LECTURE NOTES

Applied Statistics and Probability for Engineers

Zelda Zabinsky
Richard Storch
Kailash Kapur
Robert Smith

Industrial Engineering Program
University of Washington

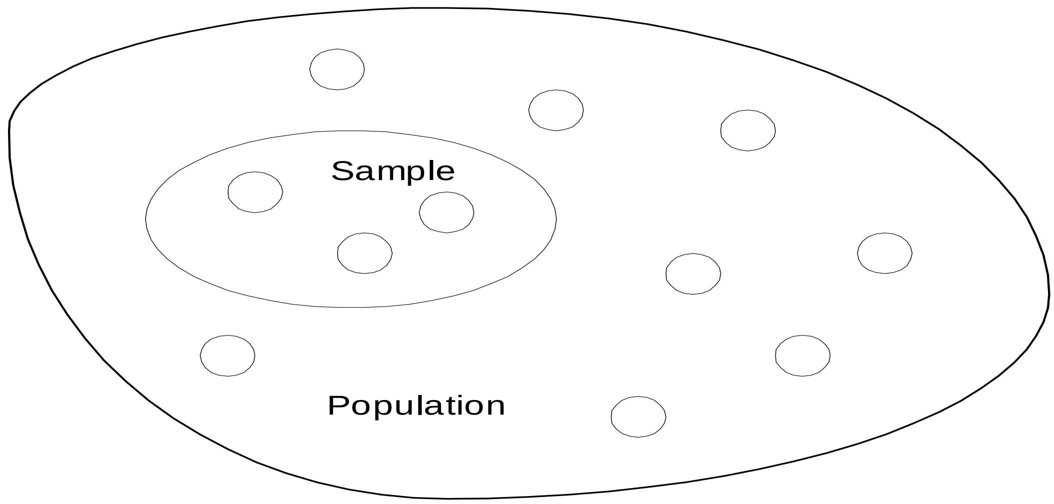
Chapter 1

Descriptive Statistics

Overview of Chapter 1

- Random Sampling
- Picturing the Distribution
- Sample Statistics

Probability and Statistics



Example: X_i = hours until failure of a particular lightbulb

Relative Frequency Histogram

Conveys a sense of how X is distributed.

Sample Mean

- Let $X_1, X_2, \ldots, X_n$ be a random sample from the distribution of X.
- The **sample mean** is a statistic and is given by

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

- The **population mean**, μ, is usually estimated by $\bar{X}$
- The **observed value** of the sample mean is the number given from the data,

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

Sample Median:

The sample median is found by arranging the random sample $X_1, X_2, \ldots, X_n$ in order, from smallest to largest, and finding the middle,

$$\tilde{x} = \begin{cases} x_{\frac{n+1}{2}} & \text{if } n \text{ is odd} \\ \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} & \text{if } n \text{ is even} \end{cases}$$

Sample Mode:

The sample mode of a random sample is the value that occurs most often.

Sample Variance and Sample Standard Deviation

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from the distribution of X. Then the **sample variance** is the statistic

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

which can be calculated by

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

The **sample standard deviation** is the statistic

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

The population variance, σ^2, and the population standard deviation σ are estimated by s^2 and s respectively.

Chapter 2

Probability

Overview of Chapter 2

- Interpreting Probabilities
- Sample Spaces and Events
- Permutations and Combinations
- Axioms of Probability
- Probability Rules
- Conditional Probability
- Independence and the Multiplication Rule
- Bayes Theorem

What is Probability?

- **Probability:** the assignment of a weight between 0 and 1 to indicate the likelihood of the occurence of an event.
- The **probability of an event** is defined in terms of an **experiment** and a **sample space**.
  - Consider an **experiment** that generates observations.
  - The **sample space** of an experiment, denoted S, is the set of all possible outcomes, or **sample points**.
  - An **event** is a subset of the sample space S.

Example: Coin Toss Experiment

- Consider an **experiment** that generates observations.
  - Example: Toss a fair coin 3 times in a row
- The **sample space** of an experiment, denoted S, is the set of all possible outcomes, or **sample points**.
  - Example: The sample space for this experiment has 8 sample points.

$$S = \left\{ \begin{array}{ll} HHH, & THH, \\ HHT, & THT, \\ HTH, & TTH, \\ HTT, & TTT \end{array} \right\}$$

Example: Continued

- An **event** is a subset of the sample space S.
  - Example: Look at 3 different events as examples.
  - The event of 3 heads,

$$A = \{HHH\}$$

  - The event of 2 heads,

$$B = \{HHT, HTH, THH\}$$

  - The event that the last toss is a head,

$$C = \{HHH, HTH, THH, TTH\}$$

Example: Continued

- The **probability** of an event indicates the likelihood the event occurs
  - Example: Look at the probability of the same 3 events.
  - The probability of getting 3 heads
    $P(A) = 1/8$
  - The probability of getting 2 heads
    $P(B) = 3/8$
  - The probability that the last toss is a head
    $P(C) = 4/8 = 1/2$

