Mathematics 2

Lecture 6: From Probability to Inference

Danica Solina
School of Mathematical & Physical Sciences

Lecture 6 Objectives

- Be able to calculate the expectation and variance of a linear combination of independent random variables.
- Understand an be able to apply the Central Limit Theorem.
- Understand the differences between calculating probabilities on X and probabilities on $\bar{X}$.
- Be able to compute and interpret a confidence interval by hand, and be able to locate a confidence interval in software output.
- Be able to compute the sample size required to construct a confidence interval of given width.

Linear Combinations of Random Variables

Reference: Devore §5.5

- Often we are interested in not just a single random variable, but a linear combination of random variables.
- Let X_1, X_2, ... ,X_n be random variables with $E(X_i) = \mu_i$ and
  $Var(X_i) = \sigma_i^2$, and the random variable Y is defined by

$$\boxed{Y = a_1 X_1 + a_2 X_2 + \ldots + a_n X_n}$$

- If the random variables are independent of each other, then the expectation and variance of the linear combination take a useful form.

Linear Combinations of Random Variables

- If the random variables X_i are independent of each other, we can write the expected value and variance of Y as

$$\boxed{\begin{aligned} \mathrm{E}(Y) &= a_1\mu_1 + a_2\mu_2 + \ldots + a_n\mu_n \\ \mathrm{Var}(Y) &= a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \ldots + a_n^2\sigma_n^2 \end{aligned}}$$

  respectively

- Furthermore, if the X_i are independent and normally distributed, then we can also say that Y is normally distributed.

Linear Combinations of Random Variables

Example

- Suppose that the random variables X_1 and X_2 denote the length and width respectively of a manufactured part. Assume that X_1 is normal with mean 2cm and variance 0.1 cm, and X_2 is normal with mean 5cm and variance 0.2 cm. Assume X_1 and X_2 are independent. What is the probability the perimeter exceeds 14.5 cm?

Linear Combinations of Random Variables

Example

- Let $X \sim N(\mu = 100, \sigma = 15)$ and $Y \sim N(\mu = 30, \sigma = 5)$ be independent random variables. Find the distribution of $X - Y$

Linear Combinations of Random Variables

Exercise

- Let X_1, X_2, and $X_3 \sim N(\mu = 9, \sigma = 4)$ be independent random variables. Find the distribution of the mean of X_1, X_2, and X_3

Central Limit Theorem

Reference: Devore §5.4

- A set of independent random variables X_1, X_2, ... , X_n with the same distribution is called a random sample from that distribution.
- The probability distribution of a statistic is called the **sampling distribution**.

Central Limit Theorem

- If we draw a random sample of size n form a population with mean μ and variance σ^2, then if n is sufficiently large, the sampling distribution of the sample mean is approximately

