Ch. 1: Overview & Descriptive Statistics

§1.1 Populations and Samples

§1.2 Pictorial Methods

§1.3 Measures of Location

§1.4 Measures of Variability

Count: slides 1–35 (35 slides).

([*] denotes content that will be given in class)

What is Statistics Used For?

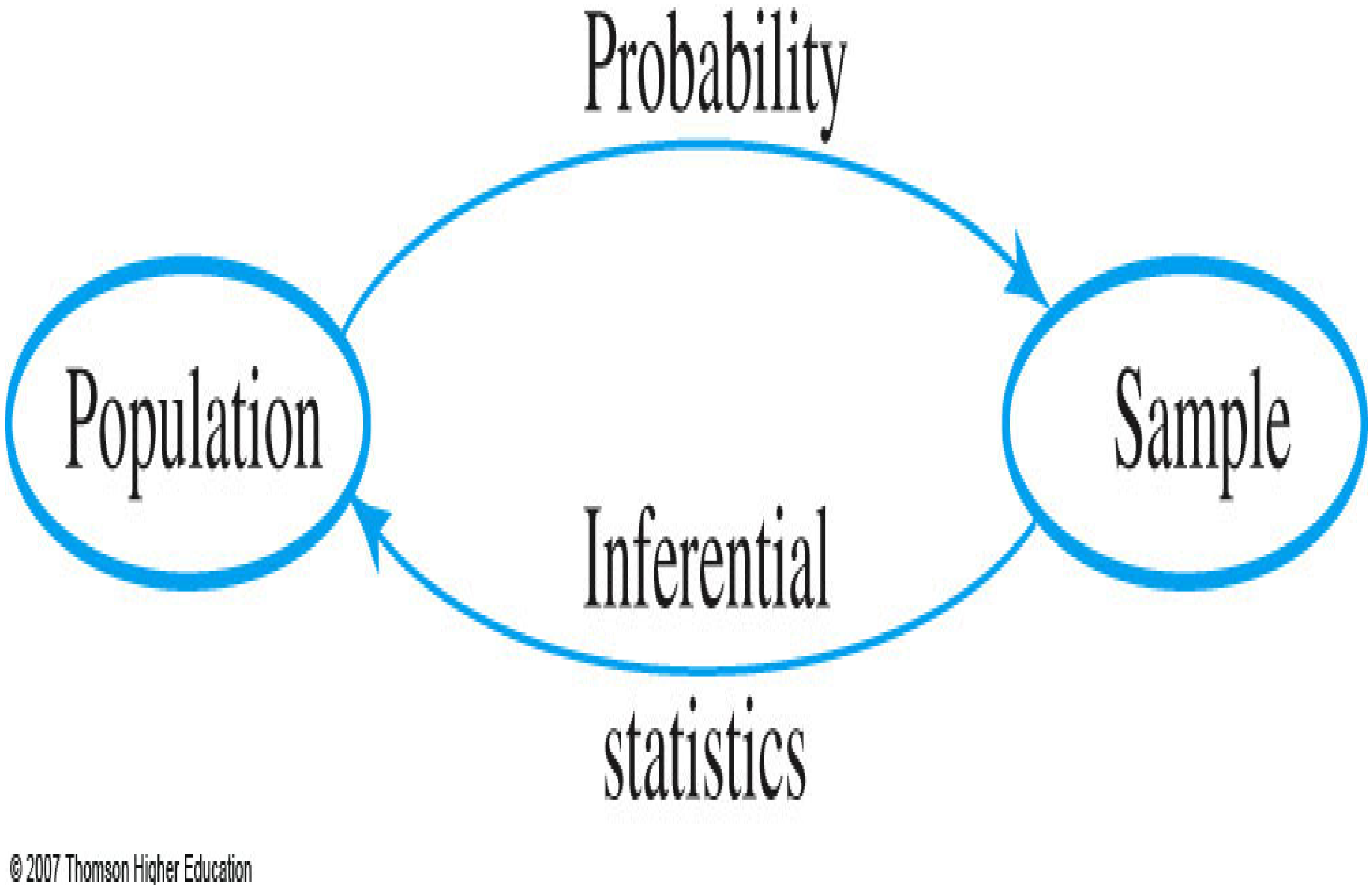
Statistics is used to:

make intelligent judgements and informed decisions in the presence of uncertainty

In more detail, some of these uses are:

- Designing experiments to collect data
- Extracting information from the data
- Making decisions and predictions in the presence of uncertainty and variation (a.k.a. data mining; predictive analytics; machine learning)

Fig 1: Population & Sample [*]



§1.1 Populations and Samples

- Population: the universal set of all objects under study, and it can be:
  - *real* or *concrete* (e.g., college students in USA)
  - *virtual* or *hypothetical* (e.g., all parts that could be produced by a specific machine)
- Sample: any subset of the population (usually collected in a prescribed manner)
- Variable: a characteristic of an object, often a *numeric* measurement (e.g., weight), but can also be a *category* (e.g., gender)

Population and Sample: Examples

- Ex 1

  Study aim : Investigate American people's preference between Republicans and Democrats

  - Population: the preference of all Americans
  - Sample: the preference of all Texans

- Ex 2

  Study aim: Determine if a given coin is fair or not

  - Population: the result from an infinite number of experiments
  - Sample: the result from 10 experiments.

Descriptive Statistics

- Descriptive statistics summarize/describe important features of the data, either graphically or numerically
- Mostly created by computer packages
- Challenger Data:
  In 1986 the space shuttle *Challenger* exploded after launch. Ex 1.1 in book (7th Ed.) gives the temperature (^0F) at each of 36 launches. Summary statistics are:

$$\text{min} = 31.00, \quad \text{median} = 67.50, \quad \text{mean} = 65.86, \quad \text{max} = 84.00$$

Challenger Data: Stem & Leaf Plot [*]

Inferential Statistics

- Descriptive statistics by themselves provide some information but do not provide *conclusions*
- *Inferential statistics* methods allow us to draw conclusions from data
- Because of *variation* in the data, we cannot draw guaranteed conclusions ...
- We must phrase the conclusions as *probabilistic* statements (confidence intervals; hypothesis tests)

Data Collection Methods

- Simple Random Sample (SRS): a sample drawn from a well-defined pop. in such a way that all possible samples of the same size have the same probability of being selected. (The simplest of all sampling schemes.)
- Stratified Sample: Separate pop. into groups (strata), and take samples from each one.
  Ex: 3 types of DVD player manufactured, so we sample from among customers that bought each type to ascertain overall satisfaction.

- Convenience Sample: Easily available and collected without systematic randomization (may entail some danger in that it may not be representative of entire pop.)
  Ex: collection of bricks stacked in such a way that those in middle harder to get, so we sample from those on top only
- Notation. Sampled data values:

$$x_1, x_2, \ldots, x_n$$

The ordered sampled values are called *order statistics*:

$$x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$$

Types of Studies and Experiments

- Enumerative vs. Analytic Studies:
  - Enumerative: finite, identifiable, unchanging collection of objects that make up the population. Ex: sample 10 furnaces from those already manufactured this year, to make inference on the lifetime of the units.
  - Analytic: pop. is not finite or identifiable. Ex: sample 10 furnaces from those to be manufactured over the course of this year.

- Designed Experiments: common in engineering; experimental conditions are purposefully manipulated to observe their effect on an outcome variable of interest.
  Ex: assign 3 corrosion protection methods randomly to each of 3 sets of 10 pipes.

- Observational Studies: are in contrast to designed experiments in that the experimental conditions cannot be purposefully manipulated. (Typically cannot ascribe *causality*.)
  Ex: observe the incidence of lung cancer among smokers and non-smokers.

§1.2 Pictorial Methods

- Histogram: plot with bars approximating the density function of continuous numeric data (called *barplot* for discrete data).

- Stem-and-Leaf plot: A type of histogram that separates digits into "stems" and "leaves".

- Probability plot (or QQ-plot): gives an indication of whether a dataset could be normally distributed (later).
- Box-and-Whisker plot: A pictorial representation of the data via a 5-number summary:

$$\text{min}, \quad \text{1st quartile } (Q_1), \quad \text{median}, \quad \text{3rd quartile } (Q_3), \quad \text{max}$$

Ex: Golf Course Yards (Stem & Leaf) [*]

Ex 1.7 in book (7th Ed.) gives total yardage of a sample of 40 golf courses. Summary statistics are:

$$\text{min} = 6433, \quad Q_1 = 6674, \quad \text{median} = 6872, \quad Q_3 = 7042, \quad \text{max} = 7280$$

Constructing Barplots & Histograms:

- Barplot: simply place a bar at each (discrete) data value with height equal to the frequency (or relative frequency) of that value.
- Histogram: divide data into equal-width *bins* (class intervals), and determine number of *bins* ($\approx \sqrt{n}$). Then set:
  - bin width $\approx \frac{x_{(n)} - x_{(1)}}{\text{number of bins}}$,
  - start 1st bin at (or just below) $x_{(1)}$
  - determine frequency of each class, and mark bin boundaries on x-axis
  - draw a rectangle of height equal to the frequency (or relative freq.) above each bin.

Ex: Power Consumption [*]

Ex 1.10 in book (7th Ed.) gives electrical energy consumption of a sample of 90 homes. Summary statistics are:

$$\min = 2.97, \quad \max = 18.26, \quad \Longrightarrow \quad \text{bin width} \approx \frac{18.26 - 2.97}{\sqrt{90}} \approx 2$$

Unequal Bin-Width Histograms

- It's possible (but unusual) to use different bin widths, e.g., highly skewed data.
- In this case it's important to make the area of each bar proportional to the frequency/relative freq. (Failure to do so is deceptive...)
- Steps (for relative freq.) are as before, except:
  - determine relative freqs. as before (after choosing bin widths),
  - rectangle height $= \frac{\text{relative freq.}}{\text{bin width}}$,

  Note: also works for equal bin-width histograms; vertical scale is then a *density* function. (See book for Exs.)

Categorical Data

- With *categorical* data we observe only the category or label of the outcome, not a numerical value.
- Graphical displays for this kind of data are limited, usually only *pie-charts* or *bar-plots*.
- Ex: Prob. 1.29 in book (7th Ed.) gives 60 obs on the type of health complaint:

  B=back pain: 7 obs.

  C=coughing: 3 obs.

  F=fatigue: 9 obs.

  J=joint swelling: 10 obs.

  M=muscle weakness: 4 obs.

  N=nose running/irritation: 6 obs.

  O=other: 21 obs.

Health Complaints Ex: Bar-plot [*]

§1.3 Measures of Location

▷ <u>Mean:</u> the sample mean is given by

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

If each data value is a point mass on the number line, then $\bar{x}$ would be the balance point (center of mass). The sample mean $\bar{x}$ is an *estimator* of the population mean μ (later).

▹ Median: the sample median is the "middle" value in the ordered data $x_{(1)} \le \cdots \le x_{(n)}$, i.e.

$$\tilde{x} = \begin{cases} x_{([n+1]/2)}, & \text{if } n \text{ is odd} \\ \frac{x_{(n/2)} + x_{(n/2+1)}}{2}, & \text{if } n \text{ is even} \end{cases}$$

Ex: median of $\{9, 4, 17\}$ is $\tilde{x} = 9$.

Ex: median of $\{4, 9, 3, 17\}$ is $\tilde{x} = (4 + 9)/2 = 6.5$.

The median is a more "representative" typical value than the mean when the data are *skewed*.

- Quartiles $\{Q_1, Q_2, Q_3\}$: divide the data into quarters (fourths). The median is the 2nd quartile: $Q_2 = \tilde{x}$. To find Q_1 (1st quartile/fourth) and Q_3 (3rd quartile/fourth):
  - Order the data: $x_{(1)} \leq \cdots \leq x_{(n)}$.
  - Separate lower half of data from upper half (include $\tilde{x}$ in both halves if n is odd).
  - Then, Q_1 is the median of the lower half, and Q_3 is the median of the upper half.

▷ Percentiles and Quantiles. The p-th *quantile* $(0 \le p \le 1)$ is any value such that a proportion p of the data is below it (and $1 - p$ above it). A *percentile* is the same, but expressed as a percentage.

So the p-th quantile is the $100p$-th percentile. If you score at the 90th percentile (0.9 quantile) on SAT, and your score is 670, then 90% of all scores are below 670.

Calculation of Quantiles [*]

- Trimmed Mean. Note that $\bar{x}$ is strongly influenced by outliers (corresponds to 0% trimming) and $\tilde{x}$ is robust to outliers but ignores the data outside of the middle (corresponds to almost 100% trimming).

  A compromise is to use a trimmed mean, where a fraction of the data at both high and low ends is "trimmed" (dropped), usually 5-25%, and the rest is averaged.

Ex: Copper Content of Bidri (Ex 1.14 7th Ed.) [*]

$n = 26$ obs on percentage copper content of Bidri wares:

$$2.0,\ 2.4,\ 2.5,\ 2.6,\ 2.6,\ 2.7,\ 2.7,\ 2.8,\ 3.0,\ 3.1,\ 3.2,\ 3.3,\ 3.3$$

$$3.4,\ 3.4,\ 3.6,\ 3.6\ 3.6,\ 3.6,\ 3.7,\ 4.4,\ 4.6,\ 4.7,\ 4.8,\ 5.3,\ 10.1$$

Numerical summaries of categorical data

Use sample proportions:

$$\hat{p} = \frac{x}{n} = \frac{\#\text{ successes}}{n}$$

Ex (Health Complaints):

Tabulate sample proportions of each type of complaint.

	B	C	F	J	M	N	O
x	7	3	9	10	4	6	21
$\hat{p}$	$\frac{7}{60}$	$\frac{3}{60}$	$\frac{9}{60}$	$\frac{10}{60}$	$\frac{4}{60}$	$\frac{6}{60}$	$\frac{21}{60}$

§1.4 Measures of Variability/Spread/Dispersion

- Variance: the sample variance is defined as

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Ignoring the division by $n-1$ instead of n (which is a correction factor to make it an *unbiased estimator* of the *population variance* σ^2, to be discussed later), s^2 is essentially the average squared distance from the mean.

- Standard Deviation: the sample standard deviation is simply the square root of the sample variance:

$$s = \sqrt{s^2}$$

Note: s is a measure of variability on the same scale as the data (unlike s^2 which is in squared units).

- The following computational formula for s^2 is easy to derive, and makes its calculation easier by being based on the sum and sum of squares:

$$s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right]$$

Properties of Variance and Standard Deviation

These measures of variability are not affected by *shifting* the data by a constant, but they are affected by *scaling*. Let s_x^2 and s_y^2 be the sample variances based on samples $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$, respectively. Then, for any constant c, we have the following:

- If $y_1 = x_1 + c, \ldots, y_n = x_n + c$, then $s_y^2 = s_x^2$ and $s_y = s_x$.
- If $y_1 = cx_1, \ldots, y_n = cx_n$, then $s_y^2 = c^2 s_x^2$ and $s_y = |c| s_x$.

Boxplots (Box-and-Whisker plots)

A simple 5-number summary of a dataset that shows location, scale, outliers, and symmetry:

- minimum: $x_{(1)}$,
- 1st quartile/lower fourth: Q_1,
- median: Q_2,
- 3rd quartile/upper fourth: Q_3,
- maximum: $x_{(n)}$.

The box spans the middle half of the data, from Q_1 to Q_3, with either a dot or a line at Q_2. The whiskers extend from the box to the min and max values that are **not** *outliers*. (The outliers are individually marked with dots.)

Def: (Interquartile Range/Fourth Spread)

$$f_s = Q_3 - Q_1$$

This is a measure of spread (variability or dispersion) over the middle 50% of the data.

Def: (Outlier) Any obs that is farther than $1.5f_s$ from the closest quartile (Q_1 or Q_3) is an outlier (or mild outlier).

An extreme outlier is any obs that is more than $3f_s$ from the closest quartile.

