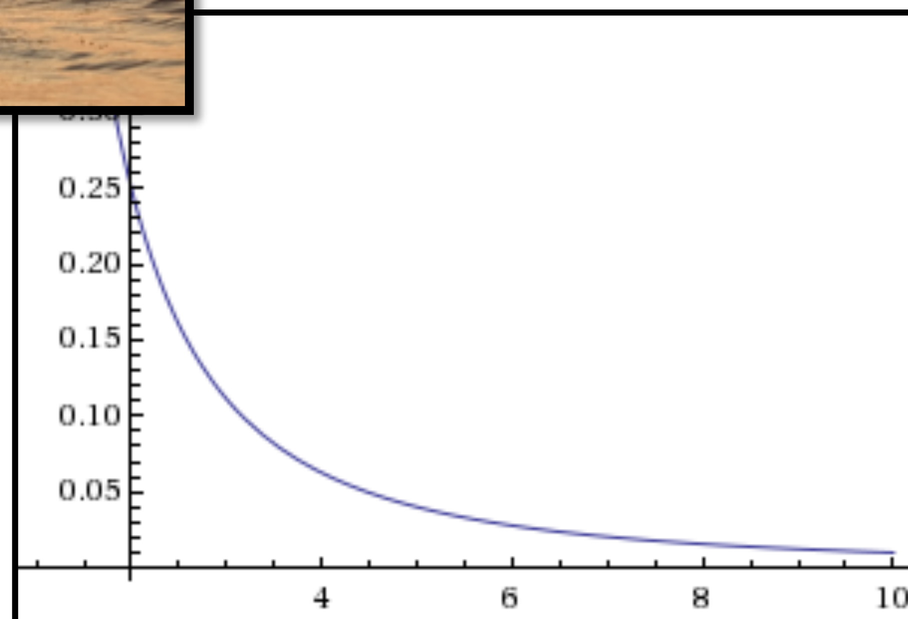
tail bounds



Markov's Inequality:

If X is a non-negative random variable, then for every c>0, we have

$$P(X \geq c) \leq \frac{E(X)}{c}$$

Chevyshev's Inequality: If X is an arbitrary random variable with $\mu = E(X)$ then for any c > 0

$$P(|X - \mu| \geq c) \leq \frac{Var(X)}{c^2}$$

Suppose X ~ Bin(n,p), μ = E[X] = pn

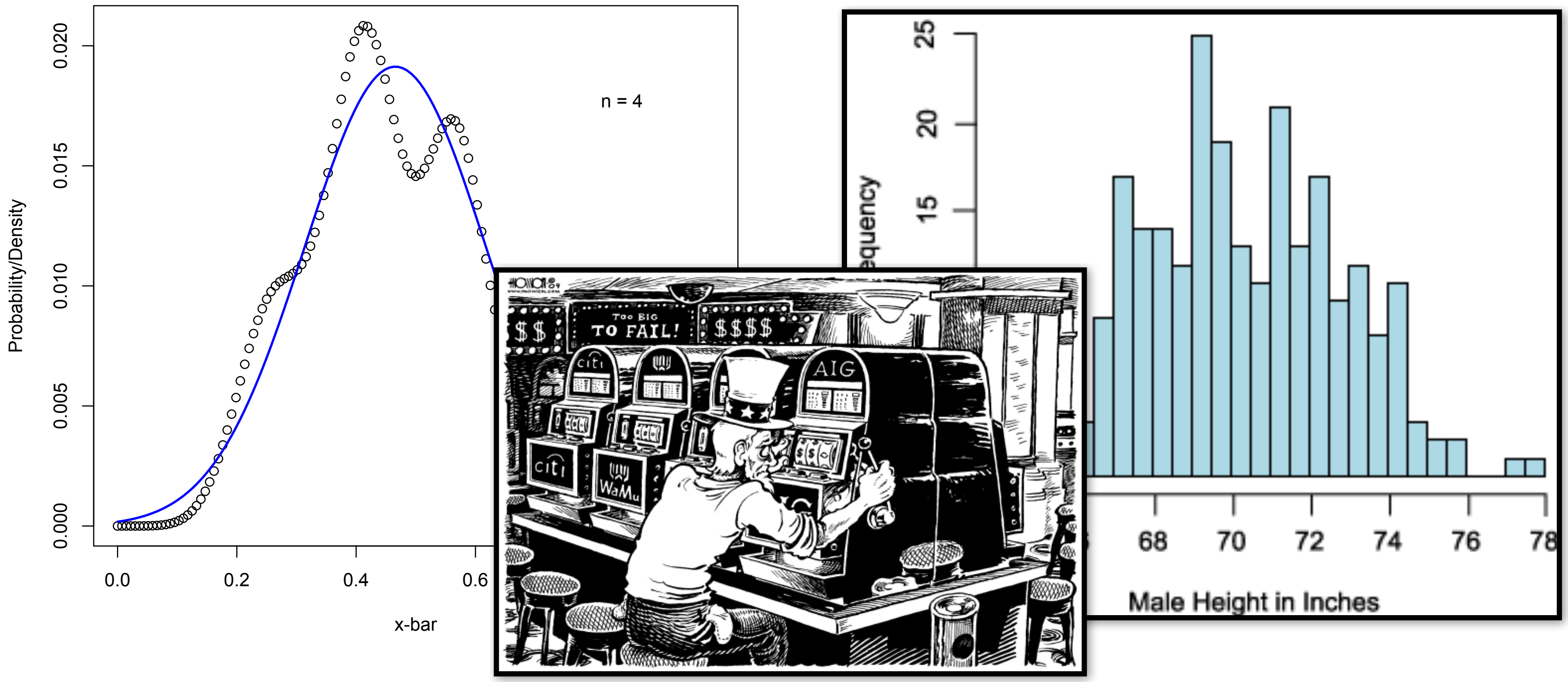
Chernoff bound:

For any δ with $0 < \delta < 1$,

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

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

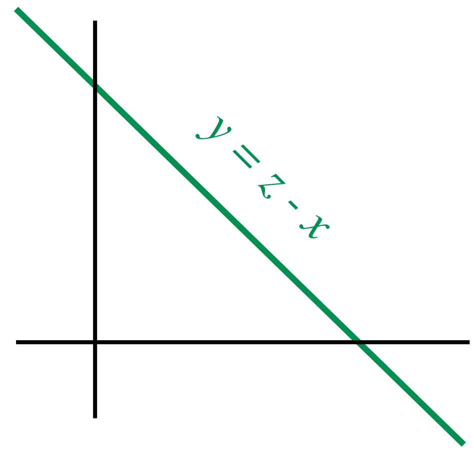
the law of large numbers & the CLT



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

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

Discrete case:



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

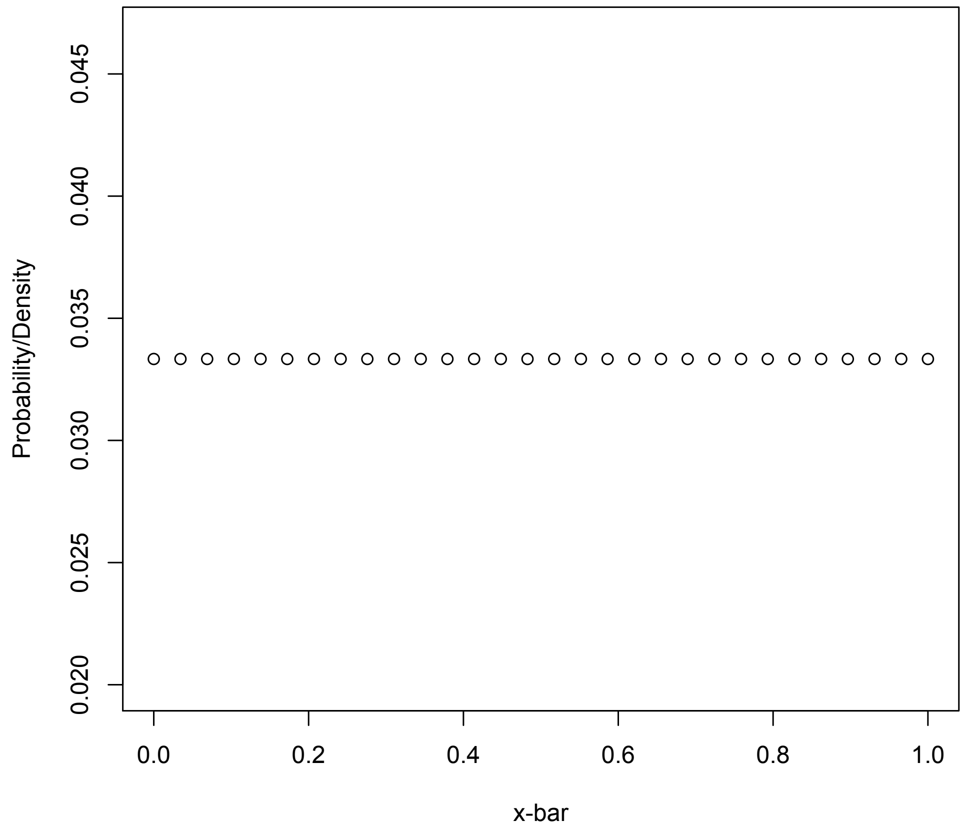
Continuous case:

