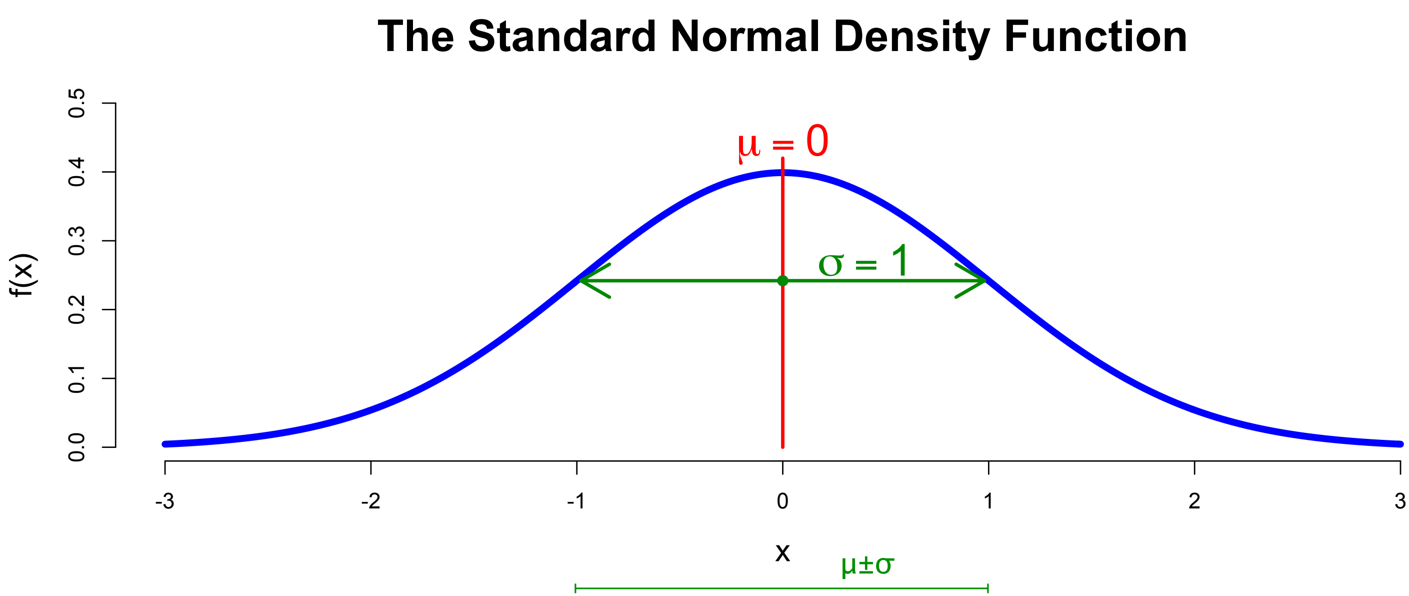
X is a normal (aka Gaussian) random variable X ~ N(μ, σ²)