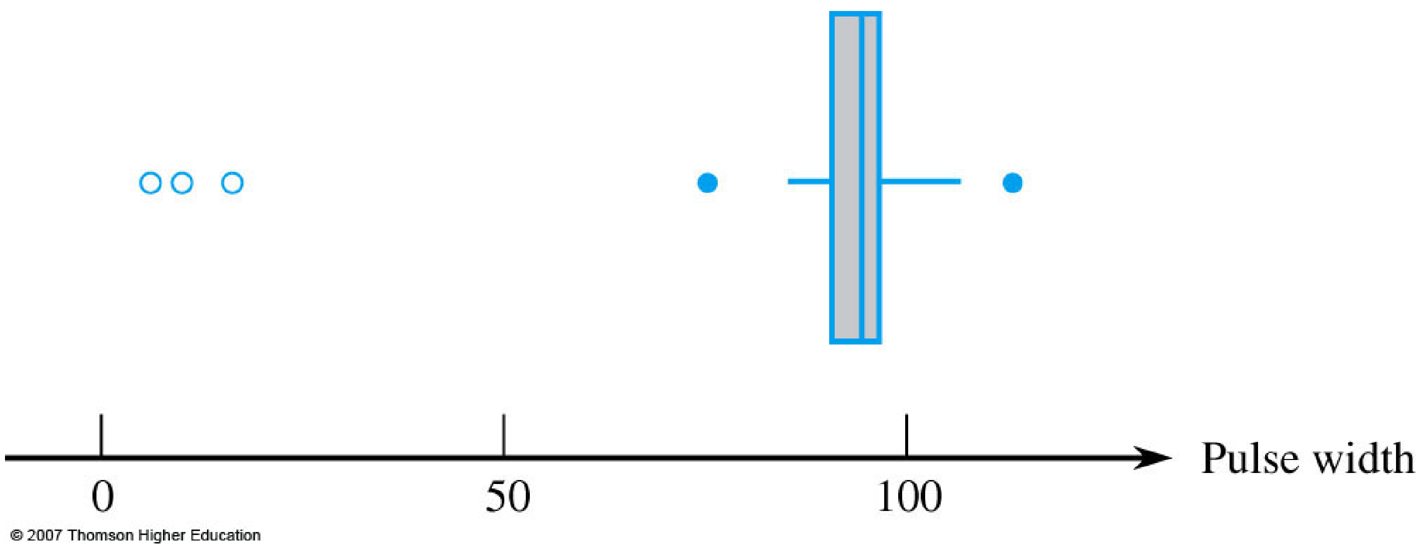
Ex: Corrosion Data (Ex 1.17 7th Ed.)[*]

$n = 19$ obs on amount of corrosion (mg) in iron parts:

40, 52, 55, 60, 70, 75, 85, 85, 90, 90, 92, 94, 94, 95, 98, 100 115, 125, 125

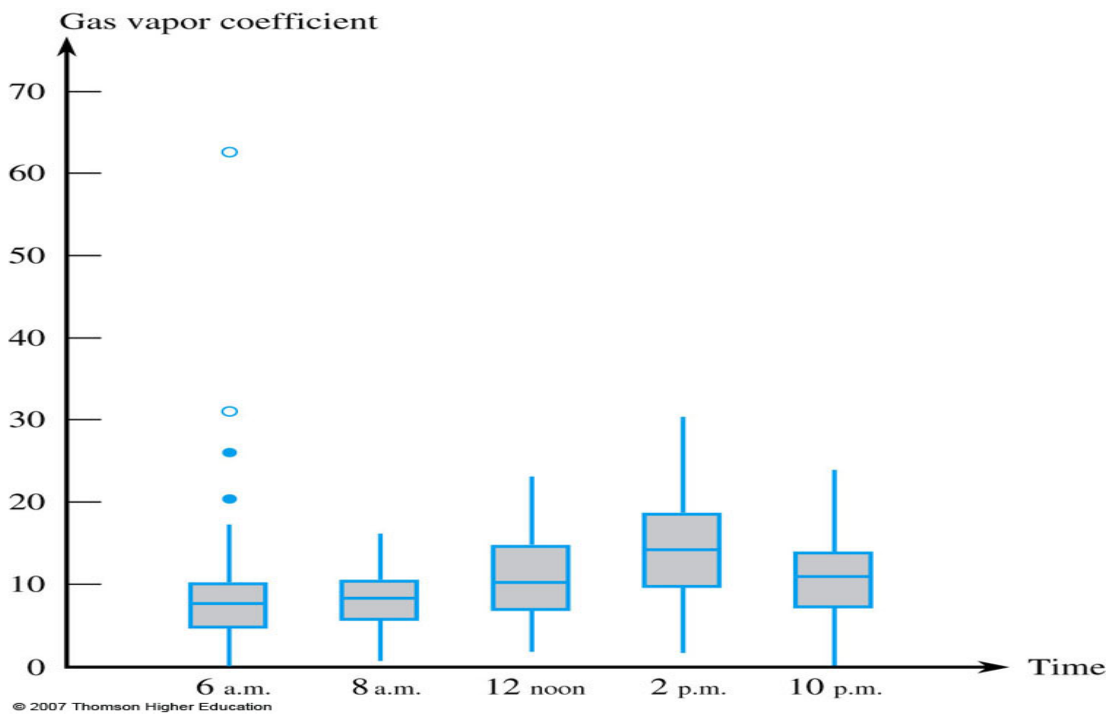
Ex: Pulse Widths Data (Ex 1.18 7th Ed.)

An example of a dataset with outliers (filled dots) and extreme outliers (open dots).



Ex: Gas Vapor Coefficient Data (Ex ??? 7th Ed.)

For two or more groups of obs of a numeric variable, *comparative* (side-by-side) boxplots allow for easy comparison of location, scale, outliers, and symmetry.



Ch. 2: Probability

§2.1 Sample Spaces and Events

§2.2 Axioms, Interpretations, and Properties of Probability

§2.3 Counting Techniques

§2.4 Conditional Probability

§2.5 Independence

Count: slides 36–64 (29 slides).

([*] denotes content that will be given in class)

Roadmap: Where is Course Headed? [*]

§2.1 Sample Spaces and Events

- Experiment: an activity or process whose outcome(s) is subject to uncertainty (random outcomes). Forms the basis for all the ensuing definitions and subsequent probability calculations.

- Sample Space ($\mathcal{S}$): the set of all possible outcomes of the experiment.

- Event: any subset of outcomes, A, of the experiment ($A \subset \mathcal{S}$). A *simple* event consists of exactly one outcome, whereas a *compound* event consists of more than one. An event occurs if the outcome of the experiment falls in the subset of outcomes that defines the event.

Examples [*]

- **Toss a coin.** $\mathcal{S} = \{H, T\}$. The event "head" is $\{H\}$.
- **Toss 2 coins.** $\mathcal{S} = \{HH, HT, TH, TT\}$. The event "exactly 1 head" is a compound event with the two outcomes $\{HT, TH\}$.
- **Battery Failures.** Experiment: test each battery as it comes off an (infinite) assembly line until we observe the 1st success. (Each battery tests as either S=success or F=failure.)
  - Sample space: $\mathcal{S} =$
  - The event "observe an S among the first 3 batteries": $A =$

Events and Operations[*]

- Empty Event: event that never occurs; empty set ϕ.
- Union: $A \cup B$ occurs if either A or B occur, or both.
- Intersection: $A \cap B$ occurs if both A and B occur.
- Disjoint Events: A and B are disjoint if $A \cap B = \phi$.
- Complement: A' occurs if A does not occur; $A' = \mathcal{S} \setminus A$.
- Venn Diagram: pictorial representation of events/sets:
-

Examples [*]

- **Toss a die.** $A = \{\text{number larger than } 3\} = \{4, 5, 6\}$, and $B = \{\text{even number}\} = \{2, 4, 6\}$.

$$A \cup B = \{ \qquad\qquad\qquad \}, \qquad A \cap B = \{ \qquad \}$$

- **Battery Failures.** Let $A = \{S, FS, FFS\}$ and $B = \{S, FFS, FFFFS\}$. Compute the outcomes in:
  - $A \cap B =$
  - $C =$
  - $C' =$

§2.2 Axioms, Interpretations, Probability Properties

- Def: (Probability) The experiment has unpredictable outcomes in any one trial, but has a predictable long-run behavior; relative frequencies of outcomes approach fixed values (over many repetitions of the experiment), and are called its *probabilities*. For a given event A, we denote its probability by $P(A)$.

- **Axioms of Probability**
  - Axiom 1: For any event A, $P(A) \geq 0$.
  - Axiom 2: $P(\mathcal{S}) = 1$.
  - Axiom 3: If $A_1, A_2, \ldots$, is a collection of *mutually exclusive* events ($A_i \cap A_j = \phi$ for $i \neq j$), then

$$P(A_1 \cup A_2 \cup \ldots) = \sum_i P(A_i)$$

Simple Properties of Probability [*]

- Property 1: $P(A) = 1 - P(A')$.
- Property 2: (Cf. with Axiom 3) For any events A and B,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Ex: In a certain suburb, 60% of households subscribe to the Metro paper, 80% subscribe to the Local paper, and 50% subscribe to both papers. If a household is selected at random, what is the probability that it subscribes to:

- at least one of the two papers;
- exactly one of the two papers.

Computing the Probability of Event A

- Define the experiment and list all the outcomes:

$$\mathcal{S} = \{s_1, \ldots, s_n\}$$

- Assign reasonable probabilities to the outcomes:

$$p_i = P(s_i), \qquad i = 1, \ldots, n$$

- Define A as a collection of outcomes: $A = \bigcup_{j=1}^{k} s_j$
- $P(A)$ is the sum of all probabilities for outcomes in A:

$$P(A) = \sum_{j=1}^{k} P(s_j) = \sum_{j=1}^{k} p_j$$

Simplification Under Equally Likely Outcomes

If all outcomes are equally likely ($p_i = p$ for all i), then:

$$P(A) = \frac{\#A}{\#\mathcal{S}} = \frac{\text{number of elements in } A}{\text{number of elements in } \mathcal{S}}$$

<u>Ex:</u> Roll a die. The probability of an odd number is:

$$P(A) = \frac{\#A}{\#\mathcal{S}} = \frac{\#\{1,3,5\}}{\#\{1,2,3,4,5,6\}} = \frac{3}{6} = \frac{1}{2}$$

<u>Ex:</u> Toss coin twice. What is probability of at least one H?

- $\mathcal{S} = \{HH, HT, TH, TT\}$: then $A = \{HT, TH, HH\}$, and since all outcomes are equally likely, $P(A) = 3/4$.
- $\mathcal{S} = \#$ H's $= \{0, 1, 2\}$: then $A = \{1, 2\}$, but $P(A) \neq \frac{2}{3}$ since outcomes NOT equally likely. Using method in previous slide:

$$P(A) = P(1) + P(2) = \frac{2}{4} + \frac{1}{4} = \frac{3}{4}$$

Ex: Roll a die twice. $A = \{$**sum of rolls** $\geq 4\}$ [*]

- **Sol. 1:** Outcomes in $\mathcal{S}$ are equally likely.

$$\mathcal{S} = \{\text{all tuples}\} = \{(1,1),(1,2),(2,1),\ldots,(6,6)\}$$

▹ **Sol. 2:** Outcomes in $\mathcal{S}$ are NOT equally likely [*] .

$$\mathcal{S} = \{\text{all sums}\} = \{2, 3, \ldots, 12\}$$

§2.3 Counting Techniques

Since usage of the equally likely outcomes method involves counting set elements, we need to learn how to "count". Our tools will be: *product rule*, *permutations*, and *combinations*.

Product Rule. Consider an ordered pair of objects (x, y), where there are n_1 choices for x and n_2 choices for y. Then the total number of distinct ordered pairs of objects that can be formed is $n_1 n_2$.

- ▷ Ex 1: Playing cards have 13 faces and 4 suits. There are thus $4 \times 13 = 52$ face-suit combinations.
- ▷ Ex 2: An 8-bit binary word is a sequence of 8 digits, each of which is either a 0 or a 1. There are thus $2^8 = 256$ binary words.

Permutations and Combinations

- **Permutation.** Any ordered sequence of k distinct objects taken from a set of $n \geq k$ distinct objects is called a *permutation* of size k. The number of such sequences is denoted by $P_{k,n}$, and can be counted by the Product Rule:

$$P_{k,n} = n(n-1)(n-2)\cdots(n-k+1) = \frac{n!}{(n-k)!}$$

- **Combination.** Any unordered sequence of k distinct objects taken from a set of $n \geq k$ distinct objects is called a *combination* of size k. The number of such sequences is denoted by $C_{k,n}$ or $\binom{n}{k}$, is read "n choose k", and can be counted by starting from the corresponding permutation and striking out all $k!$ rearrangements of the k objects:

$$C_{k,n} = \binom{n}{k} = \frac{P_{k,n}}{k!} = \frac{n!}{(n-k)!k!}$$

Examples [*]

- **Ex 1.** Select leadership roles of {P,VP,T} for a company from 10 available people.
  - In how many ways can this be done?
  - What is the probability Alex & Bob are both chosen?
- **Ex 2.** Select 3 people from 10 available to fill leadership roles in a company.
  - In how many ways can this be done?
  - What is the probability Alex & Bob are both chosen?

Ex: The Birthday Paradox [*]

One instructor and $n = 72$ students.

- What is the probability no student has same b-day as instructor (event A)?

- What is the probability all students have different b-days (event B')?

Ex: State Lottery (Lotto) [*]

Lotto officials pick 6 numbers randomly between 1 and 53. What is the probability that among your selection of 6 numbers, you match exactly k of the official numbers?

§2.4 Conditional Probability [*]

Ex: Have a group of 50 people. Of these, 26 are male, 18 of which favor proposal A. The remaining 24 are female, 12 of which favor proposal A. Randomly select one person.

- What is the probability the selected person favors proposal A?
- What is the probability the selected person favors proposal A, given that they are female?

Definition of Conditional Probability [*]

The probability of event A happening given the additional information that event B has happened, is called a *conditional probability* (cf. 2nd question of previous example), and is denoted by $P(A|B)$. This involves a reduction in the sample space which becomes just B, and can thus be obtained from the following definition.

Def: If $P(B) > 0$, then $P(A|B)$ is defined as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

and can be visualized in the following Venn diagram:

Examples [*]

- The probability that a project will be well-planned (W_P) is 0.8, and the probability that it will be well-planned and well-executed (W_E) is 0.72. What is the probability that a W_P project will be W_E?

- Roll a die twice. Given that the first number is 1, what is the probability that the total is 3?

Multiplication Rule [*]

Solving for the joint probability $P(A \cap B)$ in the definition of conditional probability gives:

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$$

Use this rule when it's easy to compute conditionals, like $P(A|B)$ or $P(B|A)$, and we have information about (or can easily compute) the corresponding marginals, $P(B)$ or $P(A)$.

Ex: Have 3 red and 2 blue balls in a box. Randomly draw 2 balls. What is the probability both are red?

Example (3 red and 2 blue balls): Continued [*]

Other ways to solve for the probability both balls are red.

- **Tree Diagram.**

- **Permutations.**

- **Combinations.**

Ex: Three Students & One Super Bowl Ticket [*]

Ticket is hidden in one of 3 boxes and they will decide who gets it as follows. One of them chooses a box, if the ticket is inside he gets it, otherwise he's out (along with the box). Then the next student picks one of the 2 remaining boxes, etc. Does it matter who goes first?

Law of Total Probability & Bayes Theorem

Let $A_1, \ldots, A_k$ be mutually exclusive events $(A_i \cap A_j = \phi)$ which are *exhaustive* $(\bigcup_{i=1}^k A_i = \mathcal{S})$. Then for any event B we have the following two important results.

- **Law of Total Probability:**

$$P(B) = \sum_{i=1}^{k} P(B \cap A_i) = \sum_{i=1}^{k} P(B|A_i)P(A_i)$$

- **Bayes Theorem:**

$$P(A_j|B) = \frac{P(A_j \cap B)}{P(B)} = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^{k} P(B|A_i)P(A_i)}$$

Ex: Machines Producing Parts [*]

Three machines produce parts with the following defective rates: $M_1 = 10\%$, $M_2 = 20\%$, $M_3 = 25\%$. The percentages of all parts produced by the machines are: $M_1 = 60\%$, $M_2 = 20\%$, $M_3 = 20\%$.

- ▷ What is the probability that a randomly chosen part is defective?
- ▷ What is the probability that a defective part is made by M_1?

Ex: Lie Detector Test [*]

Let $+/-$ denote event that lie-detector reads positive/negative. Let T/L denote the event that the subject is telling truth/lying. The reliability of the detector is: $P(+|L) = 0.88$ and $P(-|T) = 0.86$. Also, suppose most people don't lie so that $P(T) = 0.99$. If a subject has a positive reading, what is the probability that he is in fact telling the truth?

§2.5 Independence

Def: Events A and B are <u>independent</u> if

$$P(A \cap B) = P(A)P(B)$$

or equivalently

$$P(A|B) = P(A) \qquad \text{and} \qquad P(B|A) = P(B)$$

Def: Events $\{A, B, C\}$ are (mutually) <u>independent</u> if

$$P(A \cap B) = P(A)P(B), \quad P(A \cap C) = P(A)P(C), \quad P(B \cap C) = P(B)P(C),$$

and

$$P(A \cap B \cap C) = P(A)P(B)P(C)$$

(Generalizes in an obvious way to independence of $k \geq 3$ events.)

Ex: Toss Two Fair Coins [*]

$\mathcal{S} = \{HH, HT, TH, TT\}$, and we define the events:

$A = \{\text{1st coin H}\}, \quad B = \{\text{2nd coin H}\}, \quad C = \{\text{1 or 2 coins show H}\}$

- Are A and B independent? Mutually exclusive?

- Are A and C independent? Mutually exclusive?

Ex: Binary Signal Decoding [*]

A binary message consists of a single digit, either 0 or 1. Because of random noise in the channel, the message could be incorrectly received as the opposite digit with probability p. Which of the following two schemes has the highest probability of resulting in the correct transmission of a 1?

- Send the selected digit once.

- Send the selected digit 3 times in succession; message is decoded via "majority rule".

Ch. 3: Discrete Random Variables

And Probability Distributions

§3.1 Random variables

§3.2 Discrete random variables & distributions

§3.3 Expected values

§3.4 Binomial distribution

§3.5 Hypergeometric & negative binomial distribution

§3.6 Poisson distribution

Count: slides 65–97 (33 slides).

([*] denotes content that will be given in class)

§3.1 Random Variables [*]

Def: A random variable (r.v.) X is any rule that associates a real number with each outcome in the sample space

$$X : \mathcal{S} \longmapsto \mathbb{R}$$

(More abstractly: X is a function or mapping from $\mathcal{S}$ to $\mathbb{R}$.)