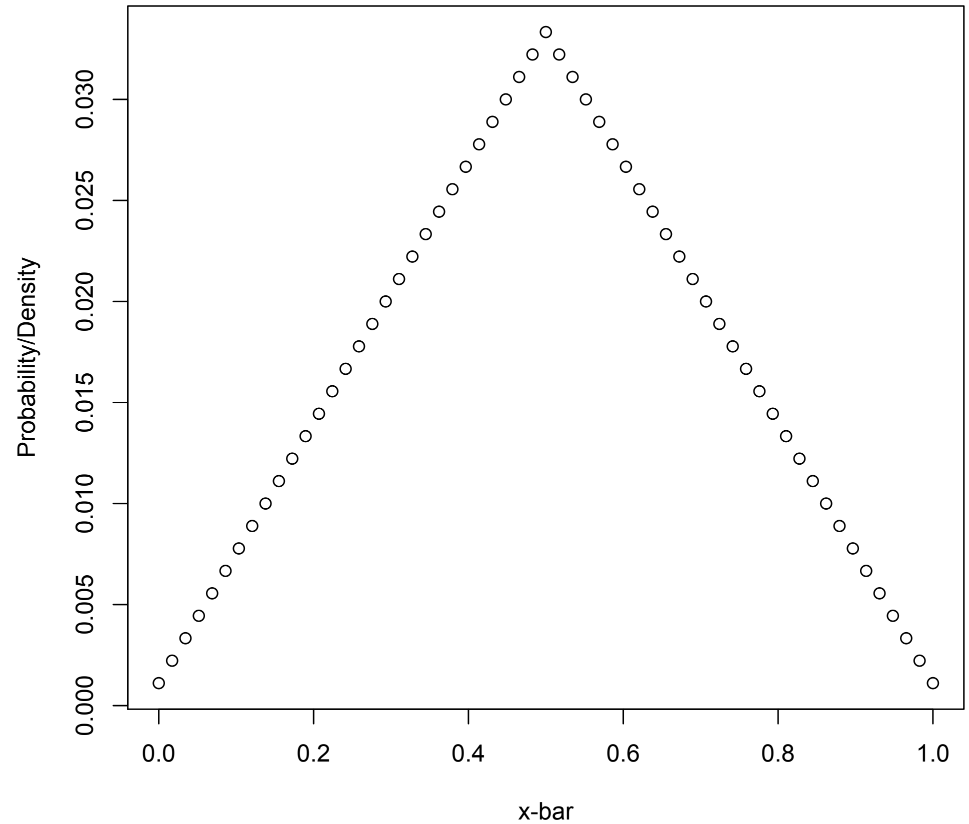
$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x) \bullet f_Y(z-x) \, dx$$

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 *triangular*:





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

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 $M(t) = E[e^{tX}]$

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

$$\frac{d}{dt}M(t)\Big|_{t=0} = E[X] \qquad \frac{d^2}{dt^2}M(t)\Big|_{t=0} = E[X^2] \quad \cdots \quad \frac{d^k}{dt^k}M(t)\Big|_{t=0} = E[X^k] \quad \cdots$$

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 indep normals is normal:*

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

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

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

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

$X_1, X_2, X_3, \ldots$

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

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

So limits as $n \to \infty$ don’t exist (except in the degenerate case where $\mu = \sigma^2 = 0$).

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/sample 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* $\sigma^2 < \infty$)