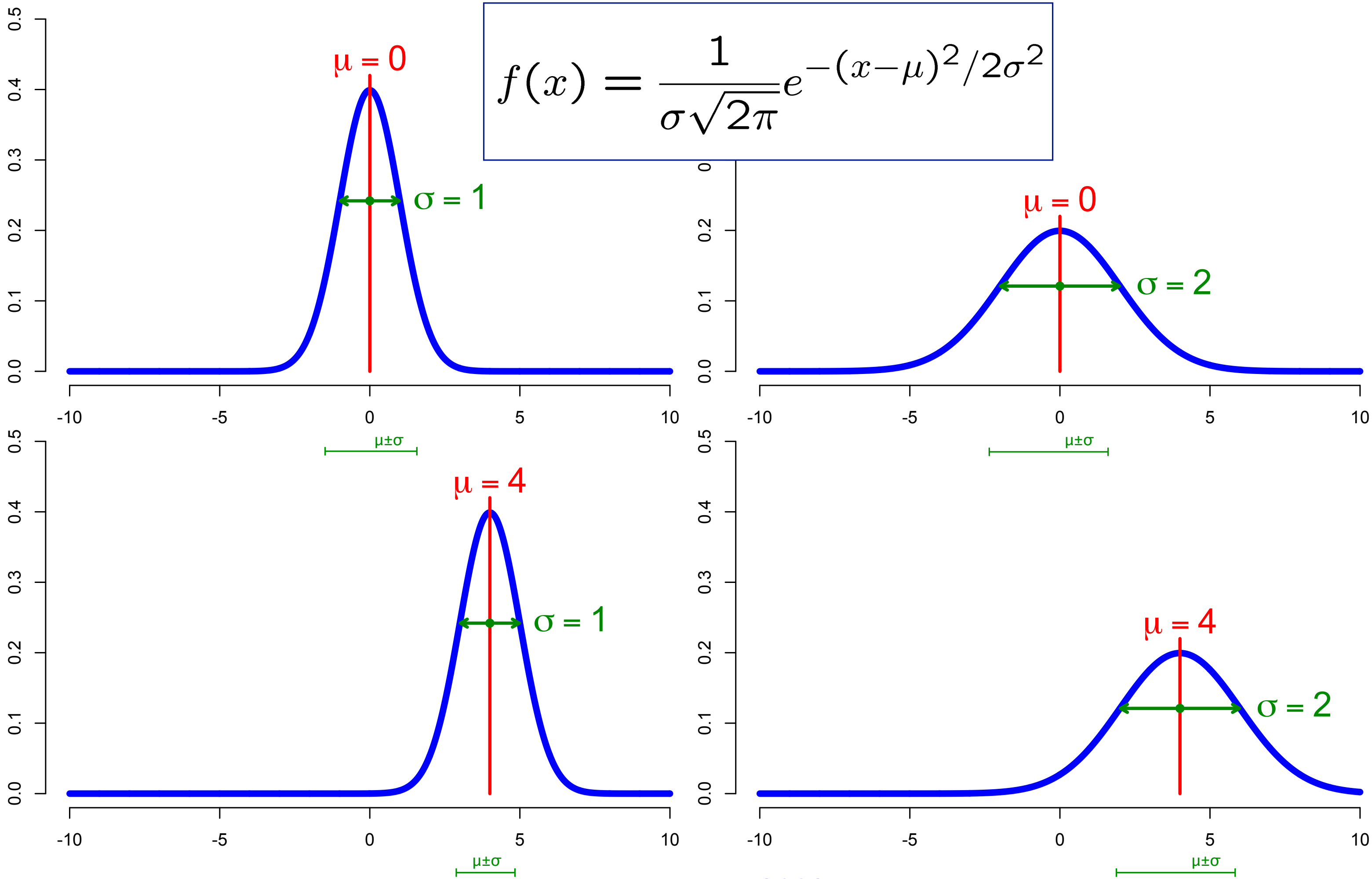
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

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



changing μ, σ

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



density at μ is ≈ .399/σ

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

$Y = aX + b$

$E[Y] = E[aX+b] =$

$Var[Y] = Var[aX+b] =$

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

$X = a_1X_1 + a_2X_2 + \dots + a_nX_n$

a_i's consts

$X_i \sim N(\mu_i, \sigma_i^2)$

X_i's are indep.

$E(X) = a_1\mu_1 + a_2\mu_2 + \dots + a_n\mu_n$

$Var(X) = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \dots + a_n^2\sigma_n^2$