Examples

- **Bernoulli Experiment:** has a binary outcome (success or failure), e.g., tossing a coin. We can define e.g., $X = 1$ if success, and $X = 0$ if failure.
- **Count Experiments:** frequently the outcome of an experiment is the number of times that a particular event happens, e.g., X is the number of traffic accidents on a given road over a year.
- **Measurement Experiments:** many of the examples already presented (temperatures at shuttle launch, material strength, golf course lengths, power consumption, etc.) are numeric outcomes on a continuous scale, e.g., X is the length of a particular golf course.

Discrete Random Variables

Def: A random variable that can only assume distinct (countable) values is said to be discrete. These can have either a finite or (countably) infinite range.

Examples of discrete r.v.'s

- **Bernoulli:** possible values are 0 or 1, denoting "failure" or "success". E.g., toss a coin.
- **Binomial:** possible values are $0, 1, 2, \ldots, n$, which counts the number of "successes" in n trials. E.g., number of heads in n tosses of a coin.
- **Geometric:** possible values are $0, 1, 2, \ldots$, which counts the number of trials until the first "success". E.g., number of tosses until the first head.

Continuous Random Variables

Def: A random variable that can (theoretically) assume any value in an interval (finite or infinte) is said to be continuous.

Examples of continuous r.v.'s

- **Measurements.** Select random location in contiguous U.S. and measure height above sea level. Could be any value in the (approximate) range of $[-300, 14500]$ feet.
- **Time to Failure.** Randomly select a new light bulb, switch it on, and record the time until it burns out.
- **Uniform Distribution.** This is the most basic of all r.v.'s, where the possible values occur with equal probability in some finite range $[a, b]$.

§3.2 Discrete Random Variables & Distributions

A discrete distribution is described by giving its **probability mass function (pmf)**, $p(x)$, either as a table, function, or plot. The meaning of the pmf is that

$$p(x) = P(X = x)$$

and is in theory defined for all x, but in practice $p(x) = 0$ except at select distinct values.

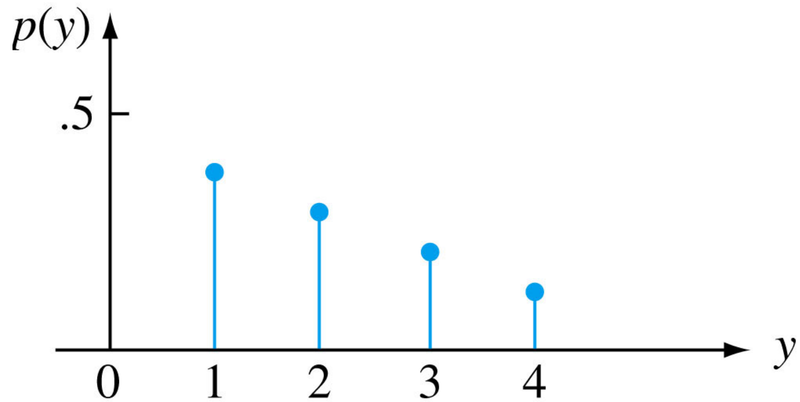
Properties of the pmf:

- $p(x) \geq 0$, for all $-\infty < x < \infty$.
- $\sum_{x=-\infty}^{\infty} p(x) = 1$.

Example: Table, Function, and Plot of a PMF

x	1	2	3	4
$p(x)$	0.4	0.3	0.2	0.1

$$p(x) = \begin{cases} \frac{5-x}{10}, & x = 1, 2, 3, 4 \\ 0, & \text{otherwise} \end{cases}$$



Parameters of a Distribution

The distribution of many familiar r.v.'s will often depend on variable quantities called *parameters*. Two examples are:

- **Bernoulli.** Single trial, results in "success" ($X = 1$) with probability p, and "failure" ($X = 0$) with probability $1 - p$. The pmf is:

$$p(x) = \begin{cases} 1-p, & x = 0 \\ p, & x = 1 \\ 0, & \text{otherwise} \end{cases}$$

- **Geometric.** Repeated (and independent) Bernoulli trials, X counts number of trials until first success. The pmf is:

$$p(x) = \begin{cases} (1-p)^{x-1}p, & x = 1, 2, \ldots \\ 0, & \text{otherwise} \end{cases}$$

The Cumulative Distribution Function (CDF)

- **Def.** The **cdf** accumulates (adds up) pmf values left of x:

$$F(x) = P(X \leq x) = \sum_{y=-\infty}^{x} p(y)$$

- **Limits.** If X has a finite range, $[a, b]$, then $F(x) = 0$ for $x < a$ and $F(x) = 1$ for $x > b$. In all cases we have:

$$\lim_{x \downarrow -\infty} F(x) = 0, \qquad \text{and} \qquad \lim_{x \uparrow \infty} F(x) = 1$$

- **Graph.** For discrete r.v.'s with pmf values on the integers, the graph of $f(x)$ is a step function, with jumps occuring at the mass points.

▹ **Usage of the cdf.** Mainly to calculate probabilities. If $a < b$ are any two numbers, and $F(x^-)$ denotes $F(\cdot)$ evaluated immediately to the left of x, we have:

- $P(a < X \leq b) = F(b) - F(a)$
- $P(a \leq X \leq b) = F(b) - F(a^-)$
- $P(a \leq X < b) = F(b^-) - F(a^-)$
- $P(a < X < b) = F(b^-) - F(a)$

For a discrete r.v. X with pmf $p(x)$ defined on the integers, $F(x)$ only changes at integer values, so that $F(a^-) = F(a-1)$, and thus:

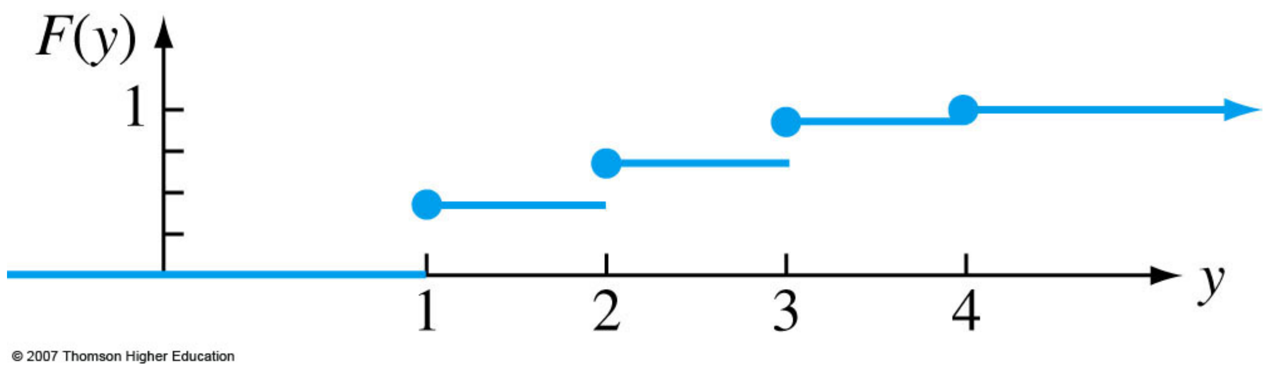
$$p(a) = P(X = a) = F(a) - F(a-1)$$

Example (Slide 71): Plot of CDF

Recall pmf of Y in tabular form:

y	1	2	3	4
$p(y)$	0.4	0.3	0.2	0.1

For discrete r.v.'s, the cdf is always a step function with jumps (mass points) at the integers:



Ex: Geometric Random Variable CDF [*]

Let X denote the number of births a couple has until the first boy. Assume the couple continues to have children indefinitely until that happens, and that the probability of a boy in any one birth is p. This is a Geometric r.v. with pmf:

$$p(x) = \begin{cases} (1-p)^{x-1}p, & x = 1, 2, \ldots \\ 0, & \text{otherwise} \end{cases}$$

The cdf of X is given by:

§3.3 Expected Values

Def: The **expected value** of discrete r.v. X with pmf $p(x)$ and values in the set $\mathcal{X}$, is the sum of its values weighted by the corresponding probabilities:

$$\mathbb{E}(X) = \mu_X = \sum_{x \in \mathcal{X}} xp(x)$$

Ex 3.17 (7th Ed.): The Apgar scores of newborns are integers in the range 0 to 10, and can be modeled as r.v. X with pmf

x	0	1	2	3	4	5	6	7	8	9	10
$p(x)$	0.002	0.001	0.002	0.005	0.02	0.04	0.18	0.37	0.25	0.12	0.01

$$\mathbb{E}(X) = 0(0.002) + 1(0.001) + \cdots + 10(0.01) = 7.15$$

Ex: Geometric Random Variable Expectation [*]

Let X be the Geometric r.v. with pmf (cf. slide 76):

$$p(x) = \begin{cases} (1-p)^{x-1}p, & x = 1, 2, \ldots \\ 0, & \text{otherwise} \end{cases}$$

We will now show that $\mathbb{E}(X) = 1/p$:

$$\mathbb{E}(X) = \sum_{x=1}^{\infty} x(1-p)^{x-1}p$$

Expected Value of a Function of X

Def: For discrete r.v. X with pmf $p(x)$ and values in the set $\mathcal{X}$, the expectation of the r.v. $h(X)$ is defined as:

$$\mathbb{E}[h(X)] = \sum_{x \in \mathcal{X}} h(x)p(x)$$

Linear Combination: For any constants a and b, it's easy to see that the expectation operator is linear, so that

$$\begin{aligned}
\mathbb{E}(aX + b) &= \sum_{x \in \mathcal{X}} (ax + b)p(x) \\
&= a \sum_{x \in \mathcal{X}} xp(x) + b \sum_{x \in \mathcal{X}} p(x) \\
&= a\,\mathbb{E}(X) + b = a\,\mu_X + b
\end{aligned}$$

The Variance of a Random Variable

Def: For discrete r.v. X with $E(X) = \mu$, pmf $p(x)$, and values in the set $\mathcal{X}$, the **variance** of X is defined as:

$$\mathbb{V}(X) = \sigma_X^2 = \sigma^2 = \mathbb{E}[(X - \mu)^2] = \sum_{x \in \mathcal{X}} (x - \mu)^2 p(x)$$

The **standard deviation** is $\sigma = \sqrt{\mathbb{V}(X)}$.

Shortcut Formula:

$$\begin{aligned} \sigma^2 &= \sum_{x \in \mathcal{X}} (x - \mu)^2 p(x) = \sum_{x \in \mathcal{X}} x^2 p(x) - 2\mu \sum_{x \in \mathcal{X}} x p(x) + \mu^2 \sum_{x \in \mathcal{X}} p(x) \\ &= \mathbb{E}(X^2) - \mu^2 \end{aligned}$$

Linear Combination: For any constants a and b:

$$\mathbb{V}(aX + b) = \sigma_{aX+b}^2 = a^2 \sigma_X^2, \qquad \text{and} \qquad \sigma_{aX+b} = |a| \sigma_X$$

§3.4 Binomial Distribution

Def: A Binomial experiment satisfies the following conditions:

(i) a fixed number of n trials are performed;

(ii) the trials are identical, and each results in a binary outome, either success (S) with probability p or failure (F) with probability $1 - p$;

(iii) the trials are idependent.

The binomial r.v. X counts the number out of the n trials that result in S (success). We write $X \sim \text{bin}(n, p)$, and its pmf is:

$$b(x; n, p) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0, 1, 2, \ldots, n$$

Mean, Variance, and CDF of the Binomial

For $X \sim \text{bin}(n, p)$, we have the following facts.

- $\mathbb{E}(X) = np$ and $\mathbb{V}(X) = np(1-p)$;
- the cdf is given by

$$F(x) = P(X \leq x) = \sum_{y=0}^{x} \binom{n}{y} p^y (1-p)^{n-y}$$

There is no closed-form expression for $F(x)$, so its values have to be obtained case-by-case. Table A.1 in the appendix of the book gives its values to 3 decimal places for a few select values of n and p.

Ex: 3.32 (7th Ed.) [*]

X is # of books out 15 that fail binding test; approx. 20% of books independently fail the test; implies that:

$$X \sim \text{bin}(n = 15, p = 0.2)$$

(a) What is the probability at most 8 fail?

(b) What is the probability exactly 8 fail?

(c) What is the probability at least 8 fail?

(d) What is the probability between 4 and 7 fail?

(e) Find the mean and variance of X.

Ex: 3.32 (7th Ed.) Continued. . . [*]

(f) Plot the pmf and cdf of $X \sim \text{bin}(15, 0.2)$

§3.5 Hypergeometric & Negative Binomial R.V.'s

- **Binomial r.v.:** requires independent trials with constant success probability p.
- **Sampling with replacement:** if we sample n objects from a population of S/F objects (total size N) by first replacing each selected object before sampling the next one, then p remains unchanged from sampling to sampling (trial to trial), and the trials are independent.
- **Sampling without replacement:** if we don't replace the objects after each trial, then p is changed for subsequent trials, which are therefore no longer independent. However: if $n/N << 1$, we can proceed as if sampling were with replacement... (an approx. binomial experiment).

The Hypergeometric R.V.

Def: In a Hypergeometric experiment, there are a total of N objects, M of which are of type S and $N - M$ of type F, and we randomly sample n objects without replacement.

The hypergeometric r.v. X counts the number of objects of type S out of the n trials. Write $X \sim \text{hypergeom}(n, M, N)$. Its pmf is:

$$h(x; n, M, N) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}, \qquad \max(0, n - N + M) \leq x \leq \min(n, M)$$

The mean and variance of X are:

$$\begin{aligned} \mathbb{E}(X) &= n\left(\frac{M}{N}\right) \\ \mathbb{V}(X) &= \left(\frac{N-n}{N-1}\right) n \left(\frac{M}{N}\right)\left(1 - \frac{M}{N}\right) \end{aligned}$$

Exs: 3.35 and 3.36 (7th Ed.) [*]

Ex 3.35: Have 12 inkjet and 8 laser printers; randomly sample 5; let X denote number of inkjets in sample. Compute the pmf of X:

Ex 3.36: In a pop. of 25 animals, 5 have been previously tagged. If 10 animals are randomly selected for inspection, compute the probability that:

- exactly 2 of the sampled animals were previously tagged;
- at most 2 of the sampled animals were previously tagged;

Binomial Approx. to Hypergeometric [*]

Recall (slide 85): if $n/N << 1$, we can proceed as if sampling were with replacement..., (an approx. binomial experiment):

$$\text{hypergeom}(n, M, N) \approx \text{bin}\,(n, p = M/N)$$

Ex: Have $500,000$ drivers in a state, $400,000$ of which are insured. Randomly sample 10 drivers, and let X be the # in the sample that are insured.

The Negative Binomial R.V.

Def: This is an experiment of repeated independent and identical trials where each trial produces "success" (S) with probability p or "failure" (F) with probability $1-p$, just like the binomial.

Let X be the number of failures until the r-th success. The distribution of X is called a *negative binomial*, and we write $X \sim \text{negbin}(r,p)$. Its pmf is:

$$nb(x;r,p) = \binom{x+r-1}{r-1} p^r (1-p)^x, \qquad x = 0, 1, 2, \ldots$$

The mean and variance of X are:

$$\mathbb{E}(X) = \frac{r(1-p)}{p}, \qquad \mathbb{V}(X) = \frac{r(1-p)}{p^2}$$

Special Case: The Geometric R.V.

Def: The special case of the negative binomial with $r = 1$ is called the geometric, and we write $X \sim \text{Geom}(p)$. I.e., letting X be the number of failures that precede the first S, the pmf is:

$$geom(x; p) = p(1-p)^x, \qquad x = 0, 1, 2, \ldots$$

with mean and variance, $\mathbb{E}(X) = (1-p)/p$ and $\mathbb{V}(X) = (1-p)/p^2$.

Note: A common reparametrization of the geometric is to let Y be the number of trials until the first S. Its pmf is:

$$geom(y; p) = p(1-p)^{y-1}, \qquad y = 1, 2, \ldots$$

with mean and variance, $\mathbb{E}(Y) = 1/p$ and $\mathbb{V}(Y) = (1-p)/p^2$.
(In fact: $Y = X + 1$.)

Ex 3.38 (7th Ed.) [*]

Want to recruit 5 couples for a study. Let $p = 0.2$ be the probability that a randomly selected couple agrees to participate.

(a) How many couples do we need to ask until 5 agree?

(a) What is the probability exactly 15 couples must be asked until 5 agree?

Ex: Winning at Table Tennis [*]

A plays B, A wins any one point with probability p, and the first player to get 21 points wins the game (ignore winning by 2 points). What is the probability A wins the next game?

§3.6 Poisson Distribution

No formal experiment; X counts number of events over space or time; have to explicitly state that X follows a Poisson r.v.; but its pmf arises naturally as the result of a certain limiting operation... (the Poisson Process, to be discussed later).

Def: The pmf of $X \sim \text{Poisson}(\lambda)$ is governed by the rate parameter $\lambda > 0$, and is given by:

$$p(x; \lambda) = \frac{\lambda^x}{x!} e^{-\lambda}, \qquad x = 0, 1, 2, \ldots$$

with mean and variance, $\mathbb{E}(X) = \lambda = \mathbb{V}(X)$.

Note: Table A.2 in appendix of book gives cdf values to 3 decimal places for a few select values of λ (called μ).

Ex 3.39 (7th Ed.) [*]

Number of creatures caught in a trap (X) is Poisson distributed with a mean of 4.5 per day. Find the following.

(a) The prob. that exactly 5 creatures are caught on a given day.