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

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

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

$X_1, X_2, X_3, \ldots$

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

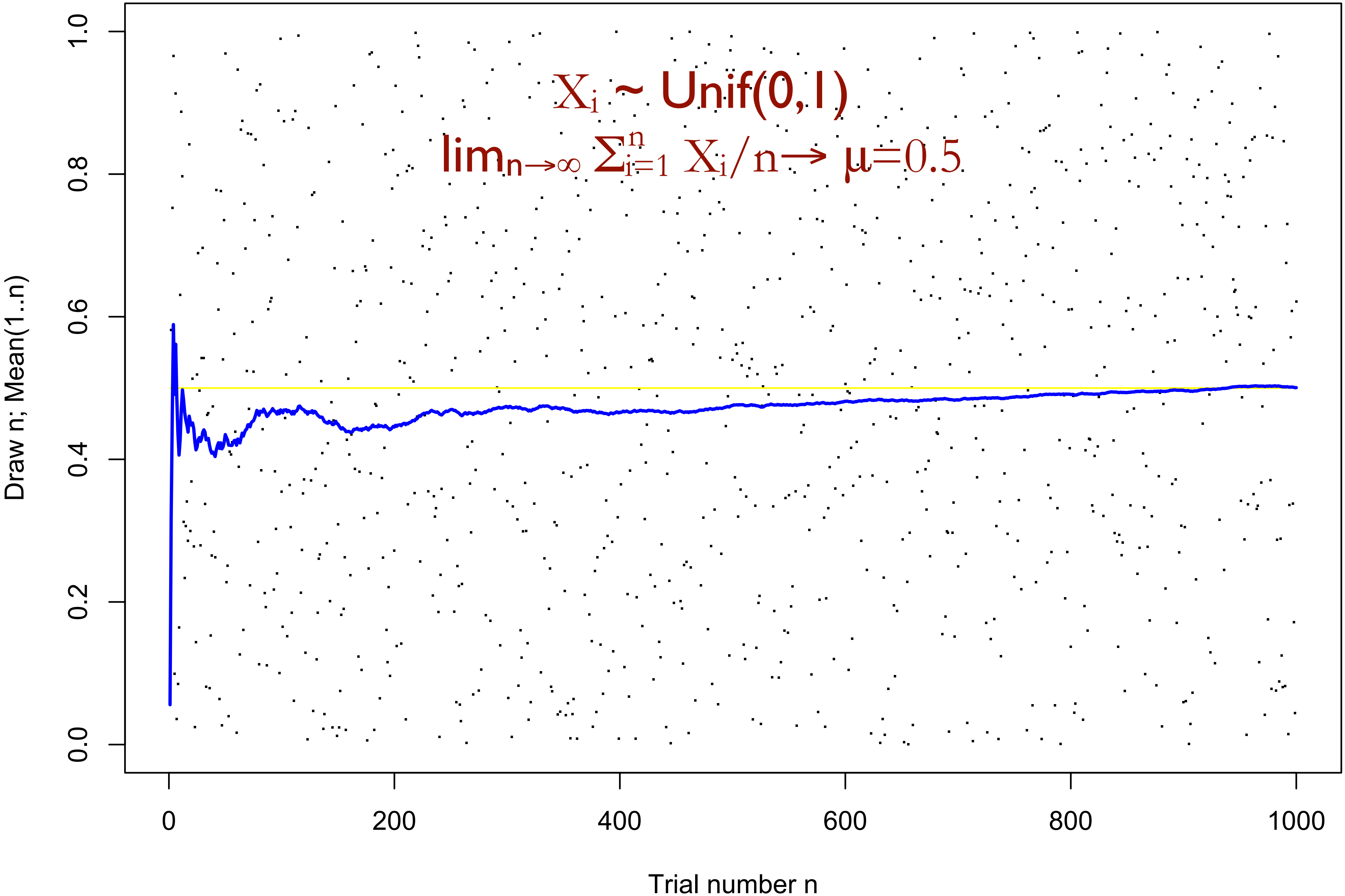
X_i has $\mu = E[X_i] < \infty$

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

Strong Law ⇒ Weak Law (but not vice versa)

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

sample mean → population mean

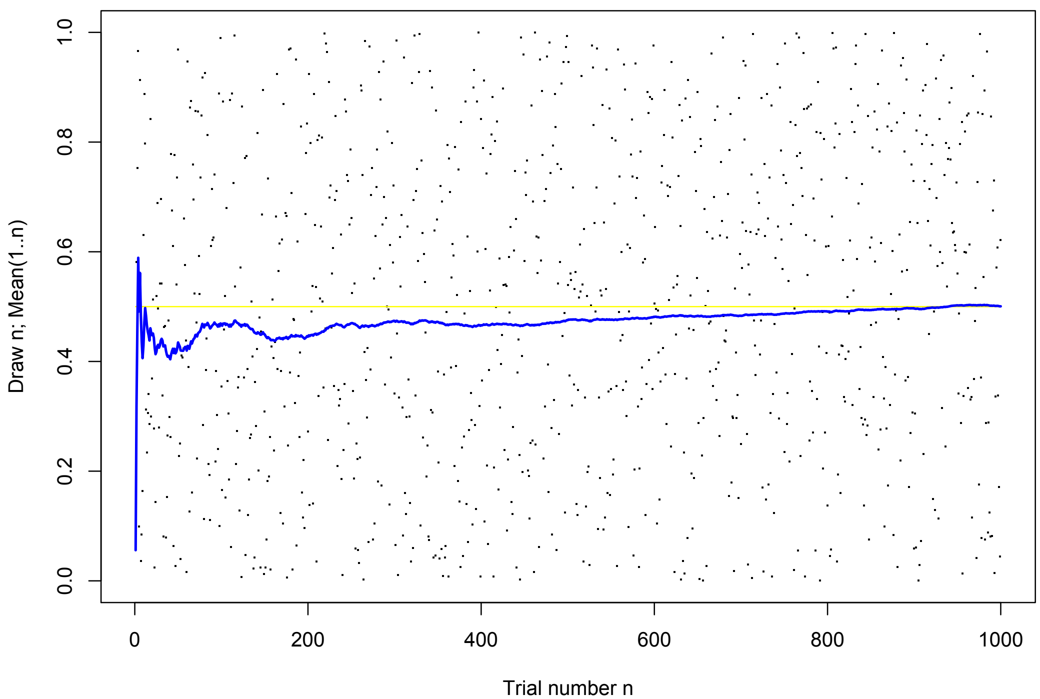


Note: $D_n = E[\,|\,\Sigma_{1 \leq i \leq n}(X_i - \mu)\,|\,]$ *grows* with n, but $D_n/n \to 0$