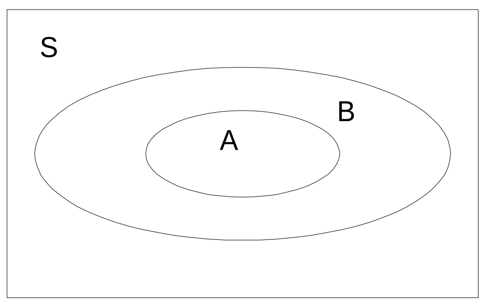
Review Sets:
Terminology & Notation

- A **set** is a well-defined collection of objects. Each object in a set is called an **element** of the set.
- The **universal set** S is the set of all objects under consideration. The **null set**, or **empty set**, $\emptyset$, contains no elements.
- Two sets are **equal** if they contain the same elements.
  - Ex: $A = \{1, 2, 3, 4\}$ equals $B = \{4, 3, 2, 1\}$

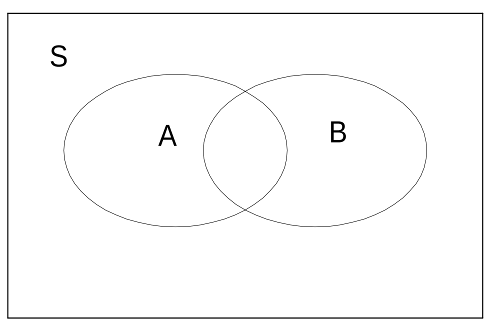
Review Set Relations

Suppose S is the universal set, with two subsets, A and B.

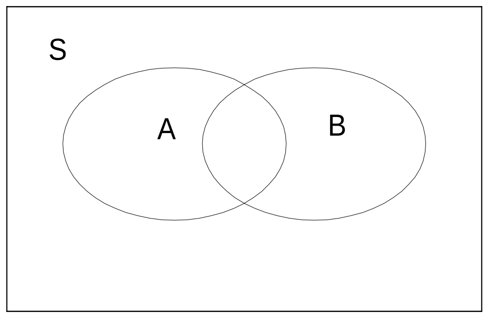
- A set, A, is a **subset** of B if all elements of A belong to B, $A \subset B$



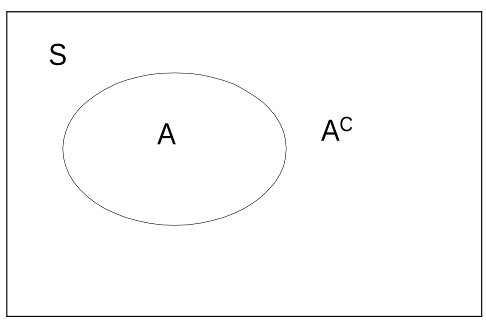
- The **union** $A \cup B = \{x \mid x \in A \text{ or } x \in B\}$



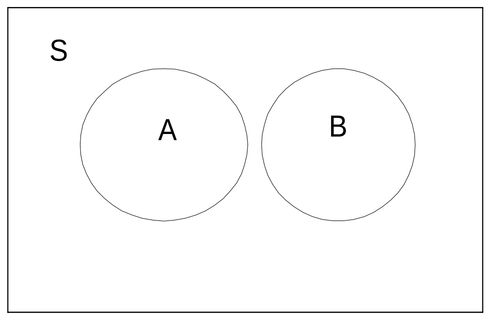
- The **intersection** $A \cap B = \{x \mid x \in A \text{ and } x \in B\}$



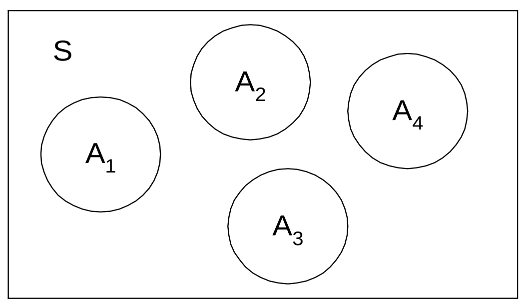
- The **complement** of A with respect to S is denoted A^C



- Sets A and B are **mutually exclusive** or **disjoint**, if and only if $A \cap B = \emptyset$.



- Any number of sets, $A_1, A_2, A_3, \ldots$ are **mutually exclusive** if and only if $A_i \cap A_j = \emptyset$ for $i \neq j$.