(b) The prob. that at most 5 creatures are caught on a given day.

(c) What is the standard deviation of X?

(d) Show that the pmf of X satisfies the properties of a pmf.

Poisson Approx. to Binomial [*]

Theorem: In the binomial $b(x; n, p)$ pmf, if $n \to \infty$ and $p \downarrow 0$ in such a way that $np \to \lambda < \infty$, then the binomial pmf converges to the Poisson pmf:

$$b(x; n, p) \longrightarrow p(x; \lambda), \qquad \text{with } \lambda = np$$

Ex 3.40 (7th Ed.): Let X be the number of typos in a 400-page book. If the probability of a typo in any one page is 0.005, find the probability that the book contains:

- ▷ exactly 1 typo;
- ▷ at most 3 typos.

The Poisson Process

A Poisson Process is generated by discrete events that occur over a continuum like time or space, subject to the following conditions:

- there exists a parameter $\alpha > 0$ such that, over short time intervals δt, $P(\text{exactly one event over } \delta t) \approx \alpha\delta t$;
- $P(\text{more than one event over } \delta t) \approx 0$;
- the number of events in an interval is independent of the number of events prior to this interval (the process has "no memory" or is "restartable").

The parameter α is called the *rate* of the process, and is the number of events per unit time. The number of events in an interval of length t has a Poisson distribution with mean $\lambda = \alpha t$.

Ex 3.42 (7th Ed.) [*]

Pulses arrive at a rate of 6/minute (according to a Poisson Process). Let X denote the number of pulses in a 30 second interval.

- What is the distribution of X?

- What is the probability that at least one pulse arrives in the next 30 seconds?

Ch. 4: Continuous Random Variables

And Probability Distributions

§4.1 Continuous random variables & probability density functions

§4.2 Cumulative distribution function & expected values

§4.3 Normal distribution

§4.4 Gamma distribution and its relatives

§4.5 Other distributions

§4.6 The normal probability plot

Count: slides 98–131 (34 slides).

([*] denotes content that will be given in class)

§4.1 Continuous Random Variables

Def: A random variable that can (theoretically) assume any value in an interval (finite or infinte) is said to be continuous.

Examples of continuous r.v.'s

- **Measurements.** Select random location in contiguous U.S. and measure height above sea level. Could be any value in the (approximate) range of $[-300, 14500]$ feet.
- **Time to Failure.** Randomly select a new light bulb, switch it on, and record the time until it burns out.
- **Uniform Distribution.** This is the most basic of all r.v.'s, where the possible values occur with equal probability in some finite range $[a, b]$.

The Probability Density Function

A **probability density function (pdf)** of a continuous r.v. X is a function $f(x)$ such that for any $a \leq b$,

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

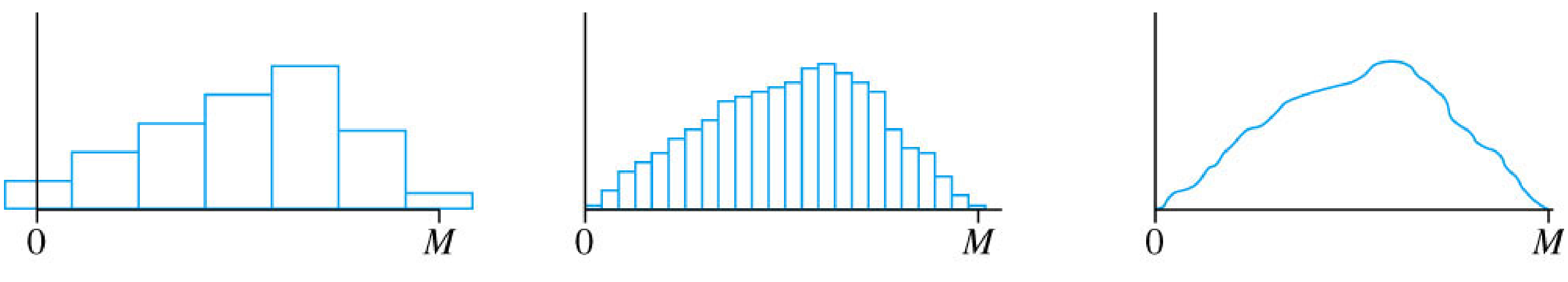
Thus, the probability that X takes on values in the interval $[a, b]$ is the area under $f(x)$. From this we get the (puzzling?) result that for any given number c, $P(X = c) = 0$.

Properties of the pdf:

- $f(x) \geq 0$, for all $-\infty < x < \infty$.
- $\int_{-\infty}^{\infty} f(x)dx = 1$.

The PDF as a Limit of Histograms

The pdf can be viewed as a limit of discrete histograms. Consider measuring the depth (X) of a lake at a set of random locations. We can "discretize" X by measuring to the nearest meter (a), centimeter (b), and so on. As we measure more finely, the resulting sequence of histograms approaches a smooth curve: the pdf (c).



(a) (b) (c)

The Uniform Distribution [*]

A continuous r.v X is said to have a uniform distribution on the interval $[A, B]$ if its pdf is:

$$f(x; A, B) = \begin{cases} \frac{1}{B-A}, & A \leq x \leq B \\ 0, & \text{otherwise} \end{cases}$$

Ex: The wait time X at the bus stop is uniformly distributed on the interval $[0, 5]$ minutes. The probability that one has to wait between 1 and 3 minutes is:

Ex 4.5 (7th Ed.) [*]

Let X be the time headway (time difference in seconds) between 2 randomly chosen consecutive cars on a freeway. The pdf of X can be modeled as follows:

$$f(x) = \begin{cases} 0.15e^{-0.15(x-0.5)}, & x \geq 0.5 \\ 0, & \text{otherwise} \end{cases}$$

- Show that $f(x)$ satisfies the properties of a pdf.

- What is the probability that the time headway is at most 5 seconds?

4.2 The CDF and Expectations [*]

Def: The cumulative distribution function (cdf) of a continuous r.v X, is the function $F(x)$ that, for every $x \in \mathbb{R}$, is defined by:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(y)dy$$

Thus $F(x)$ is the area under the pdf $f(y)$ to the left of x.

Fact: By the fundamental theorem of calculus, we have:

$$f(x) = F'(x) = \frac{dF(x)}{dx}$$

at every $x \in \mathbb{R}$ for which $F'(x)$ exists.

Ex: The cdf of $X \sim \text{Unif}(A, B)$ is (cf. slide 102):

$$F(x) = \begin{cases} 0, & x < A \\ \int_A^x \frac{1}{B-A} dy = \frac{x-A}{B-A}, & A \leq x \leq B \\ 1, & x > B \end{cases}$$

Note that $F'(x) = \frac{1}{B-A}$ for $A \leq x \leq B$, and zero otherwise, as expected!

Usage of cdf (continuous r.v. X): If $a < b$ are any two numbers, then because $P(X = x) = 0$ for any x, we have:

$$\begin{aligned} P(a \leq X \leq b) &= F(b) - F(a) \\ &= P(a < X \leq b) \\ &= P(a \leq X < b) \\ &= P(a < X < b) \end{aligned}$$

Ex 4.7 (7th Ed.) [*]

The cdf of the dynamic load X r.v. on a bridge is modeled as:

$$F(x) = \begin{cases} 0, & x < 0 \\ \frac{x}{8} + \frac{3}{16}x^2, & 0 \le x \le 2 \\ 1, & 2 < x \end{cases}$$

Find the:

- probability that the load is between 1 and 1.5;
- probability that the load exceeds 1;
- the pdf of X.

Percentiles and Quantiles (cf. Slide 23)

Def: For any $0 < p < 1$, the p-th quantile of continuous r.v. X, is the real number $\eta(p)$ that satisfies:

$$p = F(\eta(p)) \qquad \Longrightarrow \qquad \eta(p) = F^{-1}(p)$$

Remarks:

- $\eta(p)$ is a *population* quantile instead of the *sample* quantile defined in Ch. 1.
- $\eta(p)$ is the value on the x-axis s.t. $100p\%$ of the area under the pdf $f(x)$ lies to the left of $\eta(p)$.
- The (population) median $\tilde{\mu} = \eta(0.5)$ is the 0.5-quantile or 50th percentile, and is the value on the x-axis s.t. half the area under the pdf $f(x)$ lies to the left of it.

Expected Values of a Continuous R.V.

Def: The **expected value** of continuous r.v. X with pdf $f(x)$, is the integral of its values weighted by the pdf:

$$\mathbb{E}(X) = \mu_X = \int_{-\infty}^{\infty} x f(x) dx$$

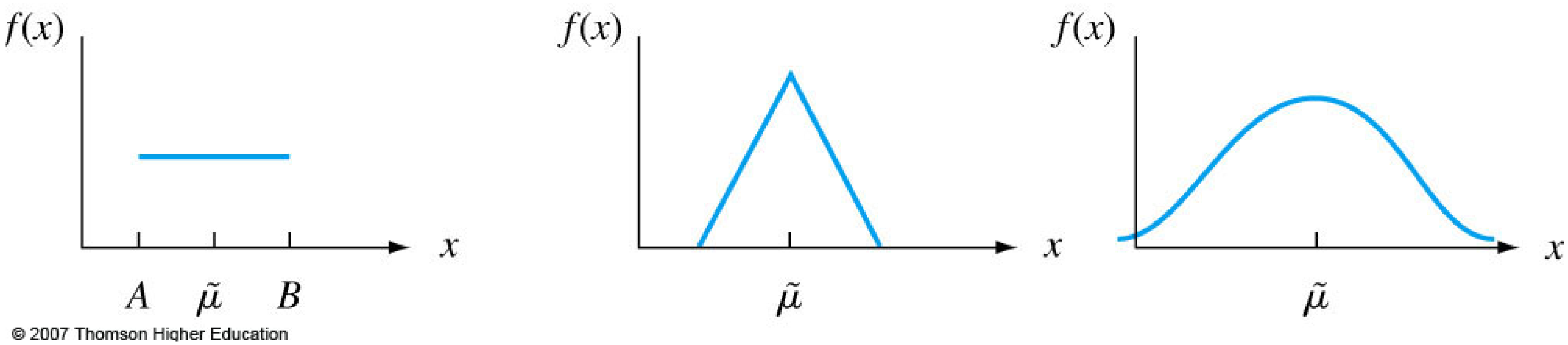
In general, the expected value of some function $h(X)$ is defined as:

$$\mathbb{E}[h(X)] = \mu_X = \int_{-\infty}^{\infty} h(x) f(x) dx$$

Remarks:

- Expectation of a linear function: $\mathbb{E}(aX + b) = a\mu + b$.

▷ For a *symmetric* distribution, which means the pdf to the left of the median $\tilde{\mu}$ is a mirror image of the pdf to the right of $\tilde{\mu}$, the mean and median coincide: $\tilde{\mu} = \mu$.



Ex 4.9 (7th Ed.) [*]

The amount (proportion) of gravel X sold in a week has pdf:

$$f(x) = \frac{3}{2}(1 - x^2), \qquad 0 \leq x \leq 1$$

Find the following:

- the cdf $F(x)$;
- the mean μ;
- the median $\tilde{\mu}$.

Ex 4.11 (7th Ed.) [*]

Species A & B compete for control of a resource. Let X be the proportion controlled by species A, and suppose $X \sim \text{Unif}(0, 1)$. Find the following:

- the function $h(x)$ that defines the larger of the two proportions controlled by each species (majority control);

- the expected proportion of the resource controlled by the species with majority control.

Variance of a Continuous R.V.

Def: For continuous r.v. X with $E(X) = \mu$ and pdf $f(x)$, the **variance** of X is defined as:

$$\mathbb{V}(X) = \sigma_X^2 = \sigma^2 = \mathbb{E}[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx$$

The **standard deviation** is $\sigma = \sqrt{\mathbb{V}(X)}$.

Shortcut Formula:

$$\begin{aligned} \sigma^2 &= \int (x-\mu)^2 f(x) dx = \int x^2 f(x) dx - 2\mu \int x f(x) dx + \mu^2 \int f(x) dx \\ &= \mathbb{E}(X^2) - \mu^2 \end{aligned}$$

Linear Combination: For any constants a and b:

$$\mathbb{V}(aX+b) = \sigma_{aX+b}^2 = a^2 \sigma_X^2, \qquad \text{and} \qquad \sigma_{aX+b} = |a| \sigma_X$$

Ex 4.9 (7th Ed.) Continued. . . [*]

For the amount of gravel X sold in a week: $\mu = 3/8$, and cdf & pdf:

$$F(x) = \begin{cases} 0, & x < 0 \\ \frac{3}{2}\left(x - \frac{x^3}{3}\right), & 0 \le x \le 1 \\ 1, & 1 < x \end{cases}, \quad f(x) = \begin{cases} \frac{3}{2}(1 - x^2), & 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases}$$

Find the following:

- the variance of X;

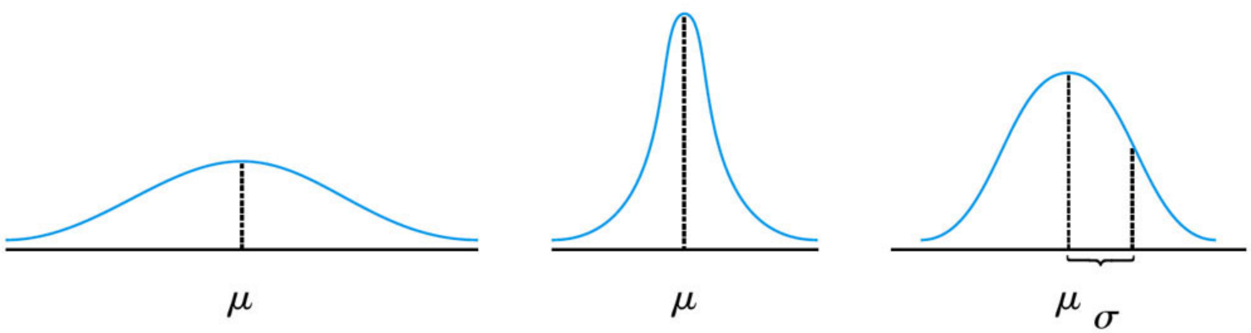
- the variance of $Y = 3X + 1$.

§4.3 The Normal Distribution

- The *Normal* or *Gaussian* distribution is central to statistics (Ch. 5).
- The pdf of $X \sim \mathcal{N}(\mu, \sigma^2)$ is:

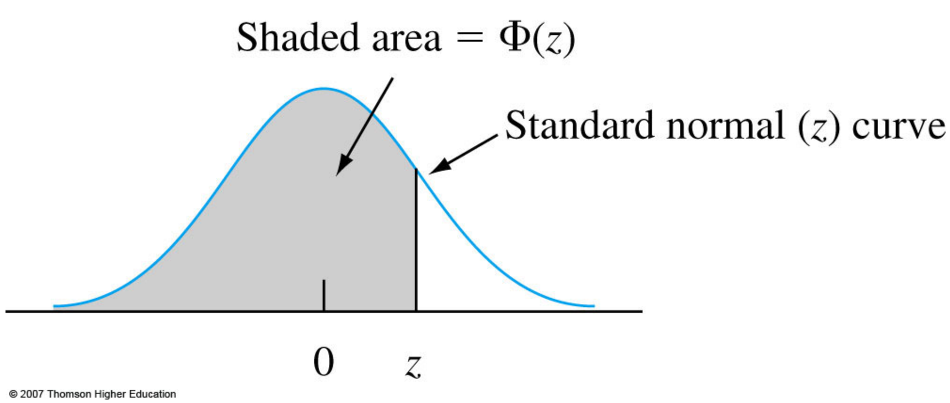
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \qquad -\infty < x < \infty$$

- Obviously: $\mathbb{E}(X) = \mu$, and this is the location of the peak.
- Not so obviously: $\mathbb{V}(X) = \sigma^2$, and this controls the width.

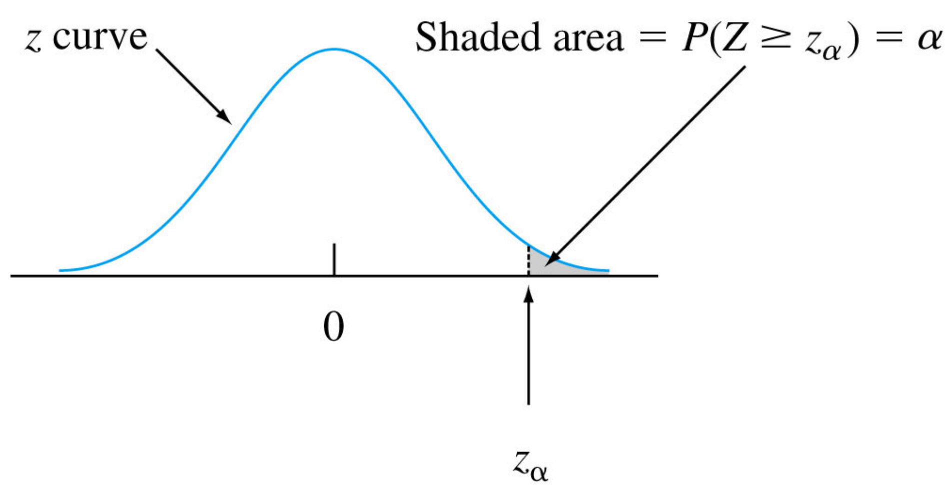


The Standard Normal Distribution

- $Z \sim \mathcal{N}(0,1)$ is called the *standard normal* distribution, and as we will see, all calculations for $X \sim \mathrm{N}(\mu, \sigma^2)$ can be related to an equivalent calculation for Z.
- The cdf of Z is denoted by $\Phi(z) = P(Z \leq z)$, the values for which are given to 4 decimal places in Table A.3.

