ECO220Y

Sampling Distributions of Sample Statistics: Sample Proportion

Readings: Chapter 10, section 10.1-10.3

Fall 2011

Lecture 9

Sampling Distributions

Recall:

- A sample is a [small] part of a population.
- A parameter is a numerical fact about the population of interest. Usually, a parameter cannot be determined exactly, but can only be estimated.
- A statistic can be computed from a sample, and used to estimate a parameter.

Today:

- Sample statistics are random variables.
- Sampling distribution is a probability distribution of a sample statistic.
- Sampling error, or noise, is the variation of estimates from sample to sample.

How to Find Sampling Distribution

1. Analytically (✓)
   - Use probability rules
   - Use Laws of Expectation and Variance
   - Use Central Limit Theorem
2. Empirically
   - Toss 2 coins many times
   - Record the value of sample statistics
   - Record frequencies of each value and probabilities - probability distribution
3. Simulations
   - Monte-Carlo simulation
   - Boot-strapping

Sample Proportion

- Mr Noxin is running for a dogcatcher and 45% of all voters favour him.
- We polled 100 people on a street and found that 30% of them favour Mr Noxin
- We polled another 100 people on a street and found that 60% of them favour Mr Noxin
- Why the difference between two samples?
- How we can reconcile 30 and 60 percent with 45 percent who favour Mr Noxin?

Population and Sample Proportions

- Mr Noxin is running for a dogcatcher and 45% of all voters favour him
- 45% is what? Answer: Proportion, or fraction of voters who favour Noxin in a population
- 45%, or 0.45 is p, population proportion, parameter, constant
- What about the proportion of voters who favour Noxin in a sample of 3 voters? 100 voters? 1000 voters?
- Fraction of voters in a sample who favour Noxin is a sample proportion, $\hat{p}$
- Sample proportion, $\hat{p}$ varies from sample to sample
- $\hat{p}$ is statistic, random variable
- Let's "imagine" what sample proportion will be in a sample of 3 voters.

Sampling Distribution of $\hat{p}$ when $n = 3$

Sample	X	$\hat{p}$	Probability
FFF	3	$\frac{3}{3} = 1$	$0.45^3 = 0.091$
FFN, FNF, NFF	2	$\frac{2}{3}$	$0.45^2 * 0.55 * 3 = 0.334$
FNN, NFN, NNF	1	$\frac{1}{3}$	$0.55^2 * 0.45 * 3 = 0.408$
NNN	0	$\frac{0}{3} = 0$	$0.55^3 = 0.166$

What if sample size is 100? 1000?

Sampling Distribution of $\hat{p}$

- $\hat{p} = \frac{X}{n}$
- X counts the number of successes $\rightarrow X \sim \textit{Binomial}$
- $\hat{p}$ is a linear transformation of X
- $\hat{p}$ is also a Binomial random variable!
- Recall that we can use Normal approximation to Binomial to compute probabilities!
- Let's find parameters of the distribution of $\hat{p}$ when n is large

Distribution of $\hat{p}$

- Parameters of $X \sim B$ are n and p, and $E[X] = np$ and $V[X] = np(1-p)$
- Since $\hat{p} = \frac{X}{n}$, then $E[\hat{p}] = E[\frac{X}{n}]$ and $V[\hat{p}] = V[\frac{X}{n}]$
- $E[\hat{p}] = E[\frac{X}{n}] = \frac{1}{n}E[X] = \frac{1}{n} * np = \frac{np}{n} = p$
- $V[\hat{p}] = V[\frac{X}{n}] = \frac{1}{n^2}V[X] = \frac{1}{n^2} * np(1-p) = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$

For large enough n

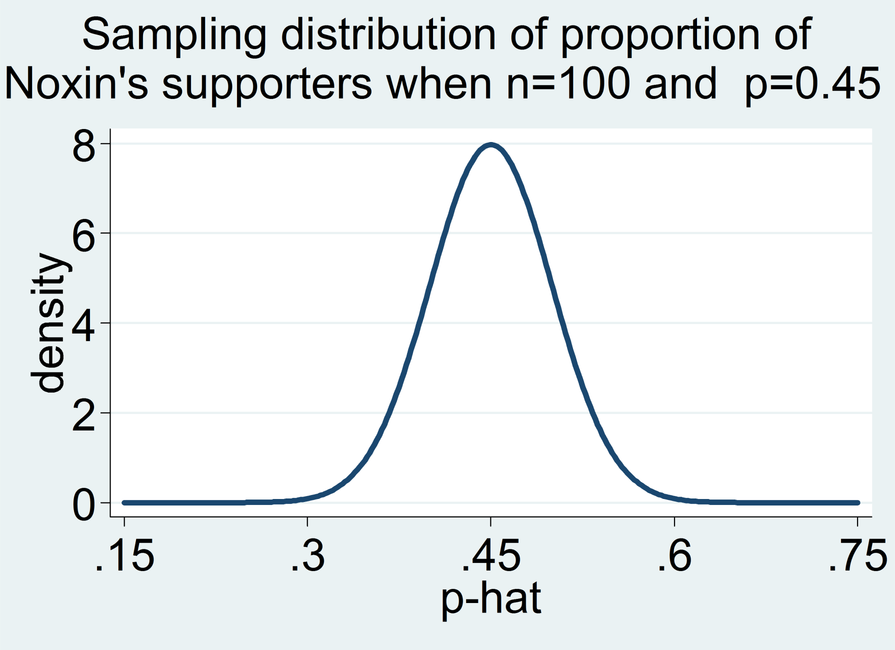
$$\boxed{\hat{p} \sim N(p, \frac{p(1-p)}{n})}$$