$X \sim N\left(\sum_{i=1}^{n} a_i\mu_i, \sum_{i=1}^{n} a_i^2\sigma_i^2\right)$

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

$Y = aX + b$

$E[Y] = E[aX+b] = a\mu + b$

$Var[Y] = Var[aX+b] = a^2\sigma^2$

$Y \sim N(a\mu + b, a^2\sigma^2)$

E[ ], Var[ ] as expected; "normality" is the surprise

Important special case: $Z = (X-\mu)/\sigma \sim N(0,1)$

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

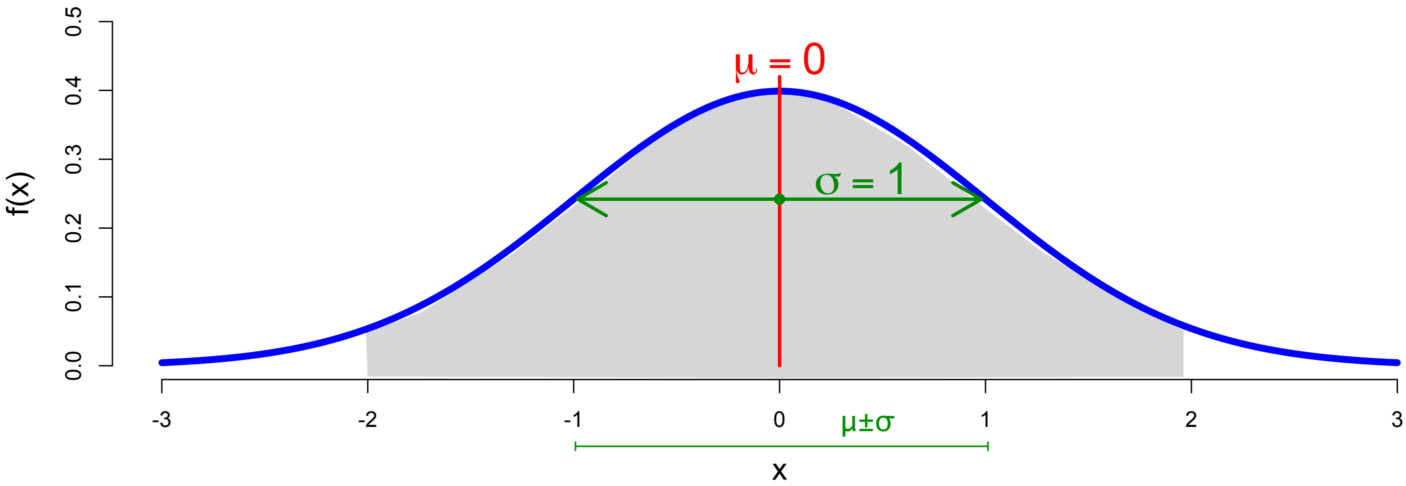
$Z \sim N(0,1)$ "*standard* (or *unit*) normal"

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

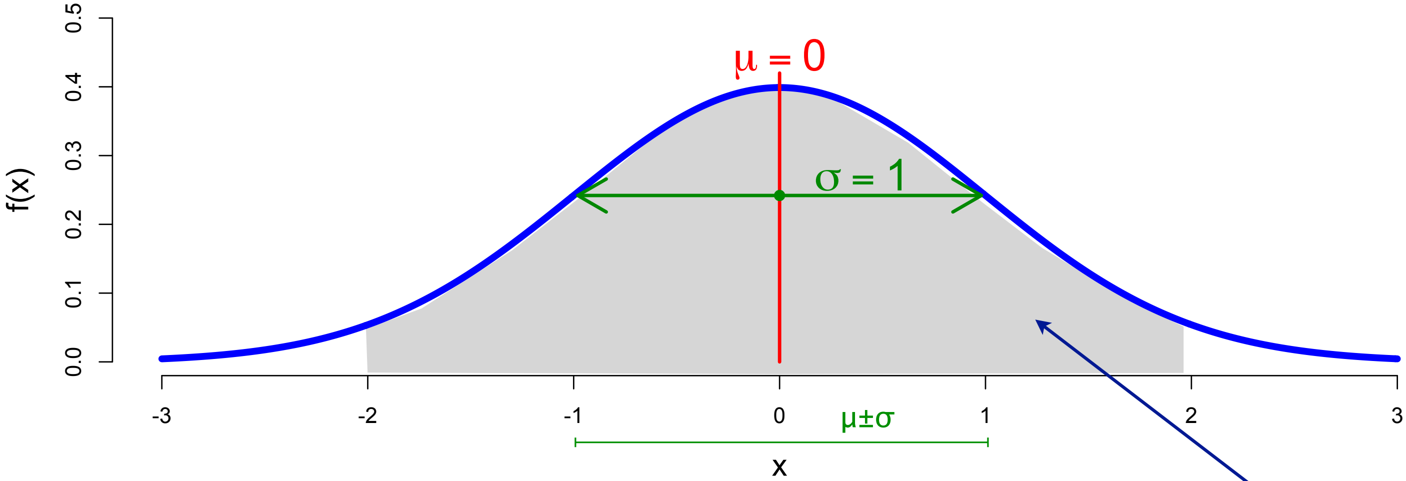
$$\Phi(z) = \Pr(Z \le z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx$$

no closed form ☹

The Standard Normal Density Function



The Standard Normal Density Function



If $Z \sim N(\mu,\sigma^2)$ what is $P(\mu-\sigma < Z < \mu+\sigma)$?

$$P(\mu - \sigma < Z < \mu + \sigma) = \Phi(1) - \Phi(-1) \approx 68\%$$

$$P(\mu - 2\sigma < Z < \mu + 2\sigma) = \Phi(2) - \Phi(-2) \approx 95\%$$

$$P(\mu - 3\sigma < Z < \mu + 3\sigma) = \Phi(3) - \Phi(-3) \approx 99\%$$

Why?