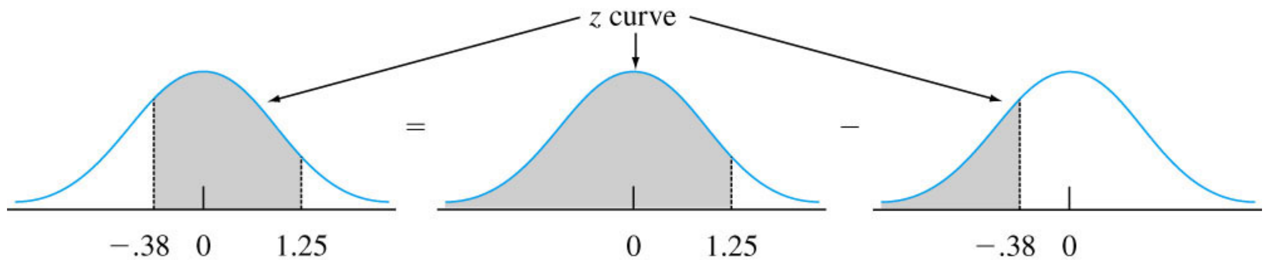


- Quantiles of $Z \sim \mathcal{N}(0, 1)$ can also be obtained from Table A.3, but it involves using the table in reverse, and approximation is needed.
- **Notation:** z_α is the $(1-\alpha)$ quantile, or $100(1-\alpha)$ percentile of Z.



Ex: Standard Normal Probabilities & Quantiles [*]

▷ $P(-0.38 < Z < 1.25) =?$



▷ $z_{0.05} =?$

Table 4.1 Standard Normal Percentiles and Critical Values

Percentile	90	95	97.5	99	99.5	99.9	99.95
α (tail area)	.1	.05	.025	.01	.005	.001	.0005
$z_\alpha = 100(1-\alpha)$th percentile	1.28	1.645	1.96	2.33	2.58	3.08	3.27

Non-Standard Normal Distributions

Facts:

- If $X \sim \mathcal{N}(\mu, \sigma^2)$ then $Z = (X - \mu)/\sigma \sim \mathcal{N}(0,1)$. Thus the computation of X probabilities can be converted to Z probabilities as:

$$\begin{aligned} P(a \le X \le b) &= P\left(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}\right) \\ &= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) \end{aligned}$$

- Since $X = \mu + \sigma Z$, the relationship between the $(1-\alpha)$ quantiles of X and Z is:

$$x_\alpha = \mu + \sigma z_\alpha$$

Ex 4.16 (7th Ed.) [*]

The reaction time X (secs) for in-traffic response to brake lights from a randomly selected driver can be modeled as $X \sim \mathcal{N}(1.25, 0.46^2)$. Find:

- the probability the reaction time (of a randomly selected driver) is between 1 and 1.75 secs;

- the reaction time that is exceeded by only 1% of the drivers.

Normal Approximation to Binomial [*]

Fact: If $X \sim \text{Bin}(n, p)$, then, provided $np \geq 10$ and $n(1-p) \geq 10$, X is approximately normal with $\mu = np$ and $\sigma^2 = np(1-p)$. This means in particular that (applying a *continuity correction*):

$$P(X \leq x) \approx \Phi\left(\frac{x + 0.5 - np}{\sqrt{np(1-p)}}\right), \qquad x = 0, 1, 2, \ldots, n$$

Ex: 25% of drivers are uninsured. Let X be the number of uninsured drivers in a random sample of 50. Compute the exact and approx. probabilities of $5 \leq X \leq 15$.

§4.4 The Gamma Distribution

For $\alpha > 0$, the *gamma function* is defined as:

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$$

Properties:

- for $\alpha > 1$, $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$
- for positive ingteger n, $\Gamma(n) = (n - 1)!$
- $\Gamma(1/2) = \sqrt{\pi}$

Def: A continuous r.v. X is said to have a *gamma* distribution with shape and scale parameters $\alpha > 0$ and $\beta > 0$, $X \sim \text{gamma}(\alpha, \beta)$, if its pdf is:

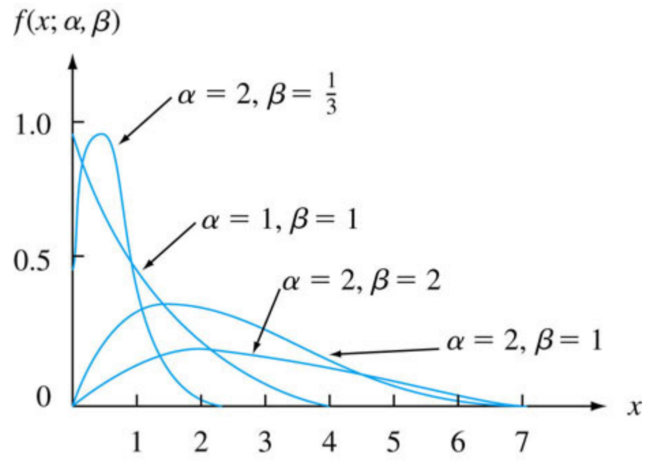
$$f(x; \alpha, \beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, \qquad x \geq 0$$

Facts about $X \sim$ gamma(α, β) [*]

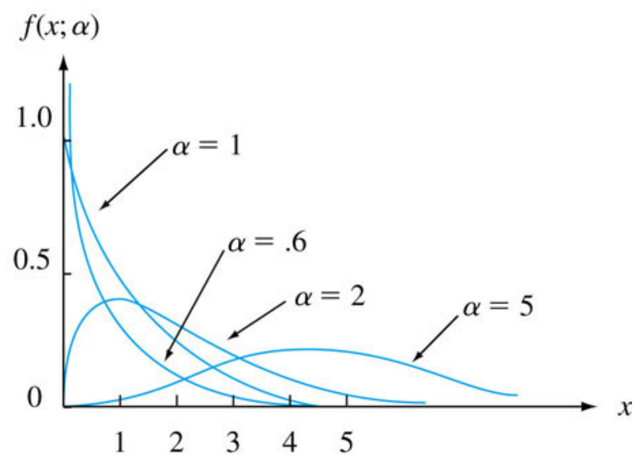
- Show that the pdf integrates to 1.

- Show that $\mathbb{E}(X) = \alpha\beta$ and $\mathbb{V}(X) = \alpha\beta^2$.

Some Gamma PDF Shapes



(a)



(b)

Computing $X \sim$ gamma(α, β) Probabilities [*]

- Compute these via the *standard gamma* distribution, i.e. $Y \sim \text{gamma}(\alpha, \beta = 1)$.

- The cdf of $Y \sim \text{gamma}(\alpha, 1)$ is given by the *incomplete gamma function* (Table A.4):

$$F(y;\alpha) = \int_0^y \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x} dx, \qquad y \geq 0$$

- **Theom:** If $X \sim \text{gamma}(\alpha, \beta)$, then for $x > 0$ the cdf of X is:

$$P(X \leq x) = F(x;\alpha,\beta) = F\left(\frac{x}{\beta};\alpha\right)$$

- **Proof:**

Ex 4.24 (7th Ed.) [*]

Survival time X (weeks) of a randomly selected mouse exposed to radiation can be modeled as: $X \sim \text{gamma}(\alpha = 8, \beta = 15)$.
Calculate:

- the expected survival time;
- the probability the mouse survives between 60 and 120 weeks;
- the probability the mouse survives at least 30 weeks.

The Exponential Distribution

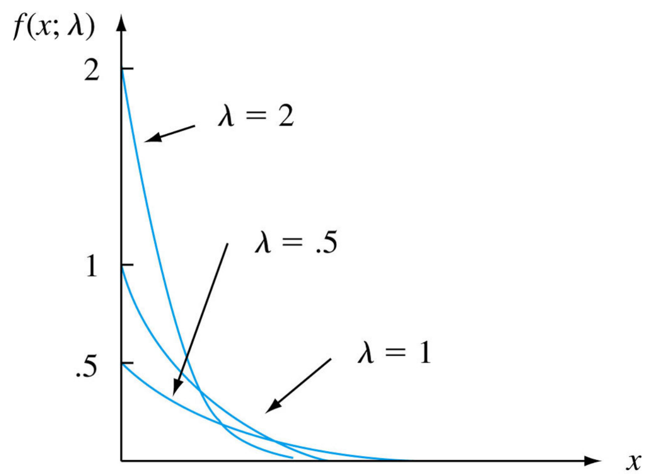
- **Def:** A gamma($\alpha = 1, \beta = 1/\lambda$) is called an *Exponential* distribution with parameter λ: $X \sim \text{Exp}(\lambda)$.
- Its pdf is:

$$f(x; \lambda) = \lambda e^{-\lambda x}, \qquad x \geq 0$$

- Its cdf is:

$$F(x; \lambda) = 1 - e^{-\lambda x}, \qquad x \geq 0$$

- $\mathbb{E}(X) = 1/\lambda$ and $\mathbb{V}(X) = 1/\lambda^2$.



The Exponential and the Poisson Process (cf. slide 96)

- An exponential is often used to model the distribution of times between successive occurrences of an event.

- This is because the exponential is closely related to the Poisson Process ...

- **Theom:** Suppose the number of events occurring in any time interval of length t follows a Poisson Process of rate α. That is:

  (i) the number of events over a time length t is Poisson with mean αt, and

  (ii) the numbers of events in non-overlapping intervals are independent of one another.

  Then, the distribution of the elapsed **time between successive events** is exponential with parameter $\lambda = \alpha$.

Ex 4.22 (7th Ed.) [*]

Calls arrive at the rate of $\alpha = 0.5$ per day according to a Poisson Process. If we let X be the number of days between successive calls, then by the above Theom, $X \sim \text{Exp}(\lambda = 0.5)$. Find:

- the expected time between calls;

- the probability that more than 2 days elapse between calls.

The Memoryless Property of the Exponential

- **Memoryless Property:** If $X \sim \text{Exp}(\lambda)$, then:

$$
\begin{aligned}
P(X \geq t + t_0 \mid X \geq t_0) &= \frac{P(X \geq t + t_0 \text{ and } X \geq t_0)}{P(X \geq t_0)} \\
&= \frac{P(X \geq t + t_0)}{P(X \geq t_0)} \\
&= \frac{e^{-\lambda(t+t_0)}}{e^{-\lambda t_0}} = e^{-\lambda t} \\
&= P(X \geq t)
\end{aligned}
$$

- LHS event says: lives at least t more units given that it has lived t_0 units. RHS event says: lives at least t units. (Thus the system "forgets" it has already lived t_0 units.)

- An amazing property! (Sadly not shared by biological systems...) A physical example is radioactive particle decay.

§4.6 Assessing Normality: The Normal Probability Plot

- If $x_{(1)} < \cdots < x_{(n)}$ are the order statistics (ordered sample), recall the sample quantile calculation from slide 24:

$$x_{(i)} \text{ corresponds approx to } p = \frac{i - 0.5}{n} \text{ quantile}$$

- A plot of the (u_i, v_i) pairs of values:

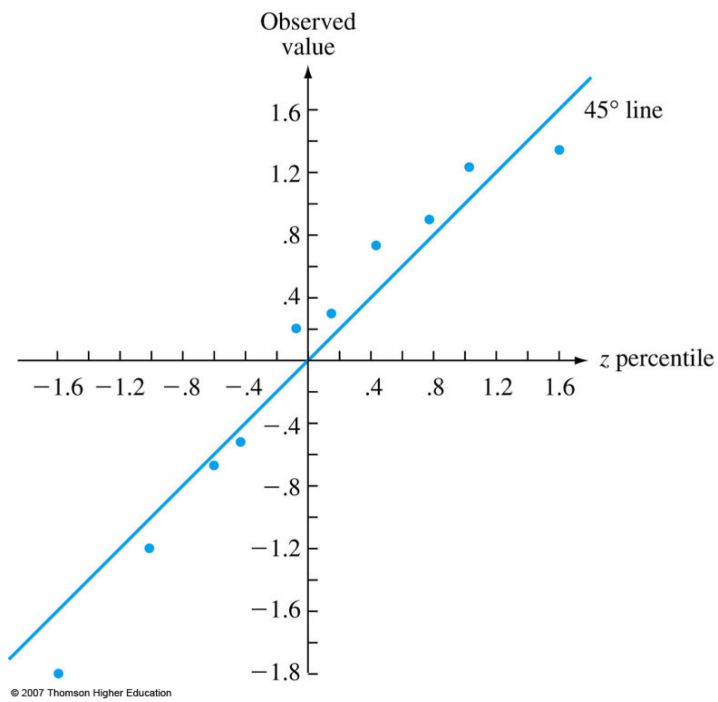
$$u_i = \left(\frac{i - 0.5}{n}\right) \mathcal{N}(0,1) \text{ quantile}, \qquad v_i = x_{(i)}, \qquad i = 1, \ldots, n$$

  is called a **normal probability plot** (or **qq-plot**).

- Use the qq-plot to assess whether the dataset $\{x_1, \ldots, x_n\}$ could plausibly have come from a $\mathcal{N}(\mu, \sigma^2)$ distribution...

- Because $X \sim \mathcal{N}(\mu, \sigma^2)$ and $Z \sim \mathcal{N}(0,1)$ are related via a linear transformation, $X = \mu + \sigma Z$, we don't need to know μ & σ...

- We just check if the (u_i, v_i) points are plausibly falling on a straight line, and if so, we would have no reason to discard the assumption of normality.
- **Ex:** this qq-plot is very typical of what would happen if the blue points did truly come from a normal (the blue line is simply a "best-fit" line through the points).



Ch. 5-6: Sampling Distributions And Concepts of Point Estimation

§5.3 Statistics and their distributions

§5.4 Distribution of the sample mean

§5.5 The distribution of linear combinations

§6.1 General concepts of point estimation

Count: slides 132–153 (22 slides).
([*] denotes content that will be given in class)

§5.3 Statistics and Their Distributions

- A **statistic** is any value that can be calculated from sample data.
- I.e., it is a function only of the sample $x_1, \ldots, x_n$, and cannot therefore depend on unknown quantities like parameters of a distribution.
- Because the statistic is a function of $x_1, \ldots, x_n$, which are the realized values of random variables $(X_1, \ldots, X_n)$, the statistic itself is a random variable.
- Prior to data collection, there is uncertainty about the value of the statistic, i.e., it has a distribution.
- Once the data are collected, we can evaluate the observed value of the statistic.

Common Statistics

- $\overline{x}$: sample mean (realized value of $\bar{X}$).
- $\tilde{x}$: sample median (realized value of $\tilde{X}$).
- s: standard deviation (realized value of S).
- $\hat{p}$: proportion (realized value of $\hat{P}$).

These are sometimes also called *point estimates* of some parameter of the underlying distribution that gave rise to the data.

Random Samples

- The distribution of a statistic is called its **sampling distribution**.
- Since evaluation of the sampling distribution is in general quite difficult, we frequently make simplifying assumptions.
- **Common simplifying assumption:** the data $X_1, \ldots, X_n$ constitute a **random sample**, that is, the X_i are **independent and identically distributed (IID)**.
- Simulation experiments allow us to approximate the distributions of complicated statistics…

Software Packages (Read §5.3)

- Allow the generation of random samples $X_1, \ldots, X_n$ from many different families of distributions. This is called **simulation**.
- Given such a sample, we can evaluate the statistic of interest, e.g., the sample mean $\bar{X}$, and then repeat the process a large number of times.
- The values of the statistic from this procedure can then be used to investigate its sampling distribution; e.g., via a histogram.
- By changing the simulation settings, we can examine how the sampling distribution changes…
- **Ex:** what happens to $\overline{X}$ as $n \to \infty$?

§5.4 The Distribution of the Sample Mean [*]

Let $X_1, \ldots, X_n$ be an IID sample from a distribution X with $\mathbb{E}(X) = \mu$ and $\mathbb{V}(X) = \sigma^2$. Consider the following statistics:

$$T_n = \sum_{i=1}^{n} X_i = \text{ sample total}, \qquad \overline{X}_n = T_n/n = \text{ sample mean}$$

- ▷ The mean and variance of T_n are:

- ▷ The mean and variance of $\bar{X}_n$ are:

- ▷ If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $\overline{X}_n$ and T_n are also normal.

