STA 291
Lecture 14, Chap 9

- **Sampling Distributions**
  - **Sampling Distribution of a sample statistic like the (sample) Proportion**

- Population has a distribution, which is fixed, usually unknown. (but we want to know)

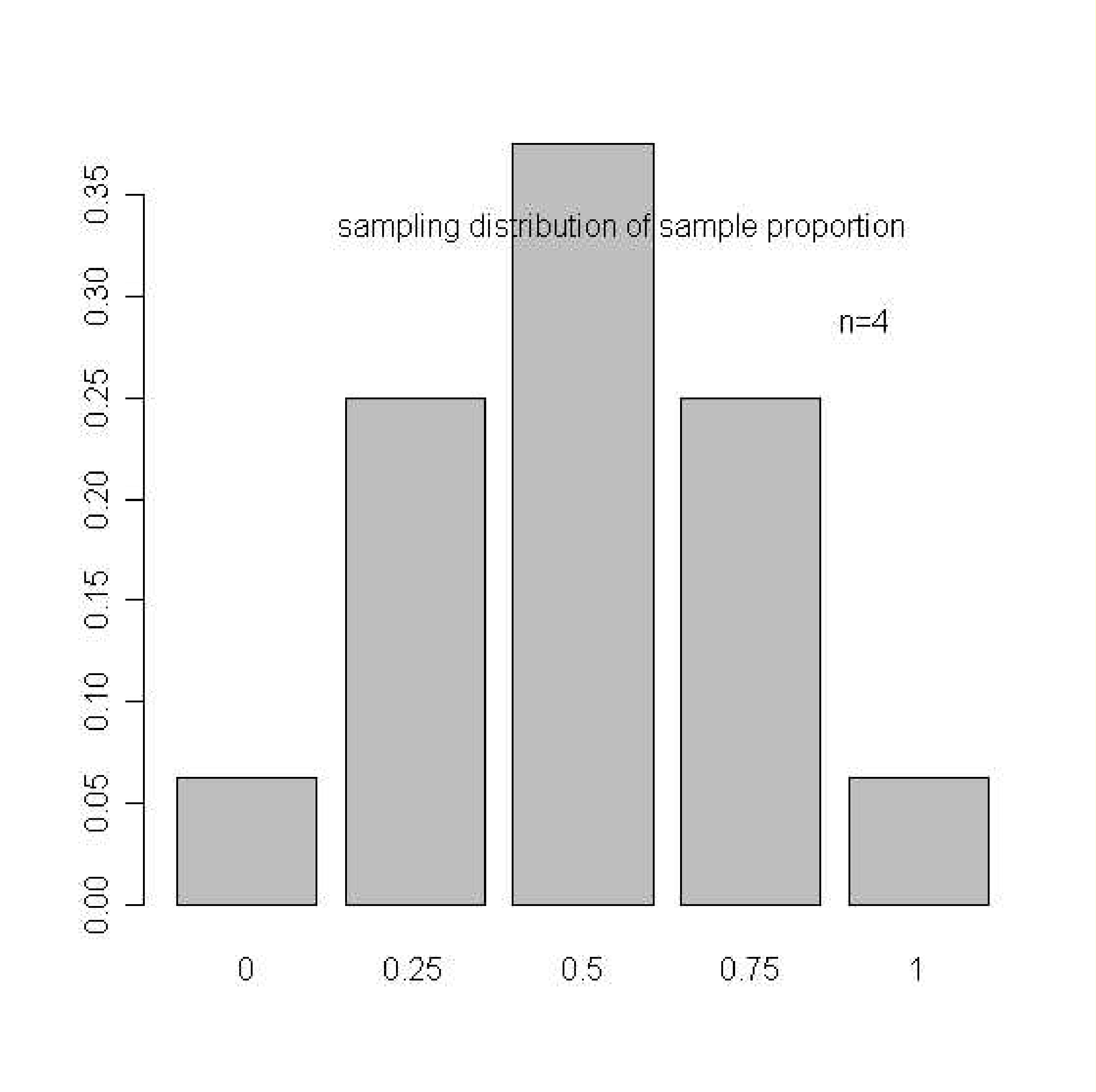
- Sample proportion changes from sample to sample (that's sampling variation). Sampling distribution describes this variation.