Justifies the “frequency” interpretation of probability

but not “Regression toward the mean”

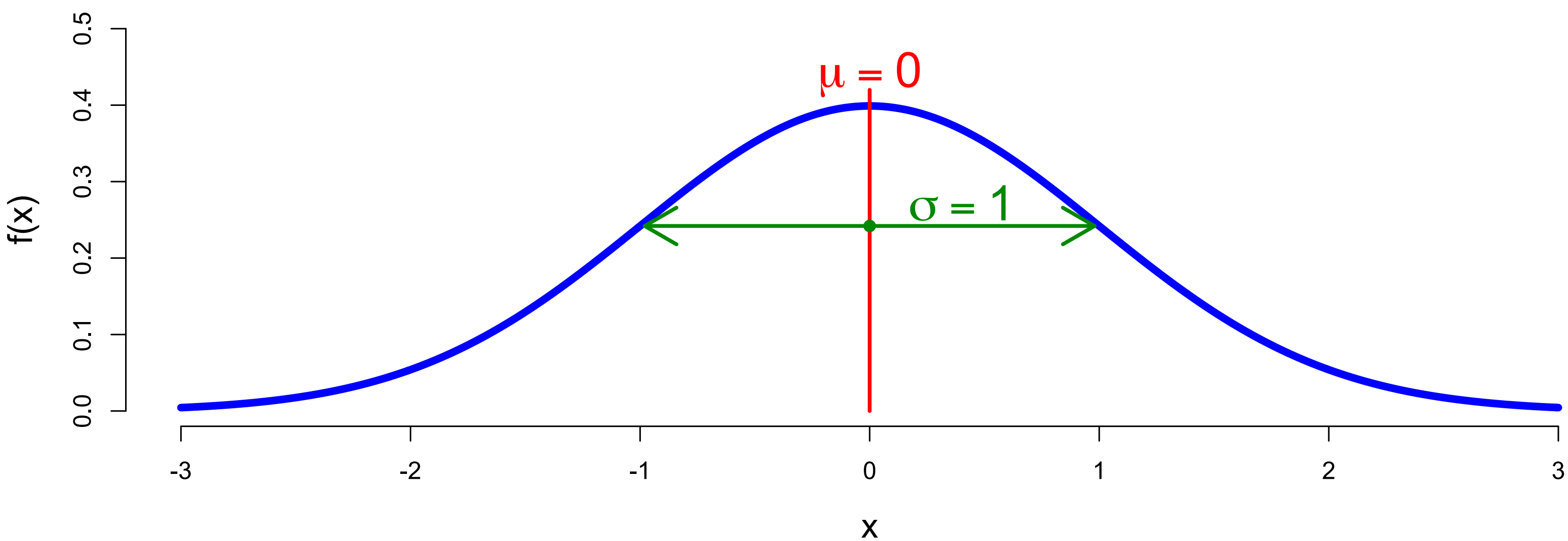
and not gambler’s fallacy: “I’m *due* for a win!”