Rule of Thumb

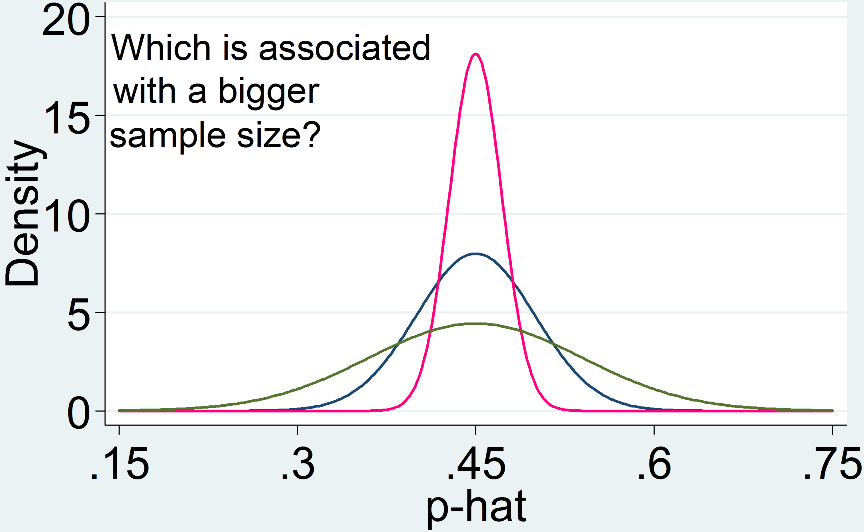
- Check if the entire interval $p \pm 3\sqrt{p(1-p)/n}$ lies within 0 and 1
- Intuition: Check whether Empirical Rule holds for that distribution
- Hint: Since $\hat{p}$ is a linear transformation of X, can use alternative rules of thumb for Binomial distribution.

Back to Mr Noxin

- In a sample with 100 voters, $\hat{p} \sim N(0.45, \frac{45*0.55}{100})$
- Rule of thumb: interval (0.30, 0.60) lies within 0 and 1 ✓
- Empirical rule: 99% of all the values should lie within 3 st. deviations from the mean
- Within 3 st.deviation in this case is between 0.30 and 0.60
- Because of the sampling error, the sample proportion varies in a sample and we may observe 30% and 60% of voters who favour Mr Noxin while the population proportion is 45%
- What about 25% or 70% of voters?


Sampling distribution of proportion of
Noxin's supporters when n=100 and p=0.45
density
0
2
4
6
8
.15
.3
.45
.6
.75
p-hat

Potential sampling distributions



Summary: Sample Proportion

- Sample proportion, $\hat{p}$, measures the proportion of "successes" in a sample
- Sample proportion is a random variable
- In samples with large enough n sample proportion is distributed normally with mean p and standard deviation $\sqrt{p(1-p)/n}$
- To check whether normal approximation works, use rule of thumb: interval $p \pm 3\sqrt{p(1-p)/n}$ lies between 0 and 1
- Note: Standard deviation of sample statistic is a measure of sampling error

Example

Assume that last year L'Oreal estimated the market share of its sunscreen product to be 30%. What is the chance that in a survey of 1000 consumers less than 280 said they prefer Ombrelle?

$$\text{Since we know that } \hat{p} \sim N\left(.3, \frac{0.3(1-0.3)}{1000}\right)\text{, we can find } P(\hat{p} < 0.28)$$

$$P(\hat{p} < 0.28 \| p = .3, \sigma_{\hat{p}} = 0.015, n = 1000) = P(Z < \frac{0.28-0.3}{0.015})$$

$$= P(Z < -1.33) \approx 0.091$$

Another way to think of it is that 99% of all values for sample proportion should lie within 3 s.d. from p.

$$p \pm 3\sqrt{\frac{p(1-p)}{n}} \Longrightarrow 0.255 \leq \hat{p} \leq 0.345$$

Back to Example

Alternatively, we can find $P(X < 280)$:

$$\bullet X \sim B(0.3, 1000)$$

$$\checkmark \text{Check rule of thumb:}$$
$$0 < (300 - 3\sqrt{300 * 0.7}, 300 + 3\sqrt{300 * 0.7}) < 1000$$

$$\bullet X \sim N(300, 210)$$

$$\bullet P(X < 280) = P(Z < \tfrac{280-300}{\sqrt{210}}) = P(X < -1.38) \approx 0.084$$