- $A_1 \cap A_2 = \emptyset$ $\quad A_2 \cap A_3 = \emptyset$
  $A_1 \cap A_3 = \emptyset$ $\quad A_2 \cap A_4 = \emptyset$
  $A_1 \cap A_4 = \emptyset$ $\quad A_3 \cap A_4 = \emptyset$

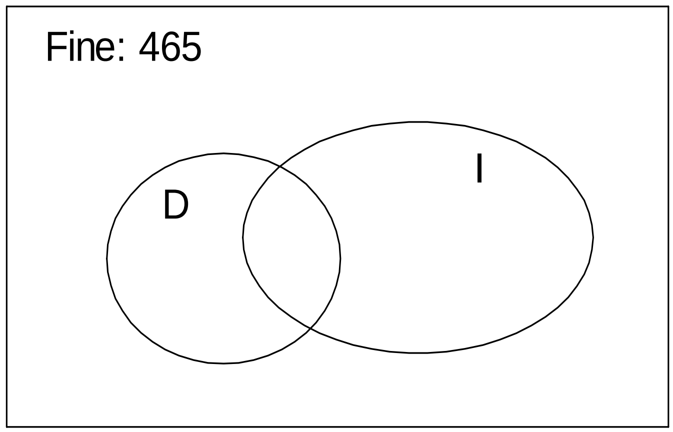
Venn Diagram Example

500 assembled machine parts are inspected before they are shipped. The inspection can detect whether:

- a part contains at least one defective component (D), and/or
- a part is improperly assembled (I)

Data:

- 15 parts contain at least one defective component
- 30 parts have been improperly assembled
- 10 parts contain at least one defective component <u>and</u> have been improperly assembled
- 465 parts are fine, i.e. have no defective components and are properly assembled



- Defective: 15 in D
- Improperly assembled: 30 in I
- Defective and Improperly assembled: 10 in $D \cap I$
- Defective and Properly assembled: 5 in $D \cap I^C$
- Improperly assembled and Not Defective: 20 in $I \cap D^C$
- Defective or Improperly assembled: 35 in $D \cup I$

Descriptive Statistics:

- What is the probability that 1 part selected at random is fine?
  $P\,(\text{a part is fine}) = 465/500 = 0.93$
  The percentage of parts that are fine is 93%.
- How many parts are rejected for shipment?
  The parts that are rejected are parts that are either defective or improperly assembled, $D \cup I = F^C$, so 35 parts are rejected for shipment.

Counting Rules

Permutation: A permutation is an arrangement of objects in a definite order.

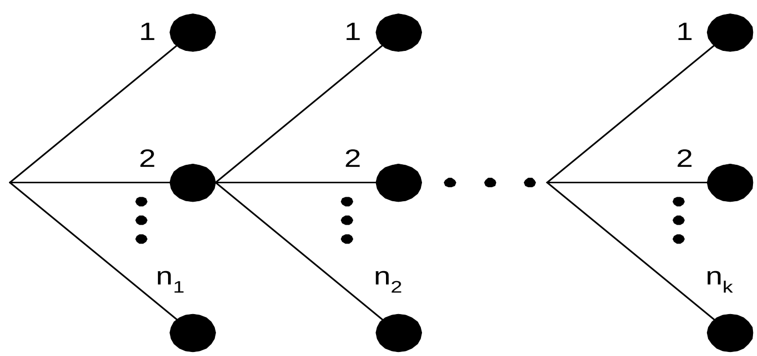
Combination: A combination is a selection of object with no regard to order.

Multiplication Principle

Consider an experiment taking place in k stages. Let n_i denote the number of ways in which stage i can occur, for $i = 1, 2, \ldots, k$.
Altogether the experiment can occur in

$$\prod_{i=1}^{k} n_i = n_1 * n_2 * \cdots * n_k \quad \text{ways.}$$

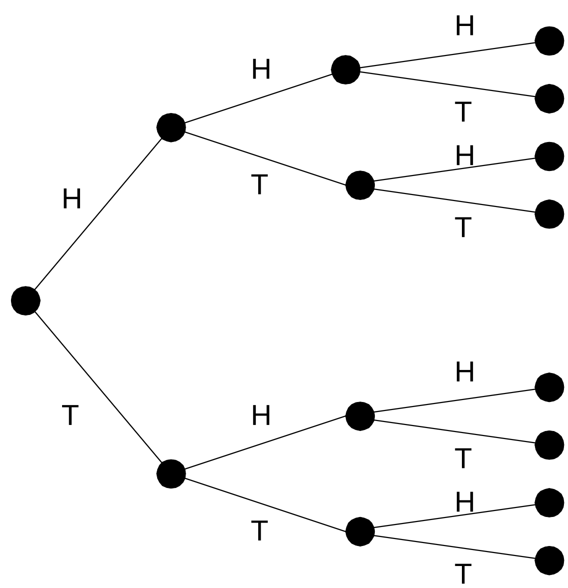
Illustrate with tree diagram, or slots



$$\frac{n_1}{\text{Stage 1}} \quad \frac{n_2}{\text{Stage 2}} \quad \cdots \quad \frac{n_k}{\text{Stage k}}$$