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

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

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



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

$X_1, X_2, X_3, \ldots$

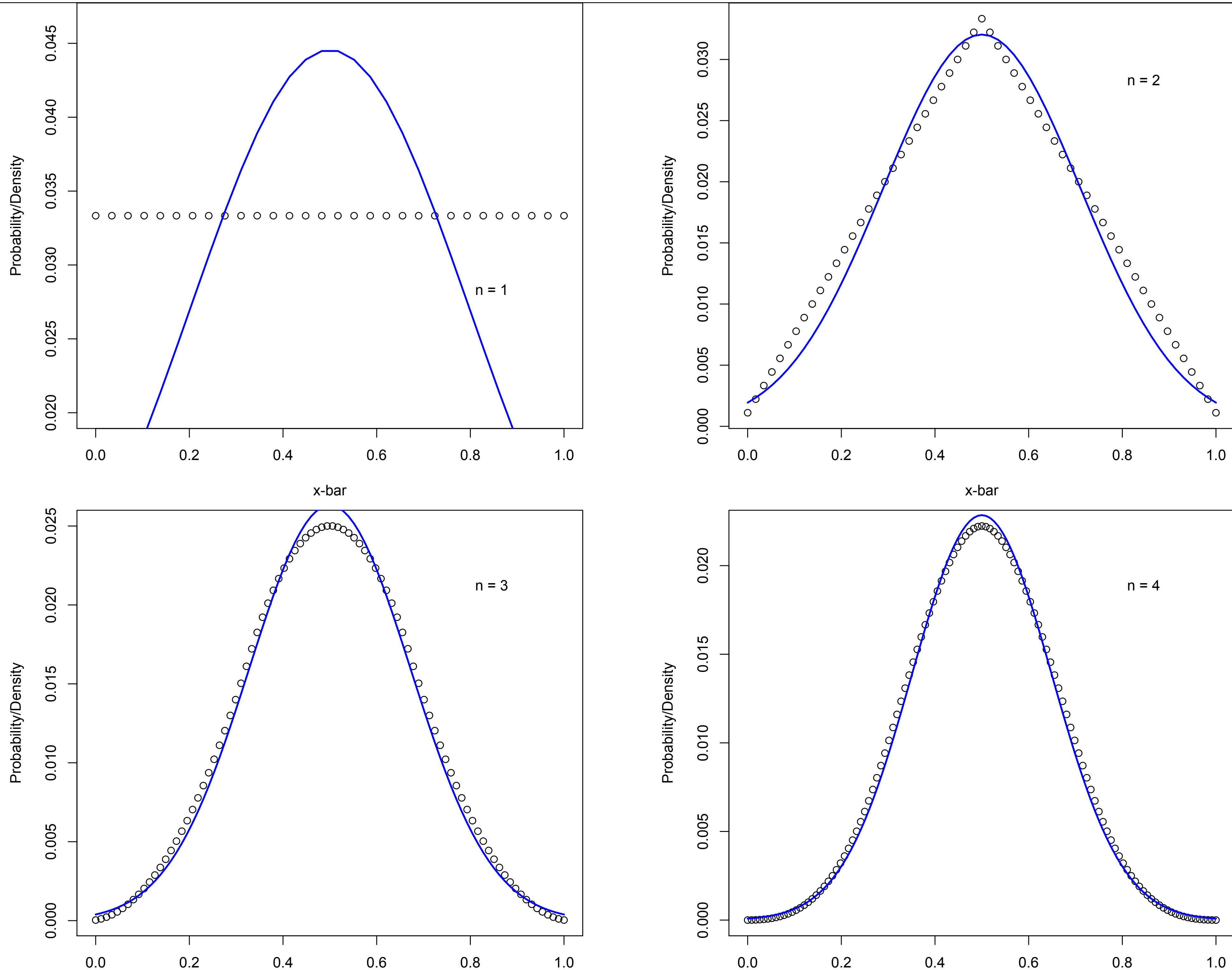
X_i has $\mu = E[X_i]$ and $\sigma^2 = Var[X_i]$

