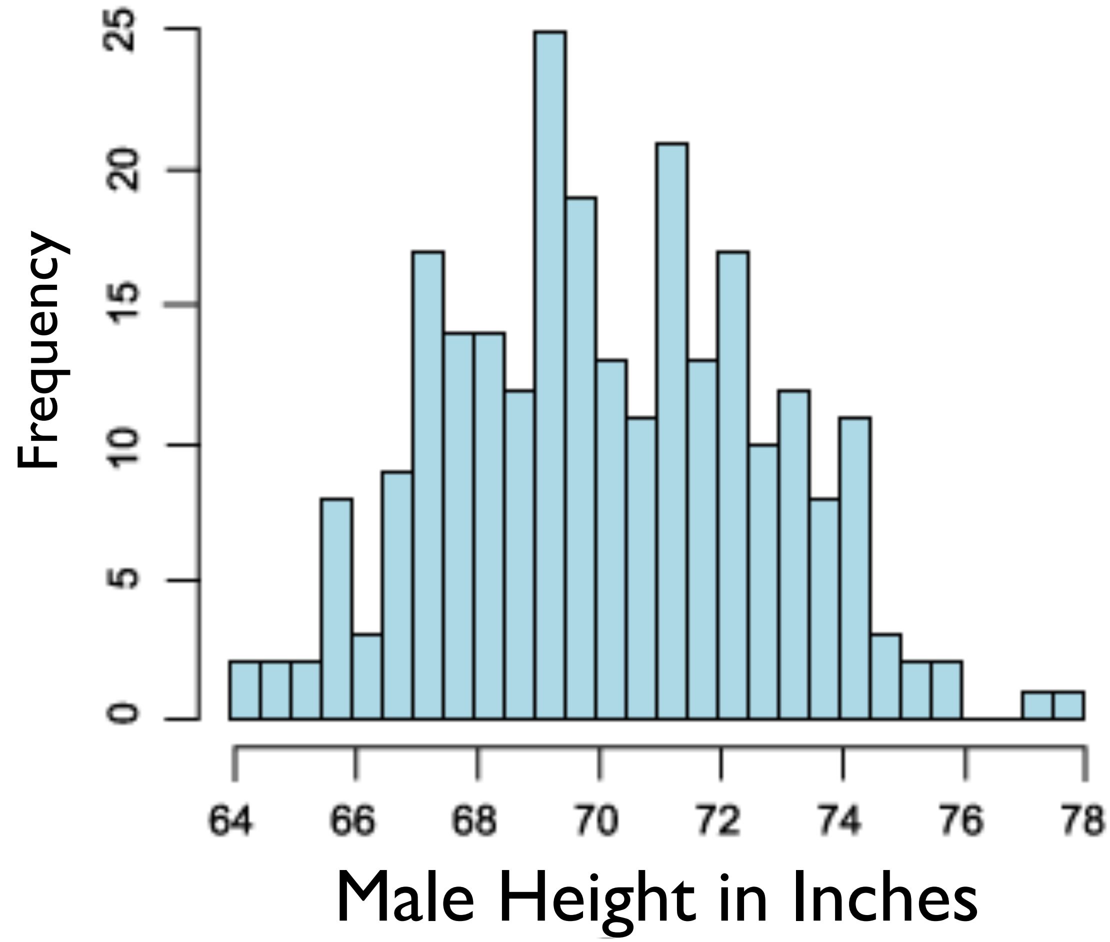
CLT in the real world

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
...

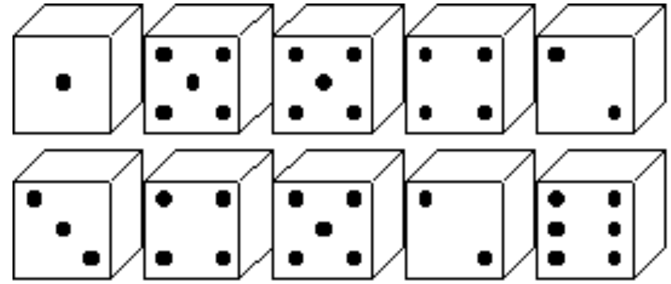
Human height is approximately normal.

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.

Roll 10 6-sided dice



X = total value of all 10 dice

Win if: X ≤ 25 or X ≥ 45

$$\mathsf{E}[X] = \mathsf{E}[\textstyle\sum_{i=1}^{10} X_i] = 10\mathsf{E}[X_1] = 10(7/2) = 35$$

$$\mathsf{Var}[X] = \mathsf{Var}[\textstyle\sum_{i=1}^{10} X_i] = 10\mathsf{Var}[X_1] = 10(35/12) = 350/12$$

$$P(\mathsf{win}) = 1 - P\left(25.5 \le X \le 44.5\right) =$$

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

$$\approx 2(1 - \Phi(1.76)) \approx 0.079$$

Poll of 100 randomly chosen voters finds that K of them favor proposition 666.

So: the *estimated proportion* in favor is $K/100 = q$

Suppose: the *true proportion* in favor is p.

Q. Give an upper bound on the probability that your estimate is off by > 10 percentage points, i.e., the probability of $|q - p| > 0.1$

A. $K = X_1 + \ldots + X_{100}$, where X_i are Bernoulli(p), so by CLT:

$K \approx$ normal with mean $100p$ and variance $100p(1-p)$; or:

$q \approx$ normal with mean p and variance $\sigma^2 = p(1-p)/100$

Letting $Z = (q-p)/\sigma$ (a standardized r.v.), then $|q - p| > 0.1 \Leftrightarrow |Z| > 0.1/\sigma$

By symmetry of the normal

$$P_{\text{Ber}}(|q - p| > 0.1) \approx 2\, P_{\text{norm}}(Z > 0.1/\sigma) = 2\,(1 - \Phi(0.1/\sigma))$$

Unfortunately, p & σ are unknown, but $\sigma^2 = p(1-p)/100$ is maximized when $p = 1/2$, so $\sigma^2 \leq 1/400$, i.e. $\sigma \leq 1/20$, hence

$$2\,(1 - \Phi(0.1/\sigma)) \leq 2(1-\Phi(2)) \approx \boxed{0.046}$$

Exercise: How much smaller can σ be if $p \neq 1/2$?

I.e., less than a 5% chance of an error as large as 10 percentage points.

Distribution of X + Y: summations, integrals (or MGF)

Distribution of X + Y ≠ distribution X or Y in general

Distribution of X + Y is normal if X and Y are normal (*)

(ditto for a few other special distributions)

Sums generally don't "converge," but averages do:

Weak Law of Large Numbers

Strong Law of Large Numbers

Most surprisingly, averages all converge to the *same* distribution:

the Central Limit Theorem says sample mean → normal

[Note that (*) essentially a prerequisite, and that (*) is exact, whereas CLT is approximate]