$N(\mu,\sigma^2)$ $N(0,1)$

$$\mu - k\sigma < \boxed{Z} < \mu + k\sigma \iff -k < \boxed{(Z-\mu)/\sigma} < +k$$

Consider i.i.d. (independent, identically distributed) random vars $X_1, X_2, X_3, \ldots$

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

Consider random variables

$$X_1 + X_2 + \ldots + X_n$$

and

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

Consider i.i.d. (independent, identically distributed) random vars $X_1, X_2, X_3, \dots$

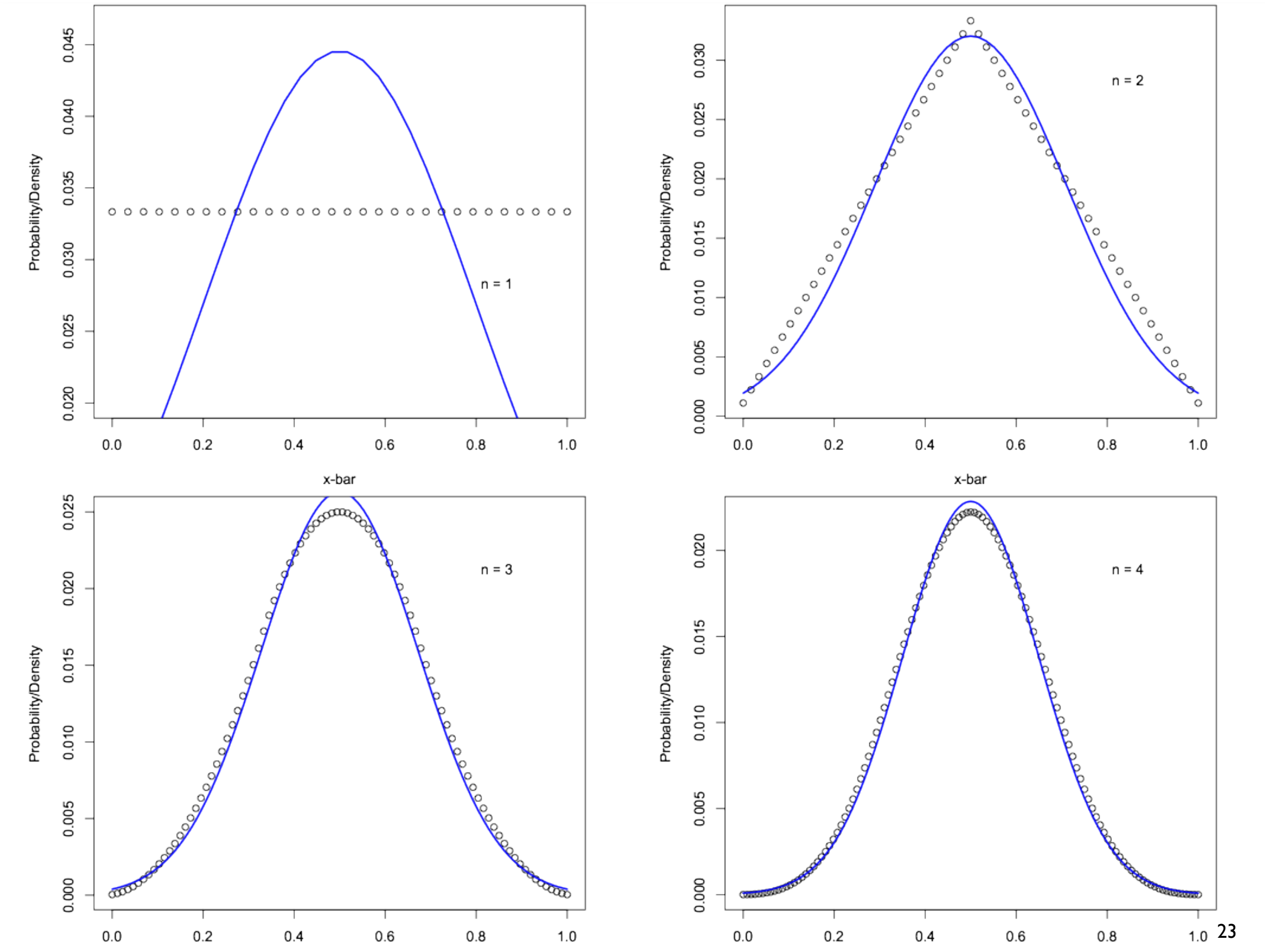
X_i has $\mu = E[X_i]$ and $\sigma^2 = Var[X_i]$