As $n \to \infty$,

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

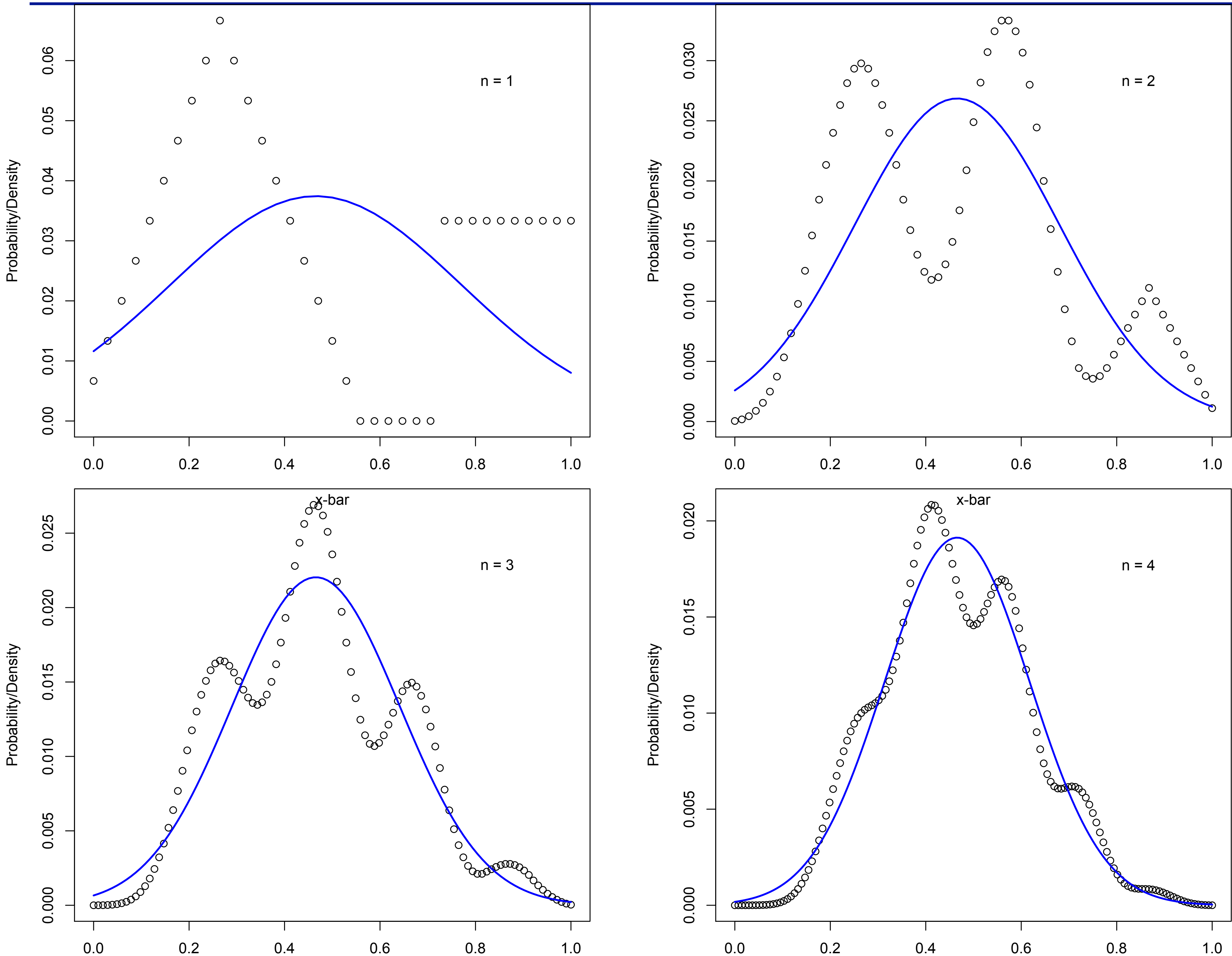
Restated: As $n \to \infty$,

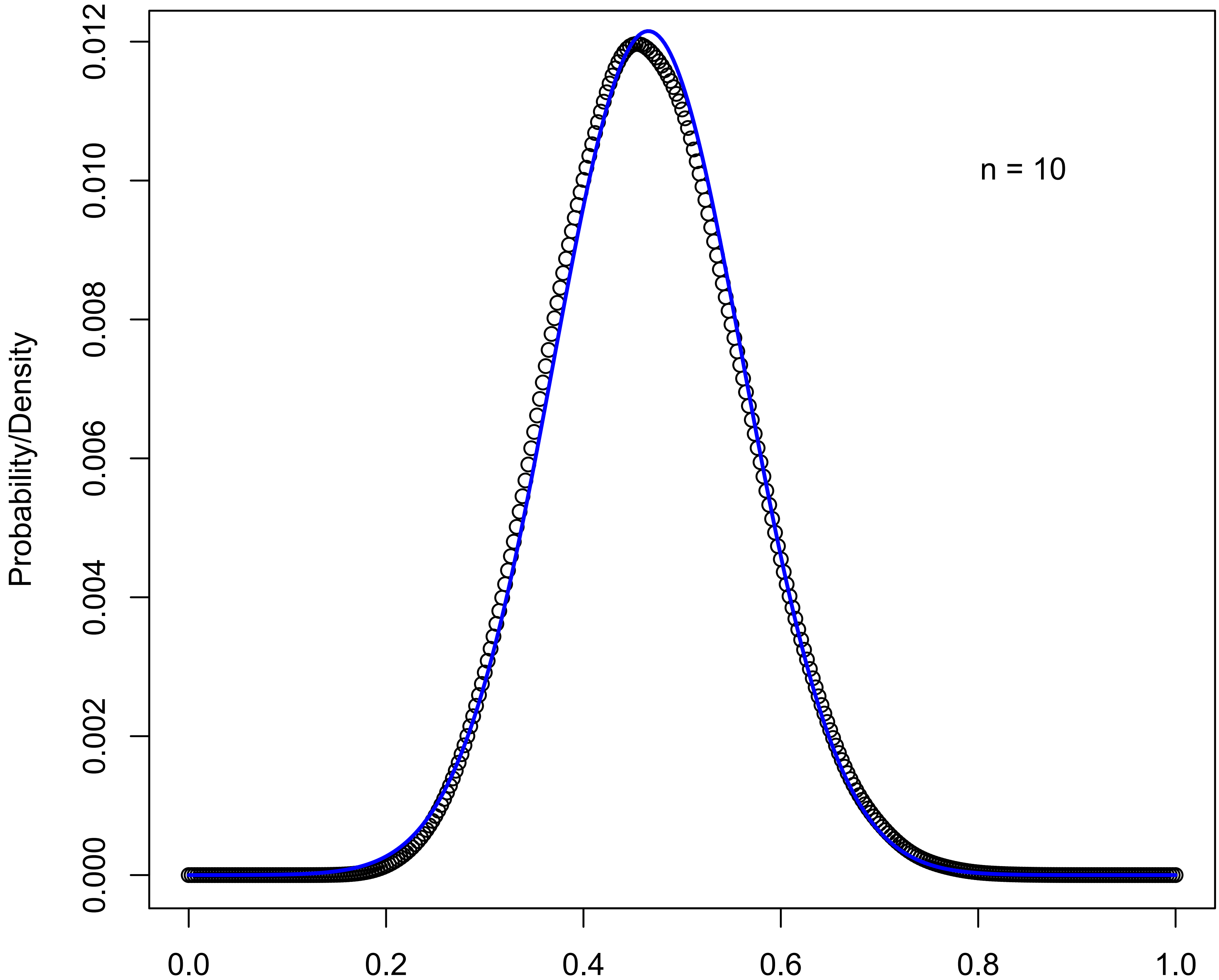
$$\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \longrightarrow N(0,1)$$