- Suppose we flip a coin 50 times and calculate the success rate or proportion.
- same as asking John to shoot 50 free throws (biased coin).

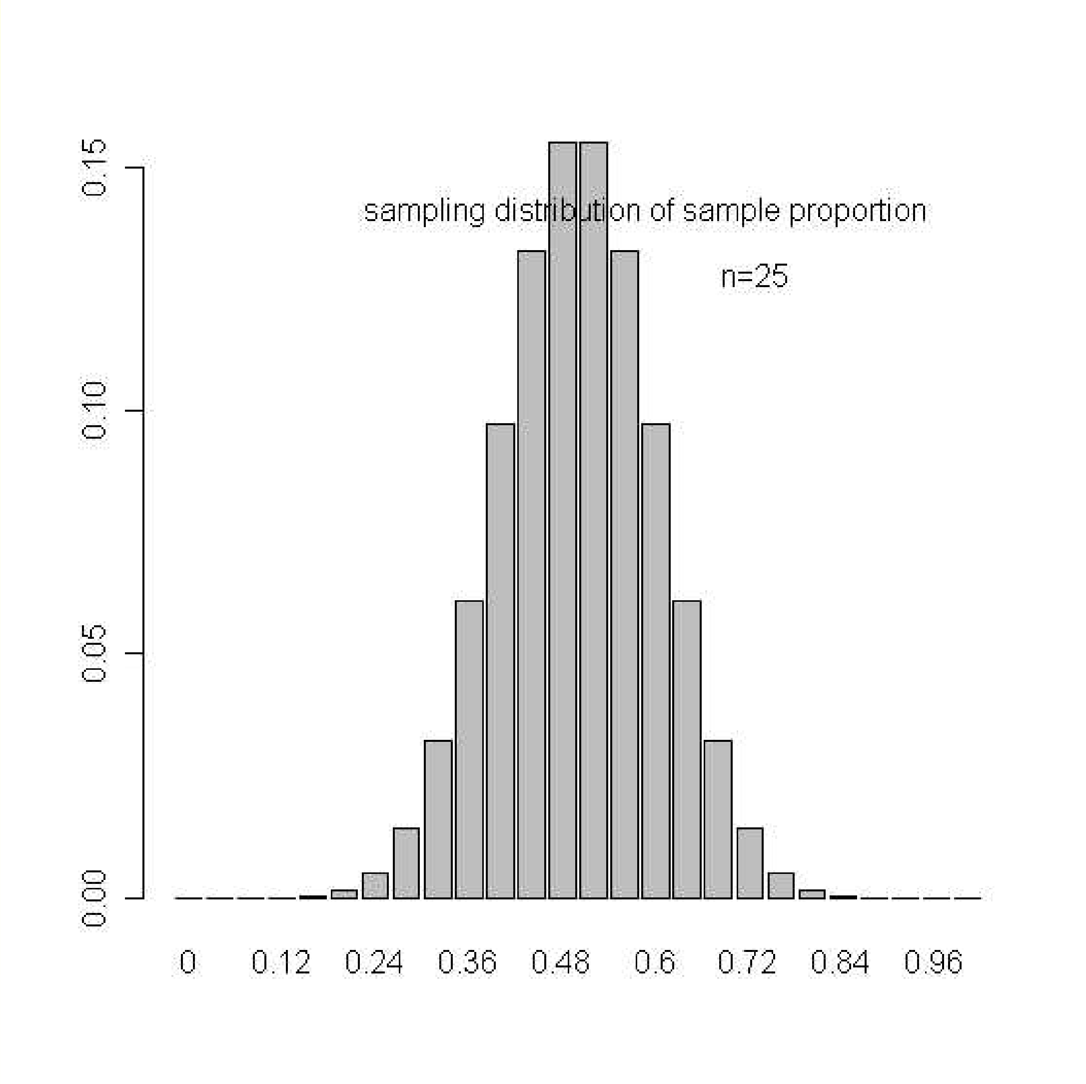
- Each time we will get a slightly different rate, due to random fluctuations.
- More we flip (say 500 times), less the fluctuation