<u>Example</u>

Toss a coin 3 times in a row



H/T	H/T	H/T
2	2	2
$\overline{\text{1st}}$	$\overline{\text{2nd}}$	$\overline{\text{3rd}}$
toss	toss	toss

$2^3 = 8$

Permutations

Suppose we have N distinct objects, and we are going to arrange, in some order, n of them. How many permutations are there?
Since we are arranging n of them, we need n slots. Also, the objects are distinct, and repetition is NOT allowed.

$$\underset{\text{1st object}}{\underline{\quad N \quad}} \quad \underset{\text{2nd object}}{\underline{\quad (N-1) \quad}} \quad \cdots \quad \underset{n\text{th object}}{\underline{(N-n+1)}}$$

$$P_n^N = N(N-1)(N-2)\ldots(N-n+1) = \frac{N!}{(N-n)!}$$

Combinations

Suppose we have N distinct objects, and we select n objects from them. (order does NOT matter). How many combinations?

$$C_n^N = \begin{pmatrix} N \\ n \end{pmatrix} = \frac{N!}{n!(N-n)!}$$

<u>Poker Hands</u>

52 cards in a deck, 4 suits, 13 cards in a suit, $\heartsuit, \diamondsuit, \clubsuit, \spadesuit$

What is the probability of being dealt 3 aces and 2 kings?

What is the probability of being dealt a run of 5 cards in sequence (A,K,Q,J,10)

Axioms of Probability

1. Let S denote a sample space for an experiment,

$$P(S) = 1$$

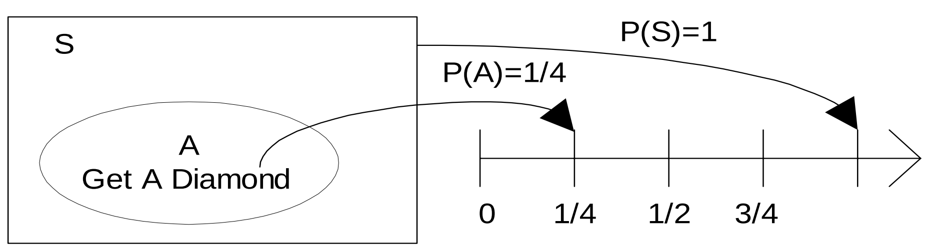
2. For every event A,

$$P(A) \geq 0$$

3. Let $A_1, A_2, \ldots, A_n, \ldots$ be a finite or infinite sequence of mutually exclusive events. Then $P(A_1 \cup A_2 \cup A_3 \cdots) =$

$$P(A_1) + P(A_2) + P(A_3) + \cdots$$

Example: Draw a card from a deck



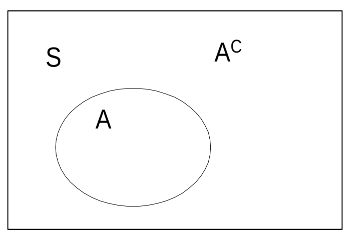
Rules of Probability

- The probability of impossible events is 0:

$$P\left(\emptyset\right) = 0$$

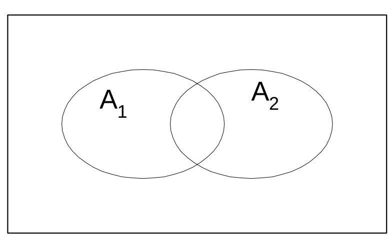
- Complement rule:

$$P\left(A^C\right) = 1 - P\left(A\right)$$



- Addition rule:

$$P\left(A_1 \cup A_2\right) = P\left(A_1\right) + P\left(A_2\right) - P\left(A_1 \cap A_2\right)$$



More Rules of Probability

Definition: Let A and B be events with $P(A) \neq 0$. The conditional probability of B given A is

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

Note:
$P(B|A)$ is undefined if $P(A) = 0$.

Multiplication Rule:

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

Independence

Definition: Two events A and B are **independent** if and only if

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

Otherwise they are **dependent**.

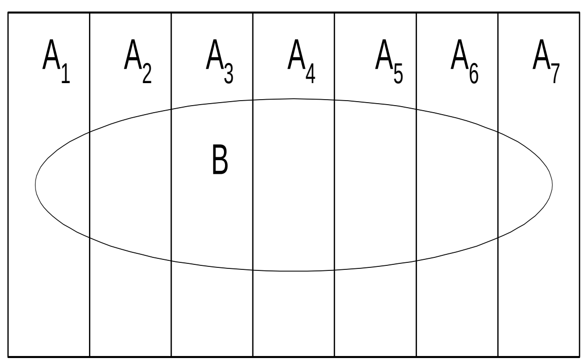
Theorem: Two events A and B are independent if and only if

$$\begin{aligned} & P(B|A) = P(B) \quad \text{if } P(A) \neq 0 \\ \text{and} \quad & P(A|B) = P(A) \quad \text{if } P(B) \neq 0 \end{aligned}$$

Bayes' Theorem

Let $A_1, A_2, \ldots, A_n$ be a collection of events which partition S. Let B be an event, $P(B) \neq 0$. Then, for any event A_j, $j = 1, 2 \ldots, n$,

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{P(B|A_1)P(A_1) + \ldots + P(B|A_n)P(A_n)}$$



Bayes' Theorem: Tree Diagram

Let $A_1, A_2, \ldots, A_n$ be a collection of events which partition S. These events form the first branches of the tree. Let B be an event, $P(B) \neq 0$. The events B and B^C are leaves off of each A_j branch.