The Central Limit Theorem (CLT) [*]

Theom: Let $X_1, \ldots, X_n$ be an IID sample from a distribution X with $\mathbb{E}(X) = \mu$ and $\mathbb{V}(X) = \sigma^2$. Then, it follows that:

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \longrightarrow \mathcal{N}(0,1), \qquad \text{as } n \to \infty$$

Remarks:

- In practice, and for sufficiently large n, this result means that: $\overline{X}_n \approx \mathcal{N}(\mu, \sigma^2/n)$.
- Result holds **regardless of the distribution of X**, which can even be discrete!
- Convergence in CLT is faster for symmetric and continuous X, but in general we can rely on the Rule of Thumb:

Simulation Exercise to Illustrate CLT [*]

- Let $X \sim \text{Unif}(-1, 1)$. Then the mean and variance of X are:

- Simulate a random sample $X_1, \ldots, X_n$ from X at each of sample sizes $n = 1, 2, \ldots, 6$, and form respective sample means:

$$\overline{X}_1, \quad \overline{X}_2, \quad \overline{X}_3, \quad \overline{X}_4, \quad \overline{X}_5, \quad \overline{X}_6$$

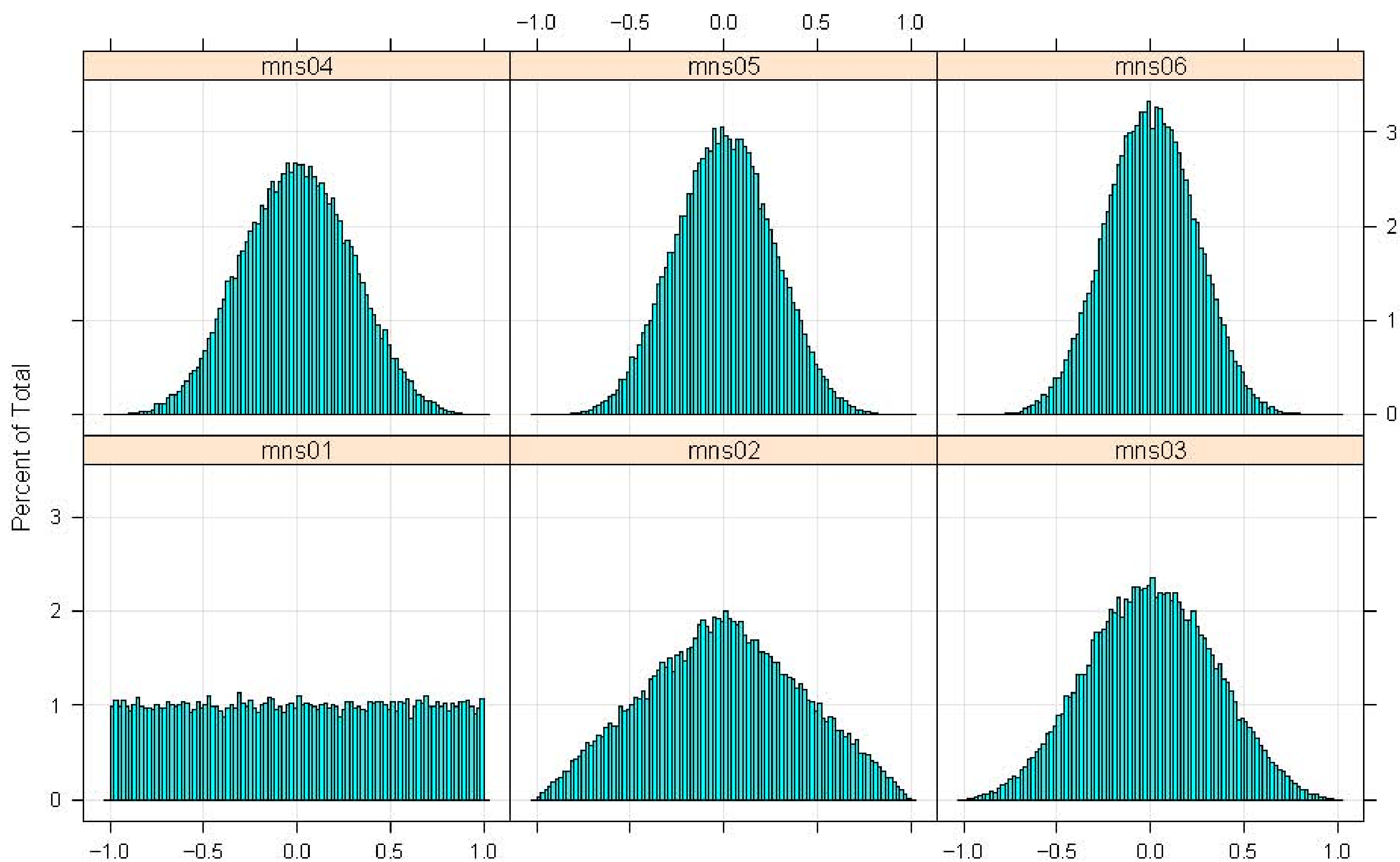
- Repeat (replicate) $50,000$ times: so we have $50,000$ sample means at each of the 6 sample sizes.

- Histograms of these will show convergence to normality with decreasing variance (next slide).

- The exact statement from the CLT is:

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} = \qquad\qquad\qquad \longrightarrow \mathcal{N}(0, 1), \quad \text{as } n \to \infty$$

Histograms of raw means of samples from U[-1,1].



§5.5 Distribution of Linear Combinations [*]

Let $X_1, \ldots, X_n$ be **any** collection of r.v.'s with respective means and variances $\mathbb{E}(X_i) = \mu_i$ and $\mathbb{V}(X_i) = \sigma_i^2$. For $a_1, \ldots, a_n$ arbitrary real numbers, consider the **linear combination**:

$$Y = a_1 X_1 + \cdots + a_n X_n$$

Facts:

- The mean Y is:

- If the X_i's are **independent**, then the variance of Y is:

- If the X_i's are **independent and normal**, then Y is normal.

The Difference Between Two RV's [*]

A special case of the previous slide is the difference:

$$Y = X_1 - X_2$$

In this case:

(i) $\mathbb{E}(Y) = \mu_1 - \mu_2$.

(ii) If X_1 and X_2 are independent, then the variance of Y is:

- In (ii), note that it's the variances that sum, not the standard deviations:

$$\sigma_Y = \sqrt{\sigma_1^2 + \sigma_2^2} \neq \sigma_1 + \sigma_2$$

Problem 5.80 (7th Ed.) [*]

The luggage weight X of a randomly selected airline passenger is a r.v. with a mean of 30 lbs and standard deviation of 6 lbs. On a particular flight 50 passengers check in luggage.

(a) How unusual is a sample mean checked luggage weight in excess of 35 lbs?

(b) Calculate the mean and standard deviation of the total luggage weight.

Problem 5.80 (7th Ed.) Continued [*]

(c) If individual luggage weights are independent and normally distributed, what is the probability that the total luggage weight is at most $1,600$ lbs?

§6.1 General Concepts of Point Estimation

- A **(point) estimate** of a parameter θ is a single value, $\hat{\theta}$, that can be regarded as a sensible value for θ.
- The corresponding r.v., $\widehat{\Theta}$, is called an **estimator**. (This parallels the concept of a r.v. X and a realized value x "drawn" from the distribution of X.)
- For common parameters there is an obvious estimator, e.g.:
  - If $X \sim \mathrm{Bin}(n, p)$, we estimate p with the proportion of successes in the sample: $\widehat{P} = X/n$.
  - If X is continuous with mean μ and variance σ^2, good estimators of these two parameters are the sample mean and variance:

$$\hat{\mu} = \overline{X} = \sum_{i=1}^{n} X_i, \qquad \hat{\sigma}^2 = S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$$

Ex 6.2 (7th Ed.): 20 obs on dielectric breakdown voltage

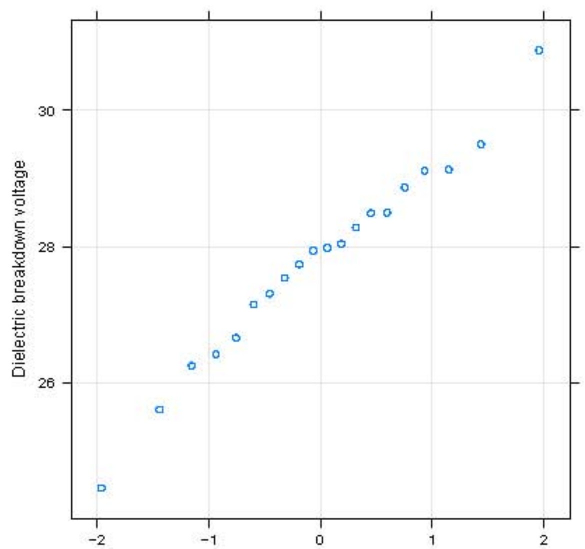
Example 6.2

A normal probability plot shows that the data closely follow a normal shape

```
> str(xmp06.02)

'data.frame': 20 obs. of  1 variable:
 $ Voltage: num  24.5 25.6 26.2 26.4 26.7 ...

> qqmath(~Voltage, xmp06.02)
```



Problem 6.2 (7th Ed.) Continued [*]

Since the distribution is symmetric (assuming normal), then we have at least 4 possible competing estimators for the mean:

(i)

(ii)

(iii)

(iv)

Question: which should we choose?

The Expectation & Variance of an Estimator [*]

- An estimator $\widehat{\Theta}$ of θ is said to be **unbiased** if: $\mathbb{E}(\widehat{\Theta}) = \theta$.
- The sample mean $\overline{X}$ is unbiased for μ if the sample is IID.
- The sample median $\widetilde{X}$ is unbiased for μ if the sample is IID **and** the distribution of the X_i is continuous and symmetric.
- The sample variance $S^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2/(n-1)$ is unbiased for σ^2 if the sample is IID. However, the sample standard deviation S is a **biased** estimator of σ, since $\mathbb{E}(S) \neq \sigma$.
- Given two unbiased estimators, we generally prefer the one with:

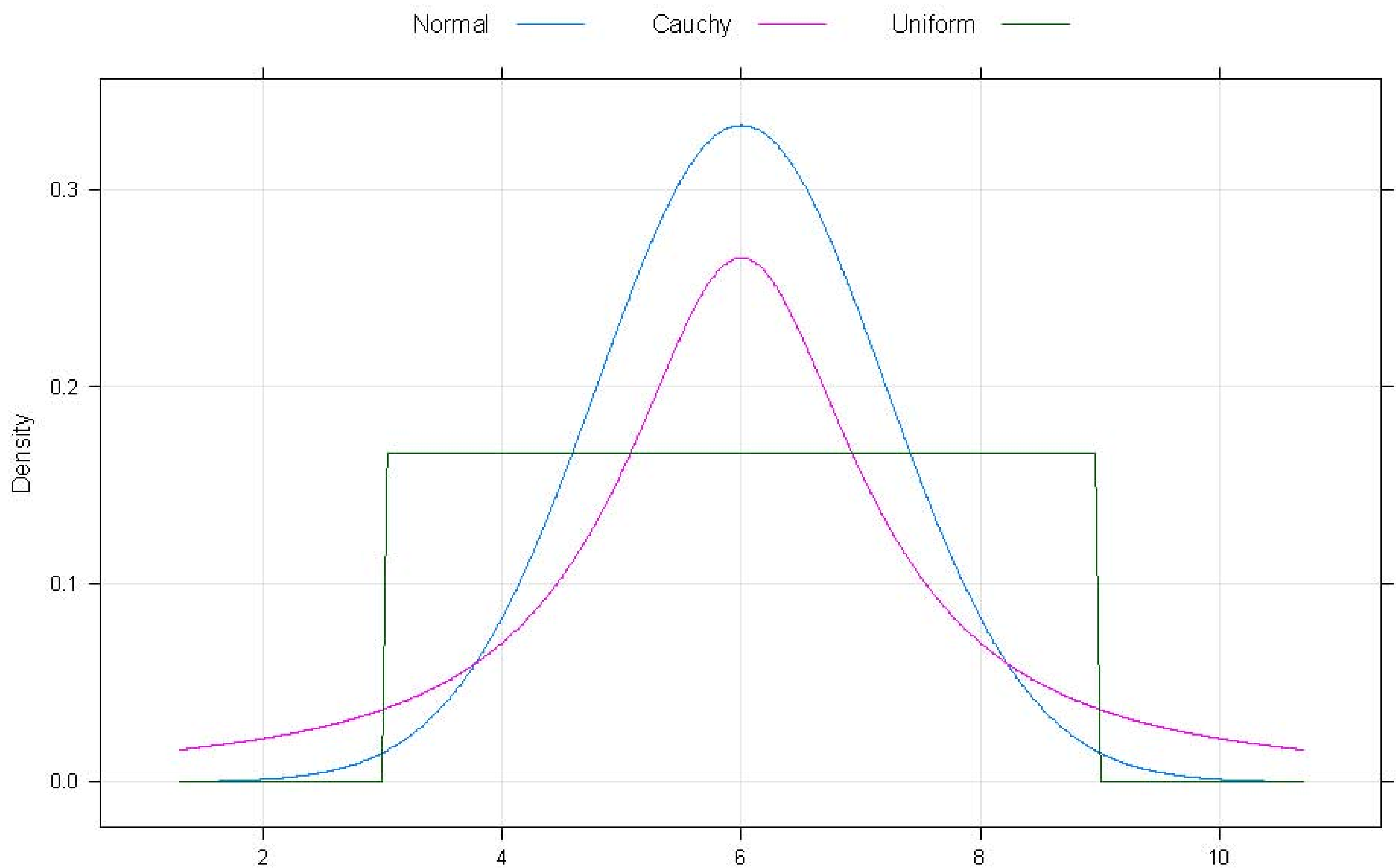
Simulation Exercise: Compare 4 Estimators of the Mean

Set $\mu = 6$ and $\sigma = 1.2$, the sample size at $n = 25$, and consider the following 4 estimators of μ:

$$\overline{X}, \quad \widetilde{X}, \quad \text{midrange} = \frac{X_{(1)} + X_{(n)}}{2}, \quad \text{trimmed mean (5\%)}$$

- **Do loop (outer):** for $X \in \{\text{Normal, Cauchy, Uniform}\}$:
  - **Do loop (inner):** repeat 50000 times:
    - draw a random sample $X_1, \ldots, X_n$ from X;
    - compute the 4 estimators of μ.

Comparison of the distributions



Results of Simulation Exercise [*]

Empirical means (standard deviations) of the 50000 replicates. Which is the best estimator in each of the 3 cases?

Dist. of X	**mean**	**median**	**midrange**	**trimmed mean**
Normal	6.000	6.000	5.996	6.000
	(0.2401)	(0.2983)	(0.4374)	(0.2422)
Cauchy	2.394	6.001	-39.19	6.010
	(735.5)	(0.3994)	(9193)	(2.000)
Uniform	6.000	5.999	6.000	6.000
	(0.3465)	(0.5770)	(0.1610)	(0.3704)

The Standard Error of an Estimator [*]

- Simulation Exercise: shows that in addition to the point estimate $\widehat{\Theta}$ of θ, we need some measure of its precision...
- The **standard deviation** of the estimate could be used, but it will usually depend on (unknown) parameters..., e.g.:
  - for $X \sim \text{Bin}(n, p)$: the usual estimator $\widehat{P} = X/n$ has a standard deviation of $\sqrt{p(1-p)/n}$;
  - for X continuous with mean μ and variance σ^2: the usual estimator of μ, $\overline{X}$, has a standard deviation of $\sigma/\sqrt{n}$.
- Using estimates for the unknown parameters in the standard deviation formula, gives the **standard error (s.e.)** of the estimator; e.g.:

Important Takeaway Messages From Chs. 5-6

- If $X_1, \ldots, X_n \sim \mathrm{IID}(\mu, \sigma^2)$, then:
  - using properties of $\mathbb{E}(\cdot)$ and $\mathbb{V}(\cdot)$: $\overline{X} \sim \left(\mu, \frac{\sigma^2}{n}\right)$
  - if n is large, using the CLT: $\overline{X} \approx \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$
- If $X \sim \mathrm{Bin}(n, p)$, then:
  - using properties of $\mathbb{E}(\cdot)$ and $\mathbb{V}(\cdot)$: $\widehat{P} \sim \left(p, \frac{p(1-p)}{n}\right)$
  - if n is large, using the CLT: $\widehat{P} \approx \mathcal{N}\left(p, \frac{p(1-p)}{n}\right)$

Ch. 7: Confidence Intervals

From A Single Sample

§7.1 Basic properties of confidence intervals

§7.2 Large sample confidence intervals

§7.3 Intervals based on a normal population distribution

§7.4 Confidence intervals for the variance & std. deviation

Count: slides 154–178 (25 slides).
([*] denotes content that will be given in class)

§7.1 Basic Properties of Confidence Intervals

With a random sample $X_1, \ldots, X_n$ drawn from the distribution of $X \sim (\mu, \sigma^2)$ as background, recall basic concepts from Ch. 6:

- **Estimate** $(\hat{\theta})$**:** a sensible value for θ.
- **Estimator** $(\widehat{\Theta})$**:** the r.v. corresponding to $\hat{\theta}$. ($\hat{\theta}$ is a draw from the distribution of $\widehat{\Theta}$.)

Now: introduce a new concept...

Def: A **confidence interval (CI)** is an *interval* (instead of point) estimate for θ which has a given probability of covering (or containing) the true value of θ.

CI For μ When σ is Known [*]

- We assume the population is normal and σ is known, i.e., $X_1, \ldots, X_n \sim \text{IID } \mathcal{N}(\mu, \sigma^2)$. Then, from Ch. 5 we have that:

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$$

- From this it follows that:

$$P\left(-1.96 < \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} < 1.96\right) = 0.95$$

- Which implies that:

$$P\left(\overline{X} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95$$

- And thus leads to the 95% CI on μ:

Example 7.2 (7th Ed.) [*]

Have $n = 31$ obs from $X \sim \mathcal{N}(\mu, \sigma = 2)$ which gives $\overline{x} = 80.0$, and want a 95% CI for μ:

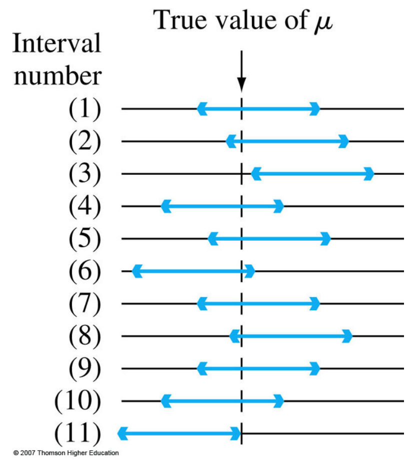
Question: what exactly does this CI mean?