- How to describe this fluctuation?
- First, use computer to simulate……
- Repeatedly draw a sample of 25, etc
- Applet:

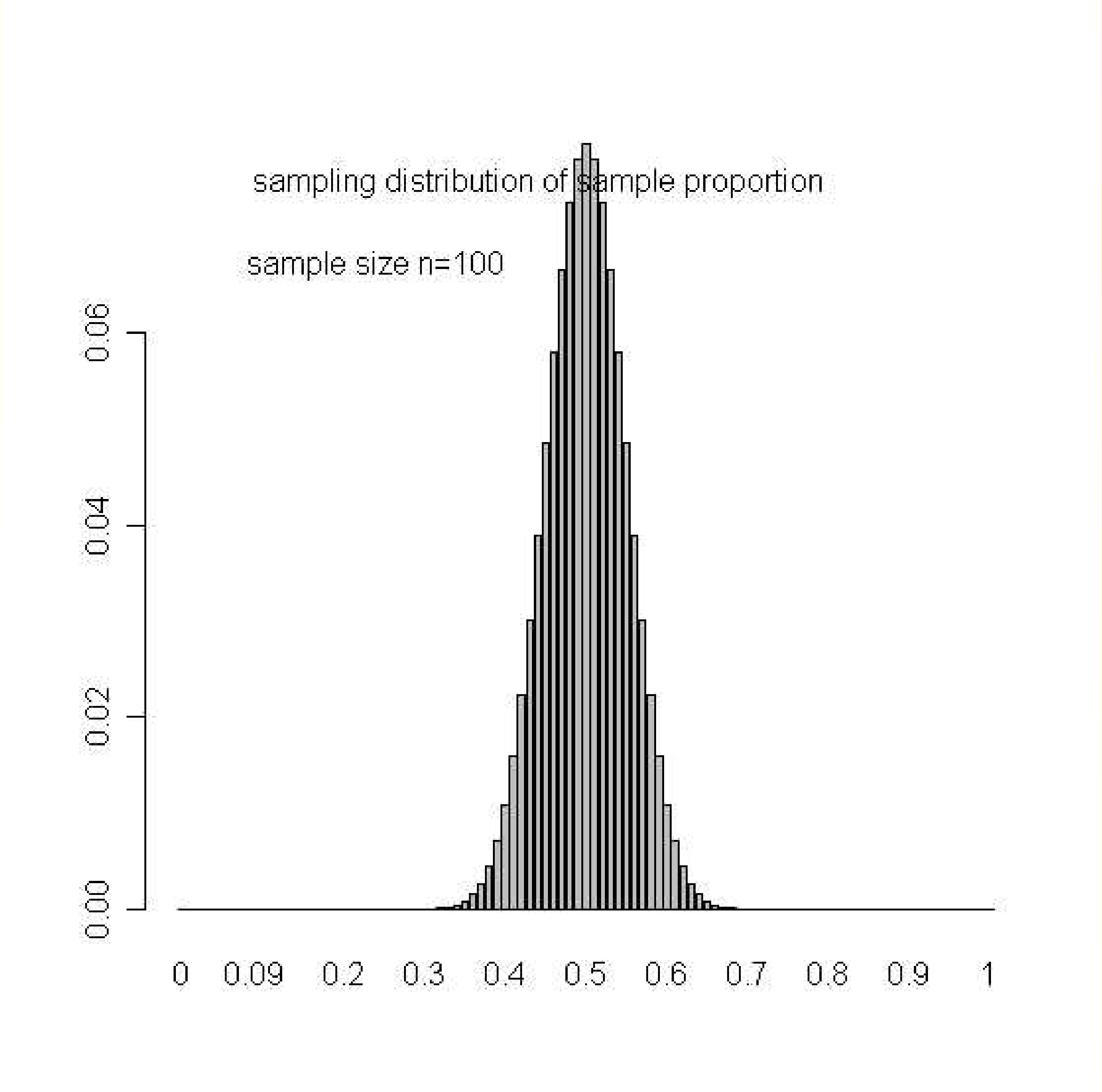
Sampling distribution of proportion



Sampling distribution: n=25



Sampling distribution: n=100



- Larger the n, less the fluctuation.

- Shape is (more or less) symmetric, bell curve.

Population distribution vs. sampling distribution

- For example, when population distribution is discrete, the sampling distribution might be (more or less) continuous.

- Number of kids per family is a discrete random variable (*discrete population*), but the sample mean can take values like 2.5 (*sampling distribution is continuous*).

Sampling Distribution: Example Details

- Flip a fair coin, with 0.5 probability of success (H). Flip the same coin 4 times.
- We can take a simple random sample of size 4 from all sta291 students.
- find if the student is AS/BE major.
- Define a variable X where

  X=1 if the student is in AS/BE, (or success)

  and X=0 otherwise

- Use the number "1" to denote success
- Use the number "0" to denote failure

Sampling Distribution: Example (contd.)

- If we take a sample of size $n=4$, the following 16 samples are possible:

(1,1,1,1); (1,1,1,0); (1,1,0,1); (1,0,1,1);
(0,1,1,1); (1,1,0,0); (1,0,1,0); (1,0,0,1);
(0,1,1,0); (0,1,0,1); (0,0,1,1); (1,0,0,0);
(0,1,0,0); (0,0,1,0); (0,0,0,1); (0,0,0,0)

- Each of these 16 samples is equally likely (SRS !) because the probability of being in AS/BE is 50% in all 291students

Sampling Distribution: Example (contd.)

- We want to find the sampling distribution of the statistic “sample proportion of students in AS/BE”