Chapter 3:

Discrete Random Variables

Overview of Chapter 3

Definition and Properties:

- Discrete random variables
- Discrete probability distributions
- Expected value, mean, variance, standard deviation

Discrete Probability Distributions:

- Uniform distribution
- Geometric distribution
- Binomial distribution
- Negative binomial distribution
- Hypergeometric distribution
- Poisson distribution

Discrete Random Variables

For a given sample space S of some experiment, a **random variable** is any rule that associates a number with each outcome in S.

Example: Age
$S = \{19, 20, 21, 22, \ldots, 29, 30^+\}$
Y = age

Example: Coin Toss three times
S = {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}
Y = number of heads

A random variable is **discrete** if its set of possible values is a discrete set, i.e. has a finite, or countably infinite, number of elements.

Example: Make ball bearings until one works
S - success, F - failure
S = {S, FS, FFS, FFFS, ...}
Y = number of ball bearings made

Discrete Probability Distribution

The **density function** for a discrete random variable Y is a function p given by

$$p(y) = P(Y = y).$$

The **cumulative distribution function** F(y), for a discrete random variable Y with density p, is defined by

$$F(y) = P(Y \leq y) = \sum_{x \leq y} p(x).$$

Ball Bearing Example

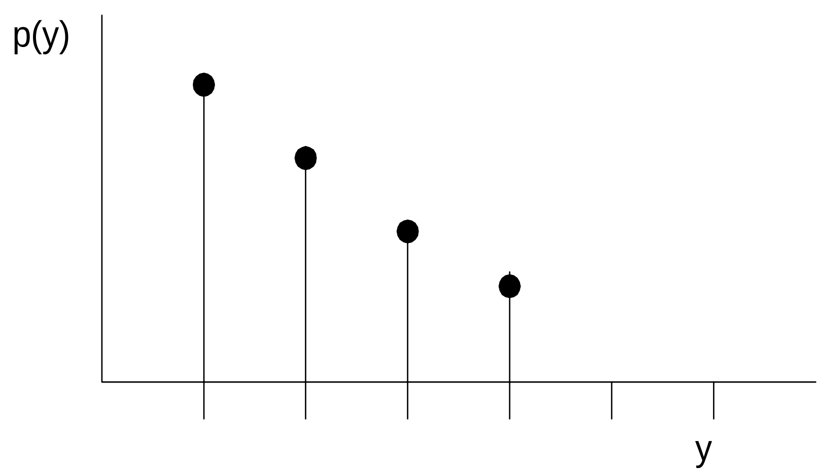
- Make ball bearings until one is good (successful)
- $Y = \#$ of ball bearings made
- Let $p =$ probability a ball bearing is successful. In this example, $p = 0.6$.
- Assume the ball bearings are independent.

$$\begin{aligned}
P(Y=1) &= p(1) = p \\
P(Y=2) &= p(2) = (1-p)p \\
P(Y=3) &= p(3) = (1-p)^2 p \\
&\vdots \\
P(Y=y) &= p(y) = (1-p)^{y-1} p
\end{aligned}$$

Ball Bearing Example continued

y	density $p(y) = P(Y = y)$	cumulative distribution $F(y) = P(Y \leq y)$
1	0.6	0.6
2	0.24	0.84
3	0.096	0.936
4	0.0384	0.9744
$\vdots$	$\vdots$	$\vdots$
y	$(0.4)^{y-1}(0.6)$	$\Sigma_{x=1}^{y}(0.4)^{x-1}(0.6)$
$\vdots$	$\vdots$	$\vdots$