Simulation Exercise: Coverage of 95% CIs for the Mean

Let $X \sim \mathcal{N}(\mu = 6.3, \sigma = 0.75)$, set sample size at $n = 10$, and recall the 95% CI for μ: $\overline{x} \pm 1.96\sigma/\sqrt{n}$.

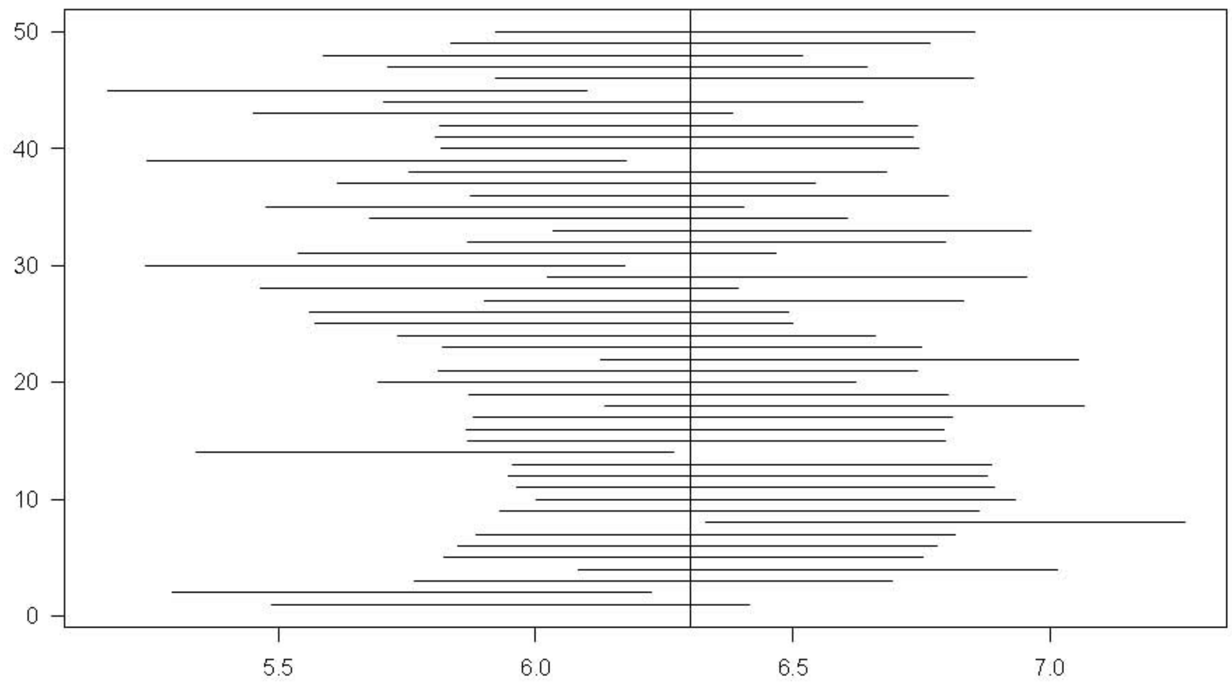
Do loop: repeat $m = 50$ times:

- draw a random sample $X_1, \ldots, X_n$ from X;
- compute a 95% CI for μ.



Plots of simulated confidence intervals

In this case only 44/50 or 88% of the intervals cover $\mu = 6.3$.



Answer to Question: If we replicate the simulation $m = \infty$ times, then 95% of the CIs would contain μ.

The General $(1-\alpha)$100% CI for the Mean [*]

When $X \sim \mathcal{N}(\mu, \sigma^2)$ and σ is known, the $(1-\alpha)100\%$ CI for μ is:

Some common values of the quantile $z_{\alpha/2}$ are:

$(1-\alpha)100\%$	α	$\alpha/2$	$z_{\alpha/2}$
80%	0.200	0.100	1.282
90%	0.100	0.050	1.645
95%	0.050	0.025	1.960
99%	0.010	0.005	2.576

Values are from: last row of Table A.5 (t distribution with $df = \infty$).

Confidence Level & Sample Size [*]

- **Confidence level.** Increasing $(1-\alpha)$ leads to a decrease in α, and consequently an increase in the appropriate quantile $z_{\alpha/2}$. In the limit, a 100% CI has $\alpha = 0$, so that $z_{\alpha/2} = \infty$, which leads to a CI of $(-\infty, \infty)$.

- **Sample size.** Since the width of the $(1-\alpha)\%100$ CI is $w = 2z_{\alpha/2}\sigma/\sqrt{n}$, solving for n gives:

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

- **Ex:** If $\sigma = 25$, how large does n need to be for a 95% CI to have width $w = 10$?

§7.2 Large Sample CIs [*]

- **Problem:** in practice σ is usually unknown, so how can we get around the need to have both a known σ and a normal X in the previous section?
- **Solution:** the CLT! Also, since n is large enough, s will be a good approx to σ, and thus:

$$Z = \frac{\overline{X} - \mu}{s/\sqrt{n}} \quad \approx \quad \mathcal{N}(0,1)$$

- **Large Sample CI:** this leads to the $(1-\alpha)\%100$ CI for μ:
- **Rule of Thumb:** this holds provided $n > 40$.

Example 7.6 (7th Ed.) [*]

Have $n = 48$ obs on breakdown voltage which gives $\overline{x} = 54.7$, $s = 5.23$, and want a 95% CI for μ:

Sample Size For Given CI Width [*]

- The formula from earlier for the n needed to get a $(1-\alpha)\%100$ CI of width w is:

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

- **Problem:** How to estimate σ, since this is to be used before collecting data?

- **Answer:** A rough estimate of σ is obtained by:

CIs for a Population Proportion [*]

For a random sample of binary data (yes/no, S/F, etc.), distribution of number of successes X is binomial: $X \sim \text{Bin}(n, p)$.

- **Estimator of p:** $\widehat{P} = \#\ \text{successes}/n = X/n$.
- **Standard Deviation of $\widehat{P}$:** $\sqrt{p(1-p)/n}$.
- **CLT:** If n is large enough we have:

$$Z = \frac{\widehat{P} - P}{\sqrt{p(1-p)}} \quad \approx \quad \mathcal{N}(0,1)$$

- **Large Sample CI:** this leads to the $(1-\alpha)\%100$ CI for p:

- **Rule of Thumb:** holds provided $n\hat{p} \geq 10$ and $n(1-\hat{p}) \geq 10$.

- **Adjusted CI for p:** Book gives a more accurate CI for p that is valid for all n:

$$\frac{\left(\hat{p} + \frac{z_{\alpha/2}^2}{2n}\right) \pm z_{\alpha/2}\sqrt{\frac{z_{\alpha/2}^2}{4n^2} + \frac{\hat{p}(1-\hat{p})}{n}}}{1 + \frac{z_{\alpha/2}^2}{n}}$$

- **Finding n for a $(1-\alpha)\%100$ CI of width w:** Using same ideas as before:

$$n \approx \left(2z_{\alpha/2}\frac{\sqrt{\hat{p}(1-\hat{p})}}{w}\right)^2$$

  - $\hat{p}$ will be unknwon if no data has yet been collected, but setting $\hat{p} = 1/2$ provides a conservative approach by giving the largest possible n.
  - A more complex formula based on the adjusted CI is available (see book), but we won't use it in this course...

Example 7.8 & 7.9 (7th Ed.) [*]

Have $n = 48$ trials in a lab, 16 of which resulted in S (ignition of substrate by cigarette). Let p denote the population proportion of all trials resulting in S.

(a) Construct a large sample 95% CI for p.

(b) Find the required n for the CI to have width 0.1.

Example 7.8 & 7.9: Continued [*]

(c) Construct an adjusted 95% CI for p.

One-Sided $(1-\alpha)\%100$ **CIs**

- Sometimes one is only interested in bounding the parameter estimate from above or below.
- The difference with two-sided CIs is that the quantile goes from $z_{\alpha/2}$ to z_α on the finite side, while the other side becomes infinite.
- **Ex:** in a one-sided CI for μ, the upper and lower CIs are:

$$\left(-\infty,\ \overline{x}+z_\alpha\frac{s}{\sqrt{n}}\right) \qquad \text{and} \qquad \left(\overline{x}-z_\alpha\frac{s}{\sqrt{n}},\ \infty\right)$$

- **Note:** in all cases, the CIs enclose an area (probability) of $(1-\alpha)$ under the appropriate normal pdf.

Example 7.10 (7th Ed.) [*]

Have $n = 48$ shear strength obs which give $\overline{x} = 17.17$ and $s = 3.28$. Construct a lower 95% confidence bound for μ.

§7.3 Intervals Based on a Normal Distribution

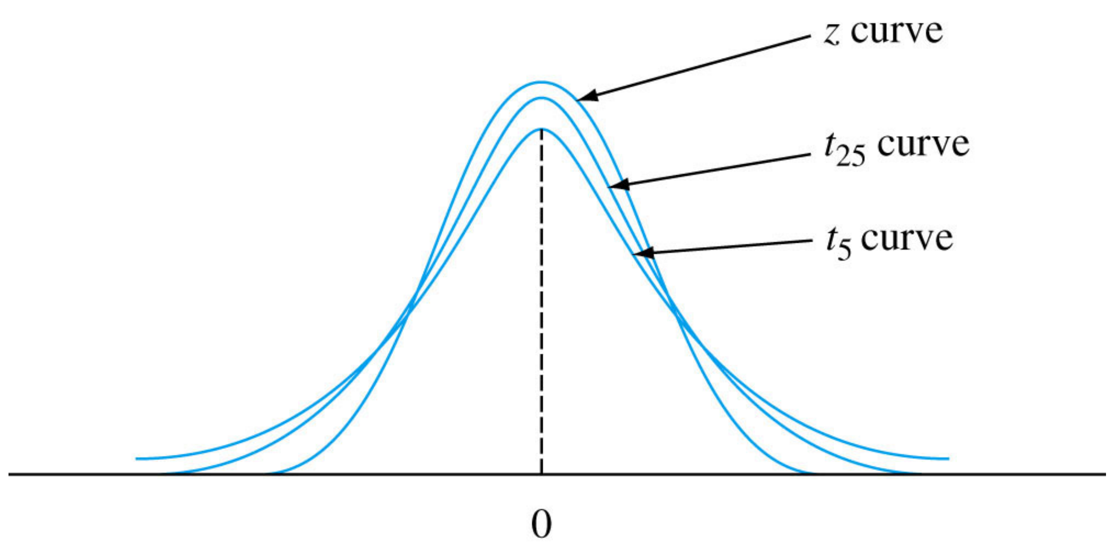
- In §7.2 we saw that large sample situations are the most versatile since they easily accommodated all unknowns.
- **Now:** tackle case when $n \leq 40$ and σ is unknown...
- **There is no magic trick:** must make assumptions to compensate for lack of information. We assume the X_i come from a **normal distribution**. This can be checked with a normal probability plot.
- t **distribution:** Assuming $X \sim \mathcal{N}(\mu, \sigma^2)$, then:

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}} \quad \sim \quad t_{n-1}$$

a t **distribution with** $(n-1)$ **degrees of freedom (df)**.

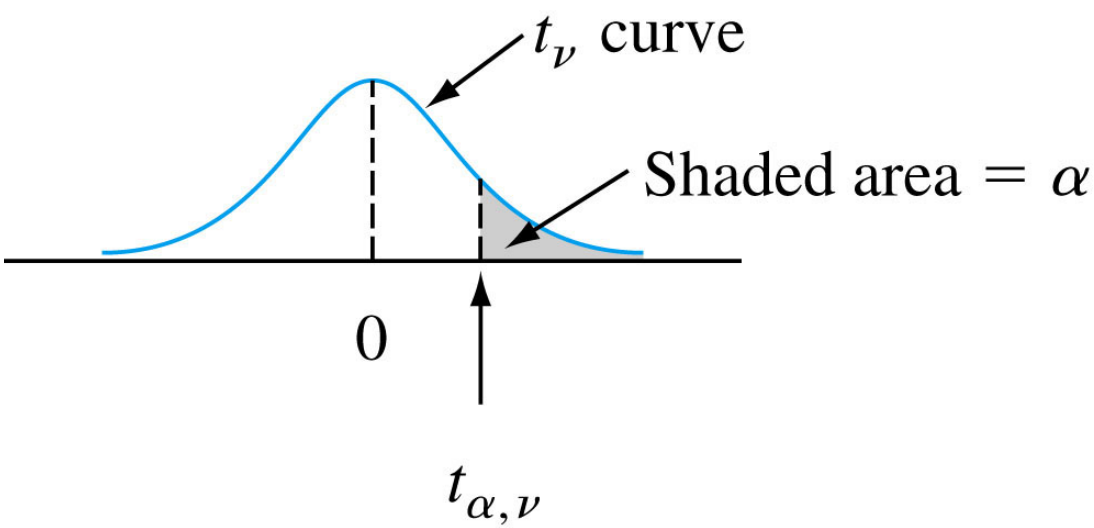
Properties of the T Distribution

- Each t_ν is centered at 0 and has the same basic shape as a $\mathcal{N}(0,1)$.
- However, the t_ν is more spread out than the $\mathcal{N}(0,1)$.
- As $\nu \uparrow \infty$, the t_ν converges to the $\mathcal{N}(0,1)$.



T Distribution Quantiles (Critical Values)

- A $t_{\alpha,\nu}$ is the $(1-\alpha)$ quantile from a t distribution with ν degrees of freedom (df). This is also called a t **critical value**.
- Critical values for selected ν and α are given in **Table A.5** (and on the inside back cover).



▷ **Recall:** in slide 160 we gave the critical values of $z_{\alpha/2}$ as:

$(1-\alpha)100\%$	α	$\alpha/2$	$z_{\alpha/2}$
80%	0.200	0.100	1.282
90%	0.100	0.050	1.645
95%	0.050	0.025	1.960
99%	0.010	0.005	2.576

▷ These values are exactly:

$$z_{\alpha/2} = t_{\alpha/2,\infty}$$

obtained from the last row of Table A.5 (a t with $df = \infty$).

T-Based $(1-\alpha)\%100$ CIs and PIs [*]

- **Two-Sided CI:** The two-sided $(1-\alpha)\%100$ CI for μ is:

- **One-Sided CI:** One-sided $(1-\alpha)\%100$ lower & upper bound CIs for μ are:

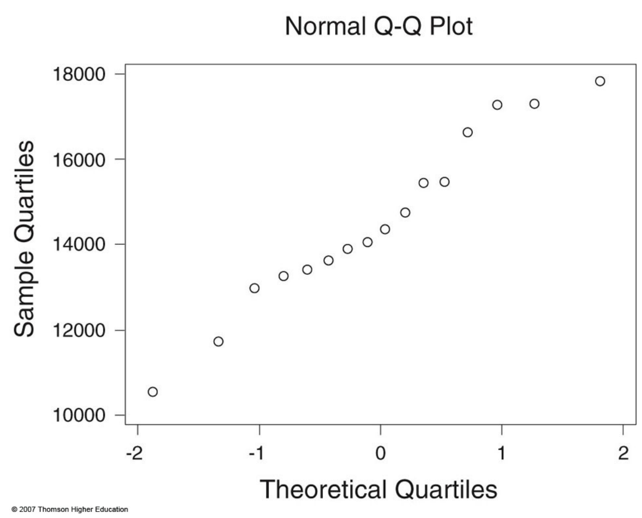
- **Prediction Interval (PI):** This is a CI on a single value of X, not $\mathbb{E}(X) = \mu$, and will thus have more uncertainty (variability). The formula for a $(1-\alpha)\%100$ PI is:

Example: Pine Lumber (7th Ed.) [*]

Have modulus of elasticity values for $n = 16$ pine lumber specimens, which give $\overline{x} = 14532.5$ and $s = 2055.67$.

(a) Construct a 95% CI for μ.

(b) Check the normality assumption.



Example 7.13: (7th Ed.) [*]

Have $n = 10$ obs on fat content of hot dogs give $\overline{x} = 21.9$ and $s = 4.134$. A 95% PI for the fat of content of my next hot dog is:

(**Note:** PIs are rarely used in practice.)

Flowchart for $(1-\alpha)\%100$ CI for μ [*]

Ch. 8: Tests of Hypotheses [*]

For A Single Sample

§8.1 Hypotheses and Test Procedures

§8.2–8.3 Tests About a Population Mean

§8.4 Tests About a Population Proportion

Count: slides 179–197 (19 slides).

([*] denotes content that will be given in class)

§8.1 Hypotheses and Test Procedures

- **Statistical Hypothesis.** This is a claim about the value of one or more population parameters; commonly used in comparative experiments. E.g., μ_1 and μ_2 are mean compressive strengths of carboard produced by two different processes, and we may want to test the claim that $\mu_1 > \mu_2$.
- **Null Hypothesis.** This claim usually represents the *status quo* of "no change" or "nothing new". It is written H_0, and we typically wish to find evidence against it.
- **Alternative Hypothesis.** This is the opposite claim to the null, usually representing an interesting change to the *status quo*. It is written H_a, and we typically wish to find evidence in favor of it.

