Math 1070-2: Spring 2008

Lecture 7

Davar Khoshnevisan

Department of Mathematics
University of Utah
http://www.math.utah.edu/~davar

February 27, 2008

An example

- *A WHO study of health: In Canada, the systolic blood pressure readings have a mean of 121 and SD of 16. A reading of 140 is considered a high blood pressure.*
- *What is the z-score for a blood pressure reading of 140?*
- Ans:

$$z = \frac{x - \mu}{\sigma} = \frac{140 - 121}{16} = 1.1875 \quad \text{(standard units)}$$

- *If systolic blood pressure in Canada has a normal distribution, what proportion of Canadians suffers from high blood pressure?*
- Ans: The area to the right of 140, which is the area—on a standard normal curve—to the right of 1.1875.
- Ans ≈ 0.1180

Normal approximation to binomials

(A recap)

- Recall: Suppose $X =$ no. of successes in a random sample of:
  - n independent success/failure trials
  - each trial has the same chance p of leading to a success.
- Then:
  - X has a binomial distribution with parameters n and p
  - The mean of this probability distribution is $\mu = np$
  - The SD of this probability distribution is $\sigma = \sqrt{np(1-p)}$
  - If n is large, then the distribution of X is approximately normally distributed with mean np and SD $\sqrt{np(1-p)}$

An example

- [The New Scientist, Jan 2002] "Tomasz Gliszczynski and Waclaw Zawadowski, statistics teachers at the Akademia Podlaska in Siedlce, received Belgian Euro coins from Poles returning from jobs in Belgium and immediately set their students spinning them."
- The bottom line: 250 tosses; 140 heads
- *Is the Belgian Euro biased, or is the previous sample due to chance variation?*
- Observation: X= no. of heads in the sample is binomial with $n = 250$ and $p = 1/2$, [assuming fair; i.e., chance variation.]
- mean = $250 \times \frac{1}{2} = 125$
- SD = $\sqrt{250 \times \frac{1}{2} \times (1 - \frac{1}{2})} \approx 7.9$
- Question: $P\{X \geq 140\} =?$

An example

- mean = $250 \times \frac{1}{2} = 125$
- SD = $\sqrt{250 \times \frac{1}{2} \times (1 - \frac{1}{2})} \approx 7.9$
- Question: $P\{X \geq 140\} =?$
- $z = \frac{140-125}{7.9} \approx 1.9$
- $P\{X \geq 140\} \approx 0.0287$ [from the normal table]

The sampling distribution

- Most people prefer proportions to total counts
- In a random success/failure sample of size n, the proportion of "successes" is the total number of successes $\div n$
- If prob. of success is p for each trial, then the total no. of successes in the sample is binom. with parameters n and p
- In this way we can find the distribution of the sample proportion ["the sampling distribution of proportions."]
- Deep Fact: If n is large, then the sampling distribution of proportions is $\approx$ normal with

$$\text{mean} = p \quad \text{and} \quad \text{SD} = \sqrt{\frac{p(1-p)}{n}}$$

- The mean = $(np)/n$ and the SD is the count-SD over n

An example

- Exit polls: 3160 California voters
- 54% supported Gov. G. Davis
- *Is this 50/50?*
- Question: If it were 50/50, then what are the chances that 0.54 of the sample votes for Gov. G. Davis?
- $p = 0.5$, $n = 1360$
- mean and SD of the sampling distribution for proportions:

$$\text{mean} = p = 0.5 \quad \text{and} \quad \text{SD} = \sqrt{\frac{p(1-p)}{n}} \approx 0.0136$$

- $P\{\text{sample proportion for GGD} \geq 0.54\} = ?$

An example

- $p = 0.5$, $n = 1360$
- mean and SD of the sampling distribution for proportions:

$$\text{mean} = p = 0.5 \quad \text{and} \quad \text{SD} = \sqrt{\frac{p(1-p)}{n}} \approx 0.0136$$

- $P\{\text{sample proportion for GGD} \geq 0.54\} = ?$
- z-score = $\frac{0.54-0.5}{0.0136} \approx 2.94$
- $P\{\text{sample proportion for GGD} \geq 0.54\} \approx 0.0016$
- Well ???? Is it 50/50?

A different problem

- What is the average age in the US?
- Take a large sample
- Sample average $\approx$ population average
- How well does the approximation work?
- The answer depends on the probab. distribution of $\bar{x}$—the sample avg
- Computation is hard

The sampling distribution for the sample mean

- $\bar{x} =$ sample average of an independent sample of size n
- Let $\mu =$ the population mean; $\sigma =$ the population SD
- The probability distribution of $\bar{x}$ ... very complicated
- Fact 1: The mean of that distribution is μ
  [This means that we expect $\bar{x}$ to be μ]
- Fact 2: The SD of that distribution is $\frac{\sigma}{\sqrt{n}}$
- Deep Fact: The probab. distribution of $\bar{x}$ is approx. normal with mean μ and SD $\frac{\sigma}{\sqrt{n}}$, if n is large
- Fact 3: If n is large then σ is close to the sample SD with high probab.
  [Useful if σ is unknown]

An example

- A university has 30,000 registered students. As part of a survey, 900 of these students are chosen at random.
- Average age in the sample = 22.3 years, SD = 4.5 years.
- Someone proposes that the average student-age at this university is 22 years. What do you think about this proposal?
- $n = 900$, $\mu = 22.3$, $\sigma = 4.5$, SD of the sampling distribution is $\frac{\sigma}{\sqrt{n}} = \frac{4.5}{\sqrt{900}} = 0.15$
- $z = \frac{22-22.3}{0.15} \approx -2$
- probab. of seeing a sample avg of 22 or less is about 0.0.028
- Well ?????

[Freedman, Pisani, Purves (1998), Ch. 23]

A lottery example

- You bet \$1 on a number between 0 and 9, randomly selected
- If correct, then win \$1
- If wrong, then win \$0
- If you play 52 times, then what is the probability that your average winnings is at least \$0.50? [i.e., total win = $0.5 \times 52 = 26$]
- Population = ?
- $\mu = (1 \times \frac{1}{10}) + (0 \times \frac{9}{10}) = 0.1$
- Fact: $\sigma = 1.5$, so the SD of the sampling distribution is $\frac{\sigma}{\sqrt{n}} = \frac{1.5}{\sqrt{52}} \approx 0.21$
- $z = \frac{0.4 - 0.1}{0.21} \approx 1.42$
- $P\{\bar{x} \geq 0.5\} \approx 0.0778$

Final remarks

- This theory is applicable **only** to random samples
- The SD of the sampling distribution is sometimes called the *standard error*
  - For sample proportions:
  $$\text{SE} = \sqrt{\frac{p(1-p)}{n}}$$
  - For sample means:
  $$\text{SE} = \frac{\sigma}{\sqrt{n}}$$
  - How large is a large n?
  - A rule of thumb for proportions: want np and $n(1-p)$ to be $\geq 15+$