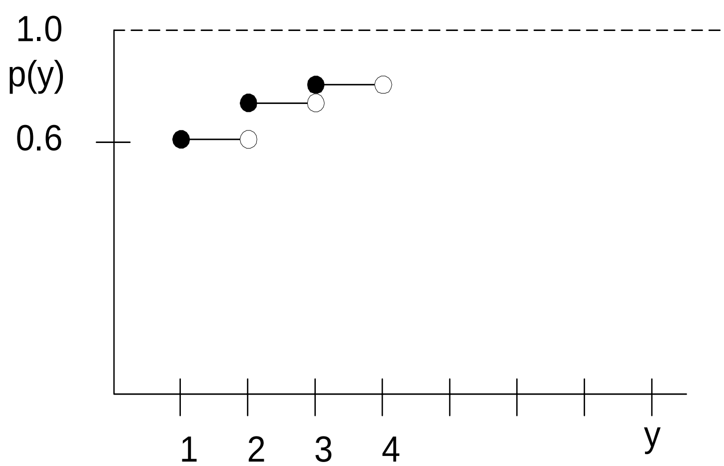
Density Distribution Function $p(y)$



Ball Bearing Example continued

y	density $p(y) = P(Y = y)$	cumulative distribution $F(y) = P(Y \le y)$
1	0.6	0.6
2	0.24	0.84
3	0.096	0.936
4	0.0384	0.9744
⋮	⋮	⋮
y	$(0.4)^{y-1}(0.6)$	$\Sigma_{x=1}^{y}(0.4)^{x-1}(0.6)$
⋮	⋮	⋮

Cumulative Distribution Function $F(Y)$



Characteristics of the Cumulative Distribution Function, $F(y)$:

- $F(Y)$ is defined for all real values, not just outcomes.
- $F(Y)$ is non-decreasing (increasing or flat)
- $\lim_{y \to \infty} F(Y) = 1$
- $\lim_{y \to -\infty} F(Y) = 0$

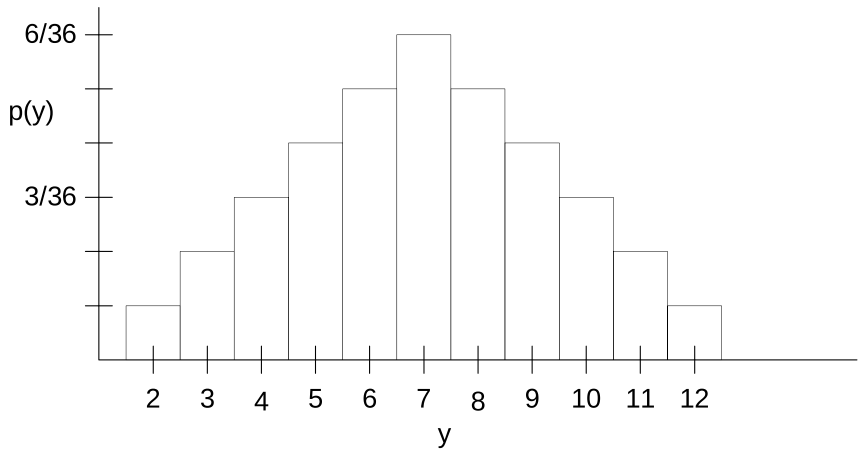
Dice Example

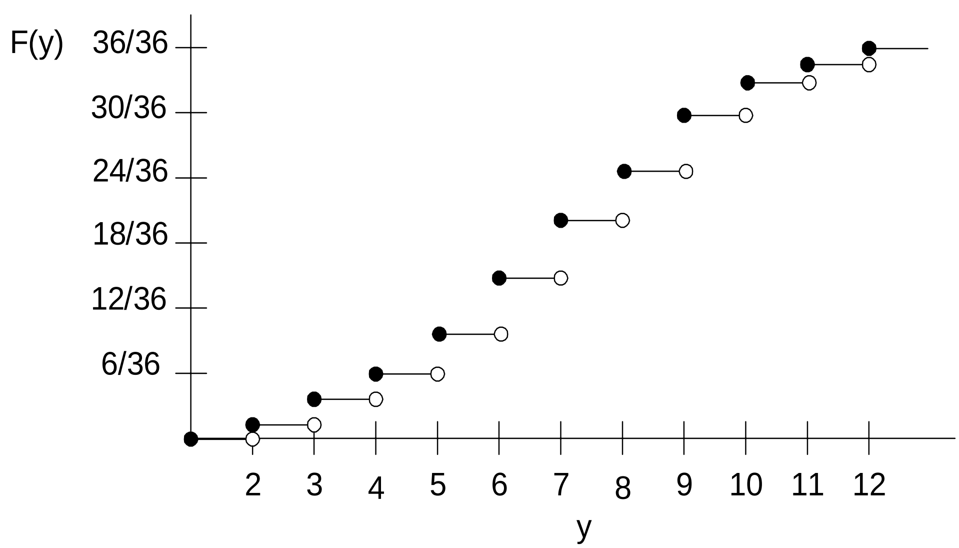
Roll a pair of dice
Y = total showing on the pair of dice

Y	density $p(y) = P(Y = y)$	cum. dist. $F(y) = P(Y \leq y)$
2	1/36	1/36
3	2/36	3/36
4	3/36	6/36
5	4/36	10/36
6	5/36	15/36
7	6/36	21/36
8	5/36	26/36
9	4/36	30/36
10	3/36	33/36
11	2/36	35/36
12	1/36	36/36

Dice Example continued

Density $p(y) = P(Y = y)$





Cumulative Distribution $F(Y) = P(Y \leq y)$

The Expected Value, or Mean, of Y

Let Y be a discrete random variable, with set of possible values D, and with density function $p(y) = P(Y = y)$.
The expected value of Y is:

$$E[Y] = \sum_{y \in D} y \cdot p(y) = \sum_{y \in D} y \cdot P(Y = y)$$

It is also called the mean, and written μ or μ_y.