$$\bar{X} \sim N\left(\mu, \sigma_{\bar{X}}^2 = \frac{\sigma^2}{N}\right)$$

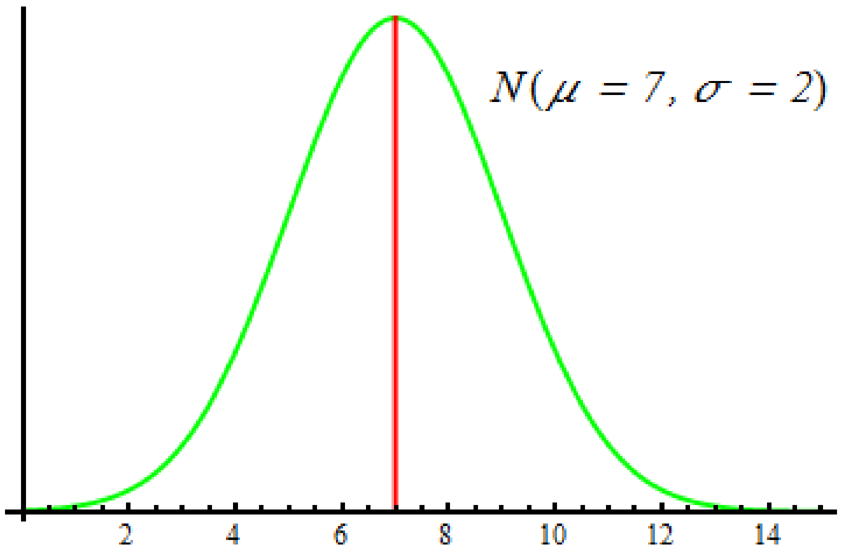
Central Limit Theorem

- When the population distribution is normal

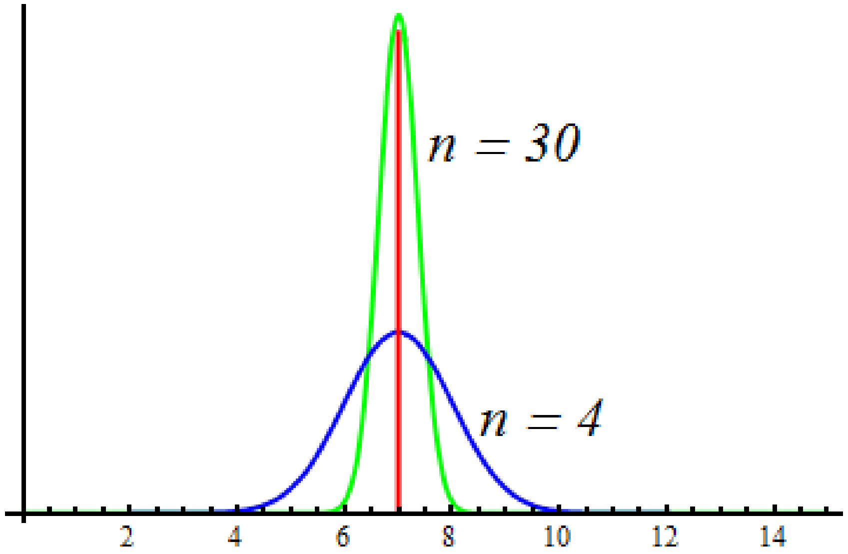
Central Tendency: $E(\bar{X}) = \mu$

Variation: $\sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}}$

Population Distribution



Sampling Distributions



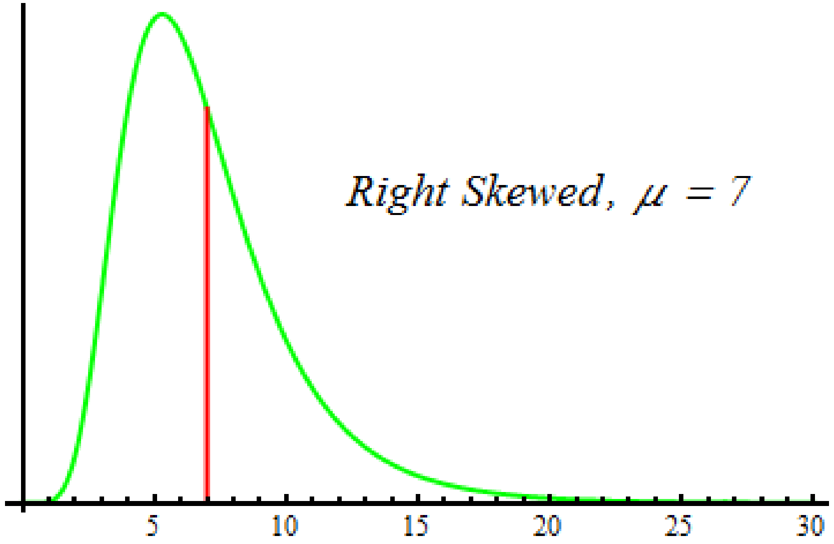
Central Limit Theorem

- When the population distribution is not symmetric

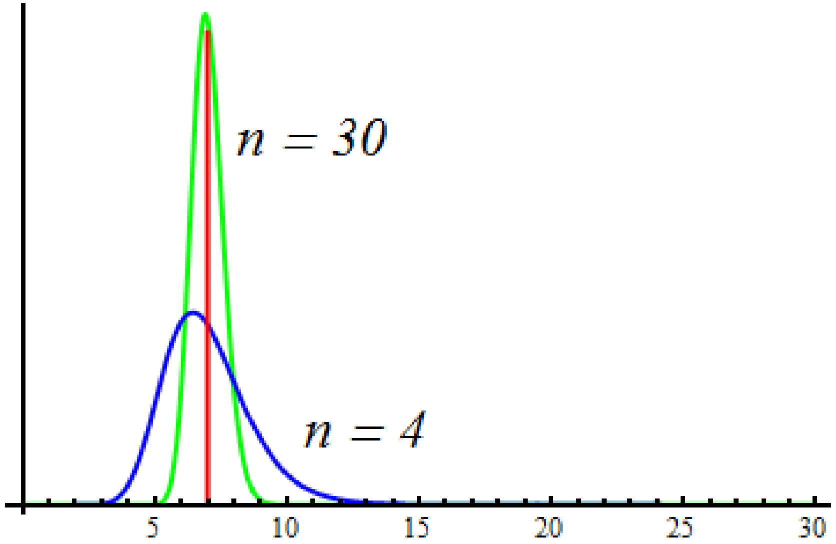
Central Tendency: $E(\bar{X}) = \mu$

Variation: $\sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}}$

Population Distribution



Sampling Distributions



How large is enough?

- For most distributions, $n > 30$
- For fairly symmetric distributions, $n > 15$

- For normal distribution, the sampling distribution of the mean is **always** normally distributed, regardless of the sample size.
  - Note: this is NOT a result of the central limit theorem but a property of sampling from a normal population distribution (as demonstrated in the example on Slide 5).

CLT

Central Limit Theorem - Example

- Soft drink cans are filled by an automatic filling machine. The mean fill volume is 12.1 fluid ounces, and the standard deviation is 0.1 fluid ounce. Assume the fill volumes of the cans are independent, normal random variables. What is the probability that the average volume of 10 cans selected from this process is less than 12 fluid ounces?