Similarity With Judicial System

- **Effect of Variability.** We can never "prove" H_a (guilt) or H_0 (innocence). All we can do is quantify (by means of a **p-value**) how unlikely it is for H_a to have given rise to the data we observed if H_0 were true...

- **Conclusions.** The result of a test is that we either:
  - **fail to reject H_0** (not enough evidence to reject it), or
  - **reject H_0** (there is enough evidence to reject it).

- **Analogy with mathematical logic.** Suppose want to prove claim: "there is no largest positive integer". The *proof by contradiction* argument starts by supposing n is the largest integer (H_0). But then $n + 1 > n$, which is a contradiction, whence the original claim must be false (reject H_0)! However, with hypothesis tests we only get "unlikely", not "impossible".

Errors in Hypothesis Testing [*]

		Truth	
		H_0	H_a
Decision	H_0	correct (prob=)	type II error (prob=)
	H_a	type I error (prob=)	correct (prob=)

- **Type I Error:** reject H_0 when true. Usually the worst kind of error, so we control it by choosing $\alpha = P(\text{Type I Error})$.
- **Type II Error:** fail to reject H_0 when H_a true. Having chosen α, we have no control over $\beta = P(\text{Type II Error})$, other than by increasing the sample size n.

Notes [*]

- The **total error** is NOT $\alpha + \beta$.
- The **power** of the test is $1 - \beta$.
- α is also called the **level of significance** of the test; sometimes abbreviated to **level α test**.
- Increasing α leads to a decrease in β (all else constant).
- Decreasing α leads to an increase in β (all else constant).
- **Ex 1:** compute α and β for the test that always accepts H_0.
- **Ex 2:** compute α and β for the test that always rejects H_0.

Steps in Performing a level α Hypothesis Test [*]

1. Define H_0 and H_a. (Implicitly assume H_0 is true.)
2. Decide what α should be. (Usually $0.01 \leq \alpha \leq 0.10$.)
3. Compute the **test statistic**: a function of the data on which to base the decision (H_0 or H_a).
4. Determine either:
   (i) the **rejection region** for the level α test (the set of all test statistic values for which H_0 will be rejected); or
   (ii) the **p-value** for the test.
5. Determine if H_0 can be rejected by checking if:
   (i)
   (ii)

The P-Value [*]

Def: The **p-value** is the probability of observing a value of the test statistic as or more "extreme" (contradictory to H_0) than what was actually observed, assuming H_0 is true.

Notes:

- P-value (a.k.a. the significance level) is the smallest level of significance at which H_0 would be rejected (see figure below):

- $0 \leq$ p-value ≤ 1: is the amount of evidence in favor of H_0.

- Don't need to know α in order to compute the p-value...

The P-Value (Continued) [*]

- P-values for z-based tests can be calculated to 4 decimal places from z-table (Table A.3).
- P-values for t-based tests can be calculated to 3 decimal places from **Table A.8**.
- **Statistical vs. practical significance.** A test produces a p-value < 0.001, and so the test is "statistically significant". Is this really important?

Example 8.2: (7th Ed.) [*]

Let X be the drying time of paint (mins), and the established knowledge (H_0) is that $X \sim \mathcal{N}(\mu = 75, \sigma = 9)$. A new additive is purported to decrease the mean drying time (σ is unchanged). Main objective: test if additive really works.

- Test:

- Data:

- CLT:

- Rej. Region:

Example 8.2: (Continued) [*]

- ▷ Type I error:

- ▷ Type II error:

§8.2–8.3 Tests About The Population Mean (μ) [*]

For Cases 1–3 below we write the null hypothesis as $H_0 : \mu = \mu_0$, and we consider the 3 types of alternatives in each case.

- **Case 1.** Normal population with known σ. The test stat is:

$$z = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} \quad \sim \quad \mathcal{N}(0,1), \text{ under } H_0$$

H_a	Rej. Region	P-value	Picture
$\mu > \mu_0$			
$\mu < \mu_0$			

▷ **Case 1.** Normal population with known σ (continued). [*]

H_a	Rej. Region	P-value	Picture
$\mu \neq \mu_0$			

H_a	Formula for $\beta = \beta(\mu')$	n for $\beta = \beta(\mu')$
1-tailed		
2-tailed		

▷ **Case 2.** Large ($n > 40$) sample (any population). CLT implies:

$$z = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} \quad \approx \quad \mathcal{N}(0,1), \text{ under } H_0$$

Testing procedures are identical to Case 1 with $\sigma \to s$.

▷ **Case 3.** Normal population with known σ. From Ch. 7: [*]

$$t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} \quad \sim \quad t(n-1), \text{ under } H_0$$

H_a	Rej. Region	P-value	Picture
$\mu \neq \mu_0$			

▹ **Case 3.** Normal population with known σ (continued). [*]

H_a	Rej. Region	P-value	Picture
$\mu > \mu_0$			
$\mu < \mu_0$			

Note: Formulas for Type II error, $\beta(\mu')$, and n, don't exist in analytical form (must resort to computer).

Problem 8.56: (7th ed.) [*]

$n = 30$ obs on % organic matter in soil specimens gives $\overline{x} = 2.481$ and $s = 1.616$. The data look normal. Is there evidence that μ differs from 3? (Test at levels $\alpha = 0.10$ and $\alpha = 0.05$.)

Example 8.8: (7th ed.) [*]

Require DCP average value for pavement to be < 30 (must conclusively show this before usage). A sample of $n = 52$ DCP obs gives $\overline{x} = 28.76$ and $s = 12.2647$. The data do NOT look normal.

- Is there enough evidence that $\mu < 30$?

Example 8.8 (Continued) [*]

- What is the power of the test if $\mu' = 28.76$? (Assume $\sigma \approx s$.)

- What n is needed to have power of 80% if $\mu' = 28.76$?

§8.4 Tests On a Population Proportion (Large Sample)

- Write the null as $H_0 : p = p_0$.
- Consider only large samples: $np_0 \geq 10$ and $n(1 - p_0) \geq 10$.
- If X is the # of successes in a sample of size n, the CLT implies $X/n \sim \mathcal{N}(p, p(1-p)/n)$, so that:

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \quad \approx \quad \mathcal{N}(0,1), \text{ under } H_0$$

- Testing procedures are identical to Case 1 of tests on the mean (see slides 189–190); just replace $\mu \to p$ and $\mu_0 \to p_0$.
- Formulas for Type II error, $\beta(p')$, and the n needed for a level α test to have $\beta = \beta(p')$, are also available, but are more complicated (see §8.4 of book).

Example 8.12: (7th ed.) [*]

A package delivery service claims to be 90% on time. A sample of $n = 225$ deliveries shows that 80% of them are on time.

- Is there enough evidence that $p < 0.9$?

- If $p' = 0.8$, how likely is it that an $\alpha = 0.01$ test based on $n = 225$ will detect such a departure from H_0?

Ch. 9: Inferences Based on Two Samples [*]

§9.1 Differences between two population means

§9.2 The two-sample t-test and confidence interval

§9.3 Analysis of paired data

§9.4 Inferences on the difference between two population proportions

Count: slides 198–212 (15 slides).

([*] denotes content that will be given in class)

§9.1: Differences Between Two Population Means

The situation is as follows.

- $X_1, \ldots, X_m$ is a random sample from $X \sim (\mu_1, \sigma_1^2)$.
- $Y_1, \ldots, Y_n$ is a random sample from $Y \sim (\mu_2, \sigma_2^2)$.
- The X and Y samples are independent from each other.
- We are interested in the difference in the (population) means:

$$\mu_1 - \mu_2 := \Delta_0$$

- The respective sample means are: $\overline{x}$ and $\overline{y}$.
- In realistic situations, the variances σ_1^2 and σ_2^2 are not known, in which case we use the corresponding sample variances: s_1^2 and s_2^2.

§9.2 Two-Sample Tests & CIs For $\mu_1 - \mu_2$ [*]

For Cases 1–3 below we write the null as

$$H_0 : \mu_1 - \mu_2 = \Delta_0$$

Given the appropriate test statistic, the procedures are then identical to Ch. 8. (We skip formulas for β and n.)

- **Case 1.** Normal populations with known σ_1 and σ_2.
  - The test stat is:

$$z = \frac{(\overline{x} - \overline{y}) - \Delta_0}{\sqrt{\sigma_1^2/m + \sigma_2^2/n}} \quad \sim \quad \mathcal{N}(0,1), \text{ under } H_0$$

  - The formula for a $(1-\alpha)100\%$ CI for $\mu_1 - \mu_2$ is:

▷ **Case 2.** Any populations with large samples ($m > 40$ and $n > 40$). [*]

- The test stat is:

$$z = \frac{(\overline{x} - \overline{y}) - \Delta_0}{\sqrt{s_1^2/m + s_2^2/n}} \quad \sim \quad \mathcal{N}(0,1), \text{ under } H_0$$

- The formula for a $(1-\alpha)100\%$ CI is:

▷ **Case 3.** Normal populations with unknown σ_1 and σ_2. [*]

- The test stat is:

$$t = \frac{(\overline{x} - \overline{y}) - \Delta_0}{\sqrt{s_1^2/m + s_2^2/n}} \quad \sim \quad t_\nu, \text{ under } H_0$$

- The formula for a $(1-\alpha)100\%$ CI is:

$$(\overline{x} - \overline{y}) \pm t_{\alpha/2,\nu} \sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}$$

- Formula for degrees of freedom ν:

- If can assume $\sigma_1 = \sigma_2$, then **pooled t procedures** apply:

Example 9.7: (7th ed.) [*]

Data on tensile strength (psi) of specimens, both when a fusion process was used (Y) and not used (X). The sample info is:

$$m = 10, \qquad \overline{x} = 2902.8, \qquad s_1 = 277.3, \quad \text{(no fusion)}$$

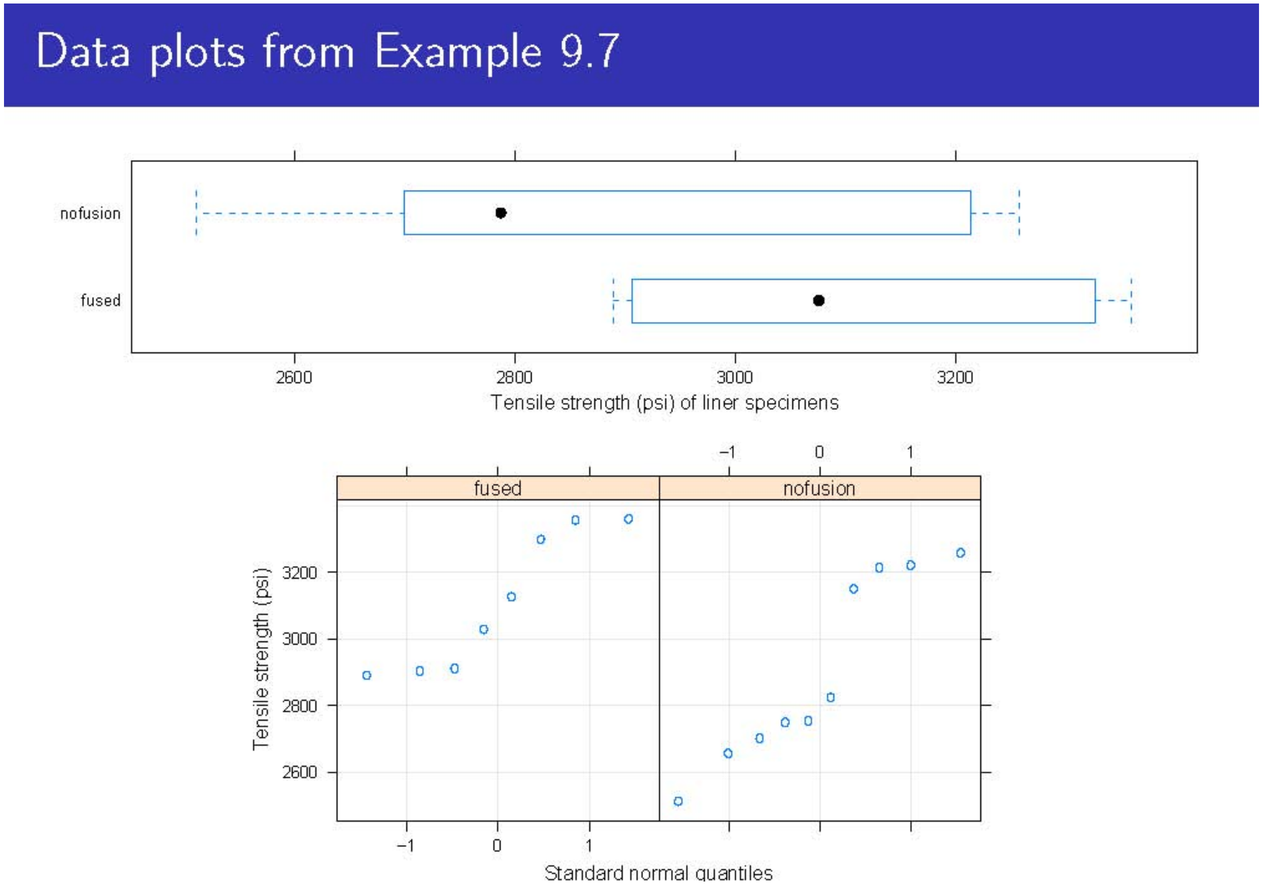
$$n = 8, \qquad \overline{y} = 3108.1, \qquad s_2 = 205.9, \quad \text{(fusion)}$$

- **Question:** Is there enough evidence that fusion leads to higher tensile strength? (Test at $\alpha = 0.05$.)

- **Answer:**

Example 9.7 (Continued) [*]

Is normality (of each sample) plausible?



Example 9.7 (Continued) [*]

- Test of $H_0 : \mu_1 - \mu_2 = 0$ vs. $H_a : \mu_1 - \mu_2 < 0$ gives:

- A 95% CI for $\mu_1 - \mu_2$ is:

§9.3 Paired Samples

- Special situation consisting of n independently selected pairs:

$$(X_1, Y_1), \ldots, (X_n, Y_n), \qquad D_i = X_i - Y_i, \quad i = 1, \ldots, n$$

- The D_i are independent, but the pair members (X_i, Y_i) are not!
- Thus the two samples are NOT independent in the sense defined in §9.1. (The X and Y samples are **dependent**.)
- The connection between X_i and Y_i is that we usually have measurements on the same experimental unit, just under different conditions.
- **Exs:** measurements on a unit taken before and after some change has been introduced, or after some time has passed.
- **Inference:** proceeds by applying one-sample procedures from Ch. 7 to the differences $D_i = X_i - Y_i$.

Example 9.9 (7th Ed.)

The amount of time-arm elevation (secs) below $30°$ is recorded for a sample of $n = 16$ individuals before and after a change in work condition. Results are tabulated as follows:

Subject	1	2	3	...	16
Before (x)	81	87	86	...	75
After (y)	78	91	78	...	62
Difference (d)	3	-4	8	...	13

From this we compute the sample mean and std. dev. of the differences: $\bar{d} = 6.75$ and $s_D = 8.234$.

Question: is there evidence of a difference in the mean after the change in the work condition?

Example 9.9 (Continued) [*]

- **Answer 1:** Test $H_0 : \mu_D = 0$ vs. $H_a : \mu_D \neq 0$.

- **Answer 2:** Construct a 99% CI for μ_D.

§9.4 Two Population Proportions

The situation is as follows.

- $X \sim \text{Bin}(m, p_1)$ and $Y \sim \text{Bin}(n, p_2)$, with X and Y independent.
- Interested in making inference on the difference of (population) proportions: $p_1 - p_2 = 0$. (Note: we will not discuss the more difficult situation of $p_1 - p_2 = \Delta_0 \neq 0$.)
- Define the sample proportions:

$$\hat{p}_1 = \frac{X}{m}, \qquad \hat{p}_2 = \frac{Y}{n}, \qquad \hat{p} = \frac{X+Y}{m+n}$$

- Both m and n are large: $m\hat{p}_1 = X \geq 10$ and $n\hat{p}_2 = Y \geq 10$.

Two-Sample Tests & CIs For $p_1 - p_2$ [*]

- For testing the null

$$H_0 : \mu_1 - \mu_2 = 0$$

the test stat is:

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(1/m + 1/n)}} \quad \sim \quad \mathcal{N}(0,1), \text{ under } H_0$$

- The formula for a $(1-\alpha)100\%$ CI is:

Example 9.11 (7th Ed.)

Does the severity of a sentence differ for defendants who plead guilty and for those who plead NOT guilty? A sample of $n = 255$ trials yielded the following results:

		Plea	
		Guilty	NOT Guilty
	Sentenced	101 $(= x)$	56 $(= y)$
Result	Pardoned	90	8
	Totals	191 $(= m)$	64 $(= n)$

Let:

$$p_1 = \text{proportion of guilty pleas that are sentenced}$$
$$p_2 = \text{proportion of NOT guilty pleas that are sentenced}$$

Example 9.11 (Continued) [*]

- **Answer 1:** Test $H_0 : p_1 - p_2 = 0$ vs. $H_a : p_1 - p_2 \neq 0$.

- **Answer 2:** Construct a 95% CI for $p_1 - p_2$.