- Note that the “sample proportion” is a special case of the “sample mean”
- The possible sample proportions are

  0/4=0, 1/4=0.25, 2/4=0.5, 3/4 =0.75, 4/4=1
- How likely are these different proportions?
- This is the sampling distribution of the statistic “sample proportion”

Sampling Distribution: Example (contd.)

Sample Proportion of Students from AS/BE	Probability
0.00	1/16=0.0625
0.25	4/16=0.25
0.50	6/16=0.375
0.75	4/16=0.25
1.00	1/16=0.0625

- This is the sampling distribution of a sample proportion with sample size n=4 and P(X=1)=0.5

3 key features

- The shape getting closer to "bell-shaped", almost continuous – become "Normal"
- The center is always at 0.5 – (the population mean)
- The variance or SD reduces as sample size n gets larger.

Sample proportion

$$\hat{p}$$

Mean of sampling distribution

- FACT:

Mean/center of the sampling distribution for sample mean or sample proportion always equal to the same for all n, and is also equal to the population mean/proportion.

$$m_{\hat{p}} = p$$

Reduce Sampling Variability

- The larger the sample size n, the smaller the variability of the sampling distribution

- The SD of the sample mean or sample proportion is called *Standard Error*

- Standard Error = SD of population/$\sqrt{n}$

Normal shape

- Shape becomes normal type. -- *Central Limit Theorem*

Interpretation

- If you take samples of size n=4, it may happen that nobody in the sample is in AS/BE
- If you take larger samples (*n*=25), it is highly unlikely that nobody in the sample is in AS/BE
- The sampling distribution is more concentrated around its mean
- The mean of the sampling distribution is the population mean: In this case, it is 0.5