Example: Y = total showing on a pair of dice

$$\begin{aligned} E[Y] &= 2(\tfrac{1}{36}) + 3(\tfrac{2}{36}) + 4(\tfrac{3}{36}) + 5(\tfrac{4}{36}) + \\ &\quad 6(\tfrac{5}{36}) + 7(\tfrac{6}{36}) + 8(\tfrac{5}{36}) + 9(\tfrac{4}{36}) + \\ &\quad 10(\tfrac{3}{36}) + 11(\tfrac{2}{36}) + 12(\tfrac{1}{36}) \\ &= 7 \end{aligned}$$

Ball Bearing Example (Geometric Distribution)

$p =$ probability a ball bearing is successful

$$
\begin{aligned}
P(Y=1) &= p \\
P(Y=2) &= (1-p)p \\
P(Y=3) &= (1-p)^2 p \\
&\vdots \\
P(Y=y) &= (1-p)^{y-1} p
\end{aligned}
$$

$$E[Y] = \sum_{y=1}^{\infty} y \cdot (1-p)^{y-1} p = \frac{1}{p}$$

Proof of the Geometric $E[Y]$

To evaluate the infinite series, notice that

$$\frac{d((1-p)^y)}{dp} = -y(1-p)^{y-1}$$

also recall that $\Sigma_{y=0}^{\infty} r^y = \frac{1}{1-r}$ for $0 \le r < 1$. Therefore, we have

$$\begin{aligned} E[Y] &= \sum_{y=1}^{\infty} y \cdot (1-p)^{y-1} p \\ &= \sum_{y=1}^{\infty} p \frac{-d((1-p)^y)}{dp} = -p \sum_{y=1}^{\infty} \frac{d((1-p)^y)}{dp} \end{aligned}$$

Interchanging the sum and the derivative

$$= -p \frac{d\left(\Sigma_{y=1}^{\infty}(1-p)^y\right)}{dp} = -p \, \frac{d\left(\frac{1}{1-(1-p)} - 1\right)}{dp}$$

$$= -p \frac{d\left(\frac{1-p}{p}\right)}{dp} = -p\left(\frac{-p-(1-p)}{p^2}\right) = -p\left(\frac{-1}{p^2}\right)$$

$$= \frac{1}{p}$$

The Expected Value of a Function

Let Y be a discrete random variable, with set of possible values D, and with density function $p(y) = P(Y = y)$.

Then the expected value of any function $H(Y)$, denoted $E[H(Y)]$, or $\mu_{H(Y)}$, is

$$E[H(Y)] = \sum_{y \in D} H(y) \cdot P(Y = y)$$

Gambling Game Example

Example: How much would you be willing to pay in order to play the following gambling game?

Game: Roll a pair of dice. If you get an even number, win \$2. If you get a 7 or an 11, win \$1.

$$H(y) = \begin{cases} \$2 \ \text{ if } y = 2,\ 4,\ 6,\ 8,\ 10,\ 12 \\ \$1 \ \text{ if } y = 7,\ 11 \\ \$0 \ \text{ otherwise} \end{cases}$$

$$\begin{aligned} E\left[H(Y)\right] &= \sum_{y=2}^{12} H(y)P\left(Y=y\right) \\ &= 2(\frac{1}{36}) + 2(\frac{3}{36}) + 2(\frac{5}{36}) + 1(\frac{6}{36}) + \\ &\quad 2(\frac{5}{36}) + 2(\frac{3}{36}) + 1(\frac{2}{36}) + 2(\frac{1}{36}) \\ &= \frac{44}{36} = 1.22 \end{aligned}$$

Property of Expected Value

If a and b are constants, then

$$E[aY+b] = aE[Y]+b$$

Proof:

$$\begin{aligned} E[aY+b] &= \sum_{y\in D}(ay+b)\cdot P(Y=y) \\ &= a\underbrace{\sum_{y\in D} y\cdot P(Y=y)}_{=E[Y]} + b\underbrace{\sum_{y\in D} P(Y=y)}_{=1} \\ &= aE[Y]+b \end{aligned}$$

Example: Suppose Y is the number of days until a machine fails, and $E[Y]=4$. Also, the cost of using the machine is \$10/day plus \$2 each time it fails. What is the cost of using the machine until it fails?

$$H(Y) = 10Y+2$$

$$E[H(Y)] = 10E[Y]+2 = \$42$$

More Properties of Expectation

- $E[aY + b] = aE[Y] + b$ $\quad$ a,b constants

- $E[cY] = cE[Y]$ $\quad$ c constant

- $E[c] = c$

- $E[Y + X] = E[Y] + E[X]$

- $E[g(Y) + h(Y)] = E[g(Y)] + E[h(Y)]$

- Notice: $E[h(Y)] \neq h(E[Y])$ for general functions.