Central Limit Theorem - Example

- Engineers responsible for the design and maintenance of aircraft pavements use pavement quality concrete. A series of tests were carried out to determine the load classification number (LCN), a measure of breaking strength, for a particular mix.
    - Suppose that the original pavement quality concrete usually has a mean LCN of 60, and standard deviation of 10. What would we expect the distribution for the mean of the 25 samples to be?
    - Under this assumption, what is the probability that the sample mean is greater than 65?

Central Limit Theorem - Example

Central Limit Theorem - Exercise

An electronics company manufactures resistors that have a mean resistance of 100 and a standard deviation of 15. Find the probability that a random sample of n=30 resistors will have an average resistance less than 95.

Estimation

Reference: Devore §7.1

- One of the most common goals when performing an experiment/study is to obtain a good estimate for a population statistic (mean, standard deviation etc.)
- While it is of some use to have a point estimate for a statistic, it is better to be able to give a range of plausible values for the statistic.
- A **Confidence Interval** is a range of values such that IF the sample was repeated many times (1- α)% would contain the population statistic.
  - α is the proportion of confidence intervals that do not contain the population mean.

Confidence Intervals

- A confidence interval provides us with a range that is quite likely to contain the true mean.

If $\bar{x}$ is the sample mean of a random sample of size n, from a population with known variance σ^2, a $(1- \alpha)\%$ confidence interval on μ is given by

$$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

where $Z_{\alpha/2}$ is the upper $\alpha/2\%$ point of the standard normal distribution.

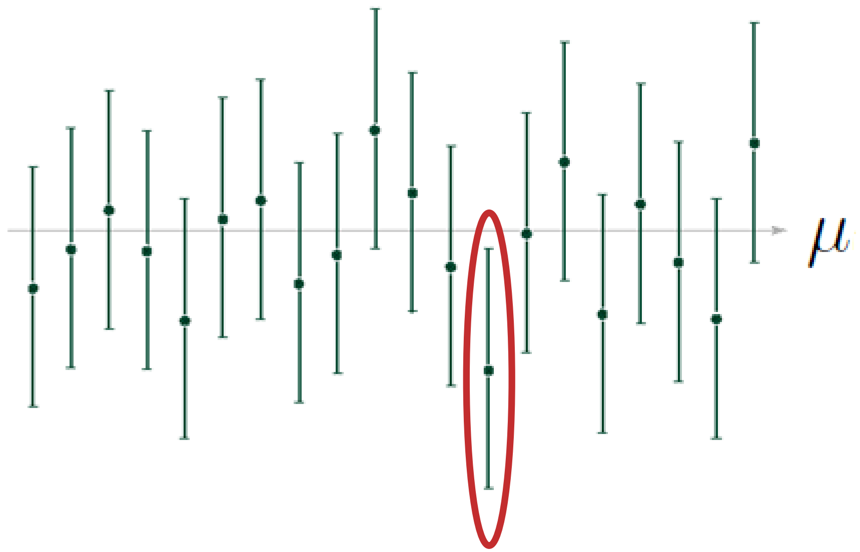
Confidence Intervals

- A table of upper percentage points for the standard normal distribution.
  - Included in the set of tables that are given in the exam.

α	z_α
0.100	1.2816
0.050	1.6449
0.025	1.9600
0.010	2.3263
0.005	2.5758
0.001	3.0902

Confidence Intervals

- We see that the 95% confidence interval doesn't always contain the true mean
  - An unusual sample.



CI

Confidence Intervals - Example

- A factory producing a chemical buys a new machine, and collects data over the next 30 days on the production from the machine.
- From past data, the standard deviation is known to be 21 tons, and over the 30 days the average daily production is 440 tons.
- What would be the best single estimate of daily production of the machine?

One-Sample Z

```
The assumed standard deviation = 21

 N     Mean  SE Mean        95% CI
30   440.00     3.83  (432.49, 447.51)
```

Confidence Intervals - Example

Find a 99% confidence interval for μ by hand.

Note: There is a trade-off between precision and confidence.

Confidence Intervals - Example

The melting points of a sample of 50 alloy filaments were collected, and had a mean of 320°C. Assume that the population standard deviation is 5°C. Calculate a 95% confidence interval for the population mean.

Determining Sample Size

- The confidence interval for μ will have a specified margin of error E when the sample size is

$$n = \left(\frac{z_{\alpha/2} \times \sigma}{E}\right)^2$$

- So if we want to be able to estimate μ to within E units either side, then we need

$$\boxed{n \geq \left(\frac{z_{\alpha/2} \times \sigma}{E}\right)^2}$$

Determining Sample Size - Example

What sample size is needed to be 90% confident of being correct within ± 5? A pilot study suggested that the standard deviation is 45.

Determining Sample Size - Example

On how many days would we need to collect data if we want to estimate average yield of the new machine with 90% confidence with a margin of error of $\pm$ 2?

Lecture 6 Revision Exercises

Use software where appropriate

- Devore §5.4 Q49, 51, 53, 55b
- Devore §5.5 Q59, 61, 65, 67
- Devore §7.1 Q1, 2, 3, 4, 5, 6

Montgomery, Runger and Hubele (Old Text)

- Q3.173 - 3.181, 3.195 – 208, 4.35, 4.36, 4.44, 4.45, 4.46