- The standard deviation of the sampling distribution of the mean is called “standard error” to distinguish it from the population standard deviation

Standard Error

- Intuitively, larger samples yield more precise estimates
- Example:
  - X=1 if student is in AS/BE, X=0 otherwise
  - The population distribution of X has mean p=0.5 and standard deviation

$$\sqrt{p(1-p)} = 0.5$$

Standard Error

- Example (contd.):
  - For a sample of size n=4, the standard error of $\bar{X}$ is

$$s_{\bar{X}} = \frac{s}{\sqrt{n}} = \frac{0.5}{\sqrt{4}} = 0.25$$

  - For a sample of size n=25,

$$s_{\bar{X}} = \frac{s}{\sqrt{n}} = \frac{0.5}{\sqrt{25}} = 0.1$$

  - Because of the approximately normal shape of the distribution, we would expect $\bar{X}$ to be within 3 standard errors of the mean (with 99.7% probability)

Central Limit Theorem

- For random sampling (SRS), as the sample size *n* grows, the sampling distribution of the sample mean $\bar{X}$ approaches a normal distribution
- **Amazing: This is the case even if the population distribution is discrete or highly skewed**
- The Central Limit Theorem can be proved mathematically
- We will verify it experimentally in the lab sessions

Attendance Survey Question

- On a 4"x6" index card
  - Please write down your name and section number
  - Today's Question:

Do you believe in the "Hot hand" claim?

Extra: March Madness

- Some statistics related to sports: how does a ranking system for basketball team work? Also the ranking of tennis players, chess players, etc.

Theory of “hot hand”

- One theory says, if a player gets hot (hit several 3 point in a row) then he is more likely to hit the next one. Because he has a hot hand……

Effect of Sample Size

- The larger the sample size n, the smaller the standard deviation of the sampling distribution for the sample mean

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

  - Larger sample size = better precision
- As the sample size grows, the sampling distribution of the sample mean approaches a normal distribution
  - Usually, for about n=30 or above, the sampling distribution is close to normal
  - This is called the "Central Limit Theorem"

Using the Sampling Distribution

- In practice, you only take ***one*** sample
- The knowledge about the sampling distribution helps to determine whether the result from the sample is reasonable given the population distribution/model.

- For the random variable X we defined

- Sample mean = sample proportion

- For example, our model was

P(randomly selected student is in AS/BE)=0.5

- If the sample mean is very unreasonable given the model, then the model is probably wrong

Sampling Distribution of the Sample Mean

- When we calculate the sample mean $\bar{X}$, we do not know how close it is to the population mean m
- Because m is unknown, in most cases.
- On the other hand, if ***n*** is large, $\bar{X}$ ought to be very close to m since the sampling distribution

must getting more concentrated.

- The sampling distribution tells us with which probability the sample mean falls within, say, 2 SD of the sample mean (empirical rule says 95%)

- How do we get to know SD of the sample mean – standard error?

Parameters of the Sampling Distribution

- If we take random samples of size ***n*** from a population with population mean m and population standard deviation s, then the sampling distribution of $\bar{X}$
  - has mean $$m_{\bar{x}} = m$$
  - and standard error $$s_{\bar{X}} = \frac{s}{\sqrt{n}}$$

Central Limit Theorem

- Usually, the sampling distribution of $\bar{X}$ is approximately normal for n 30 or larger
- In addition, we know that the parameters of the sampling distribution

Central Limit Theorem

- For example:

  If the sample size is n=36, then with 95% probability, the sample mean falls between

$$m - 2\frac{s}{\sqrt{n}} = m - \frac{2}{6}s \ = m - 0.333s$$

$$\text{and } m + 2\frac{s}{\sqrt{n}} = m + \frac{2}{6}s \ = m + 0.333s$$

$$(m = \text{population mean, } s \ = \text{population standard deviation})$$