As $n \to \infty$,

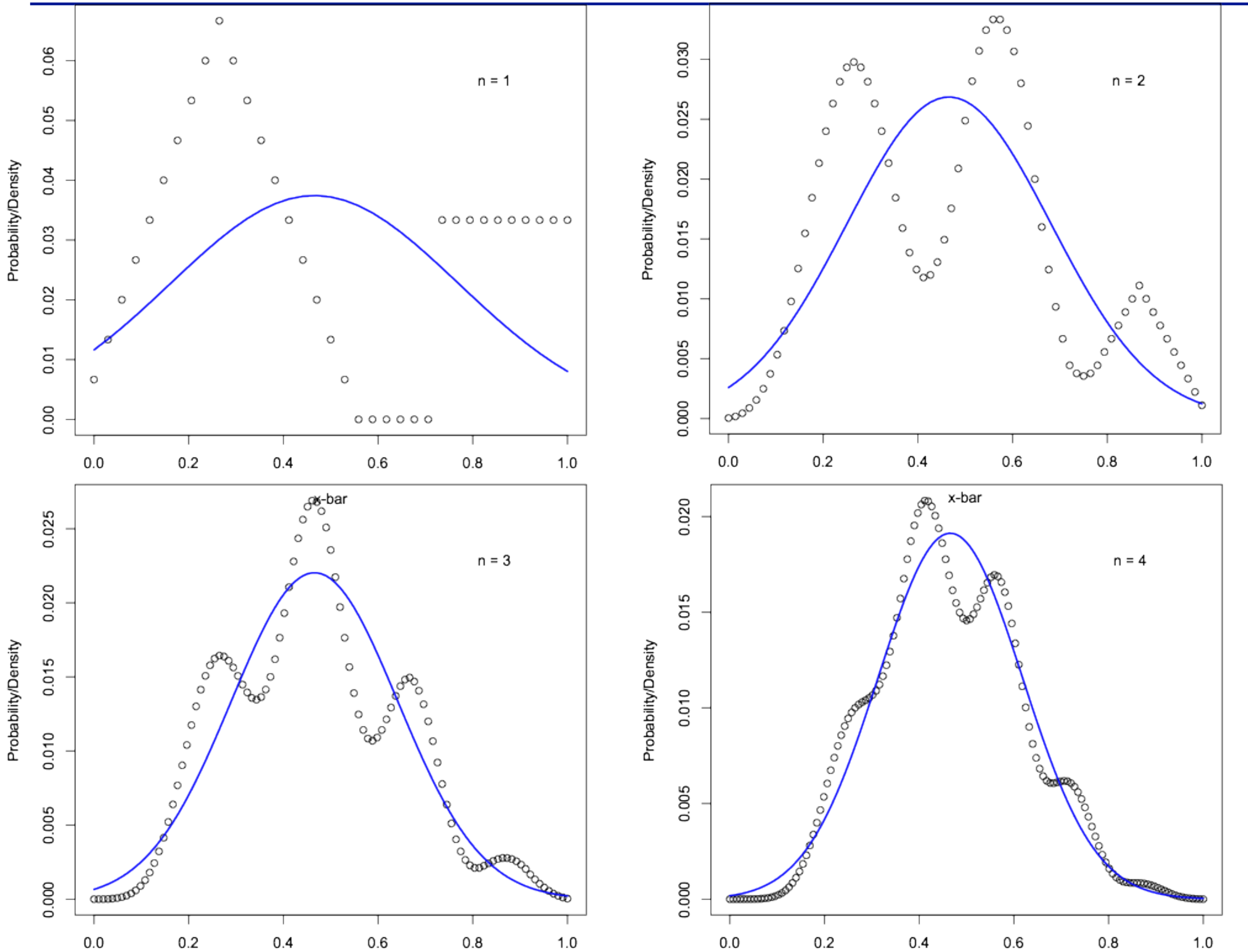
$$\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \longrightarrow N(0,1)$$