n = 1
n = 2
n = 3
n = 4
Probability/Density
x-bar

CLT applies even to even wacky distributions

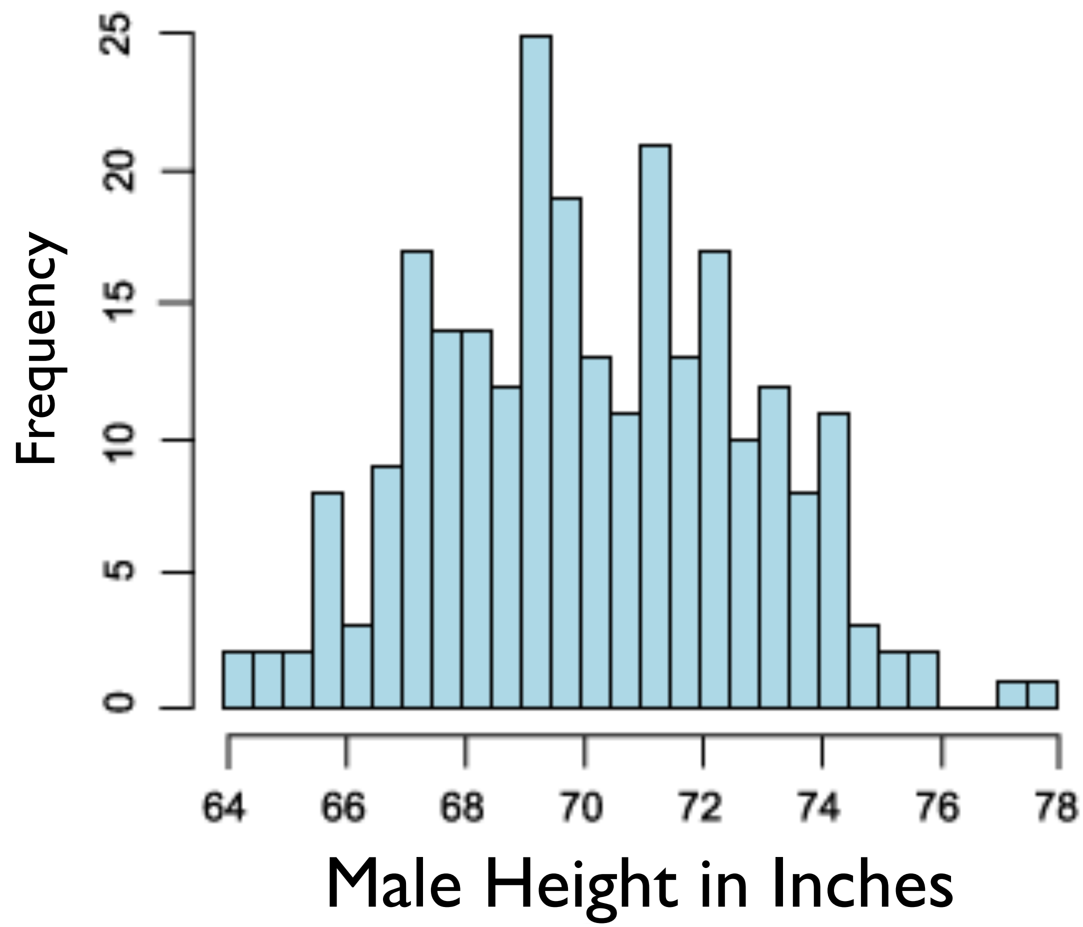



Probability/Density
0.000
0.002
0.004
0.006
0.008
0.010
0.012
0.0
0.2
0.4
0.6
0.8
1.0
n = 10

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.

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