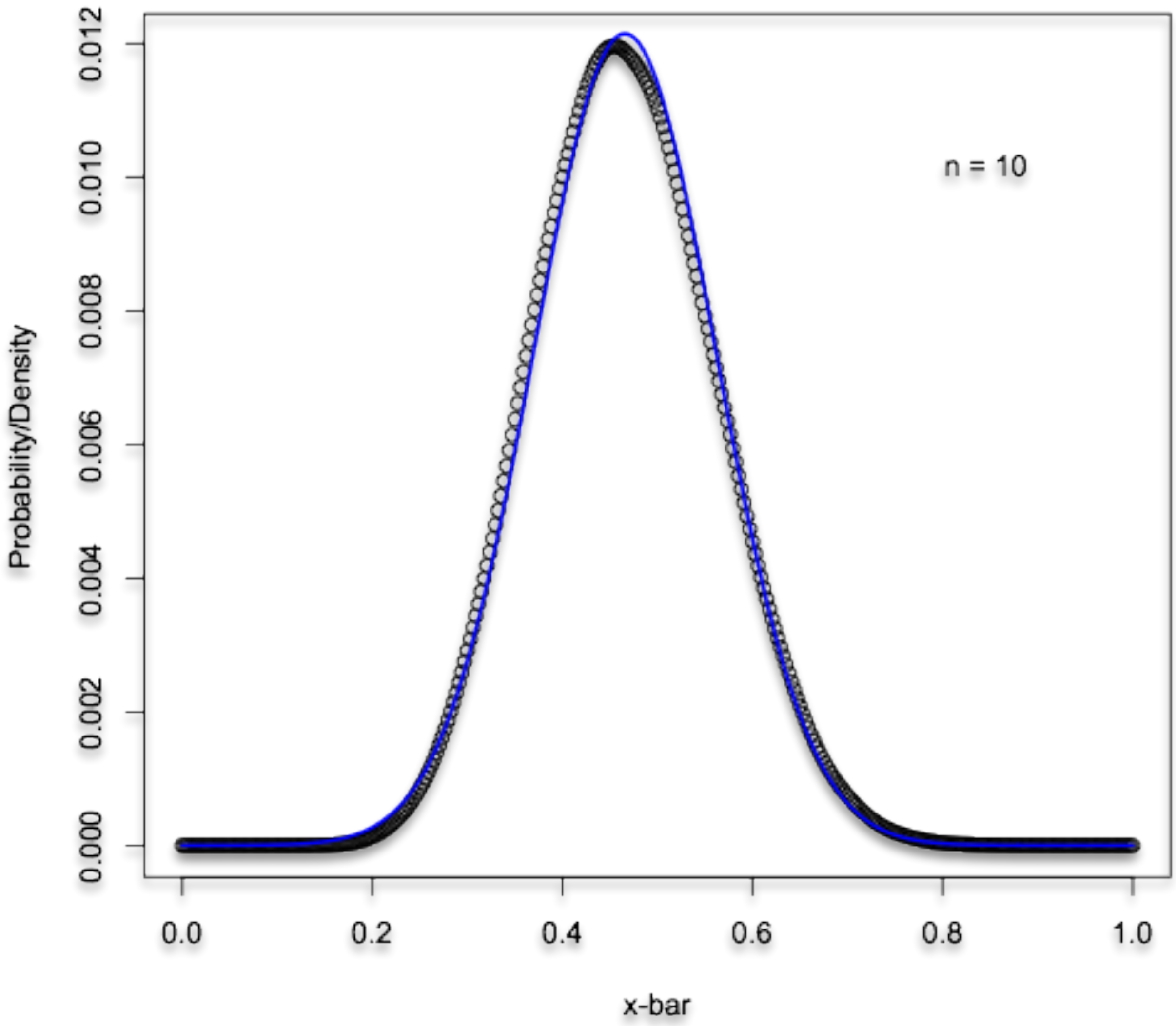
Restated: As $n \to \infty$,

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


Probability/Density
x-bar
n = 1
n = 2
n = 3
n = 4
0.0 0.2 0.4 0.6 0.8 1.0
0.020 0.025 0.030 0.035 0.040 0.045
0.000 0.005 0.010 0.015 0.020 0.025 0.030
0.000 0.005 0.010 0.015 0.020 0.025
0.000 0.005 0.010 0.015 0.020
23

CLT applies even to even wacky distributions




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

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
...