- For independent random variables X, Y, $E[Y \cdot X] = E[Y] \cdot E[X]$

<u>The Variance of Y</u>

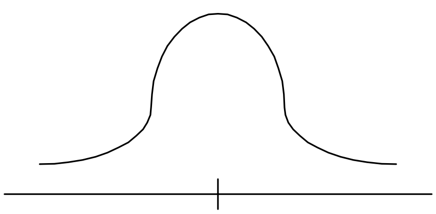
Let Y be a discrete random variable with density $p(y)$ and $E[Y] = \mu$. The variance of Y, denoted $Var(Y)$, or σ_Y^2, or σ^2 is

$$\begin{aligned} Var(Y) &= \sum_{y \in D} (y-\mu)^2 p(y) \\ &= E\left[(Y-\mu)^2\right] \end{aligned}$$

The standard deviation of Y, σ_Y or σ_Y, equals the square root of the variance of Y

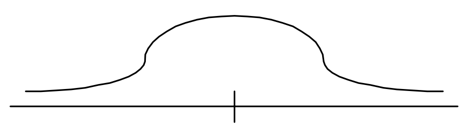
$$\sigma = \sqrt{\sigma^2}$$

Small σ^2 or σ



Most values close to mean

Large σ^2 or σ



Most values far from mean

Variance Examples

Example: Y = total showing on a pair of dice.

$E[Y] = 7$

$$Var(Y) = (2-7)^2 \tfrac{1}{36} + (3-7)^2 \tfrac{2}{36} + \cdots + (12-7)^2 \tfrac{1}{36} \approx 5.833$$

$\sigma \approx 2.415$

Example: Y = # of ball bearings made, stop when one is successful

p = probability a ball bearing is successful

$P(Y = y) = (1-p)^{y-1} p \quad$ for $y = 1, 2, \ldots$

$E[Y] = \frac{1}{p}$

$Var(Y) = \sigma^2 = \frac{1-p}{p^2}$

Geometric distribution with parameter p.

Special Expectations

Moments about the origin:

The expected value of Y^k is called the k^{th} moment about the origin of Y,

$$E\left[Y^k\right] = \Sigma_{y \in D}\, y^k p(y)$$

when $k = 1$, this is simply the mean,

$$\mu = E\left[Y\right]$$

Moments about the mean:

The function $H(Y) = (Y - \mu)^k$ is called the k^{th} moment about the mean of Y

$$E\left[(Y - \mu)^k\right] = \Sigma_{y \in D}(y - \mu)^k p(y)$$

when $k = 2$, this is the 2^{nd} moment about the mean, or the variance.

Moment Generating Function

$$m_Y(t) = E\left[e^{tY}\right]$$

Rules for Variance

- $Var(c) = 0$ c constant

- $Var(cY) = c^2 Var(Y)$

- $Var(aY + b) = a^2 Var(Y)$ a,b constants

- $Var(Y) = E\left[(Y-\mu)^2\right] = E\left[Y^2\right] - \mu^2$

$$\begin{aligned} E\left[(Y-\mu)^2\right] &= E\left[Y^2 - 2\mu Y + \mu^2\right] \\ &= E\left[Y^2\right] - 2\mu E[Y] + \mu^2 \\ &= E\left[Y^2\right] - \mu^2 \end{aligned}$$

- If Y and X are independent,
  then $Var(Y + X) = Var(Y) + Var(X)$

Discrete Distributions

Uniform Discrete Distribution:

If Y assumes the values $y_1, y_2, \ldots, y_n$ with equal probability, then it has a discrete uniform distribution

$$p(y) = \frac{1}{n} \qquad \text{for } y = y_1, y_2, \ldots, y_n$$

$$E[Y] = \frac{\sum_{i=1}^{n} y_i}{n}$$

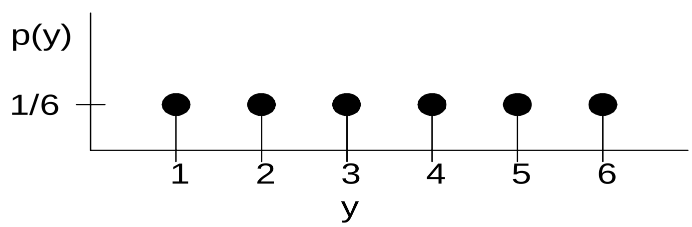
$$Var(Y) = \frac{\sum_{i=1}^{n} (y_i - \mu)^2}{n}$$

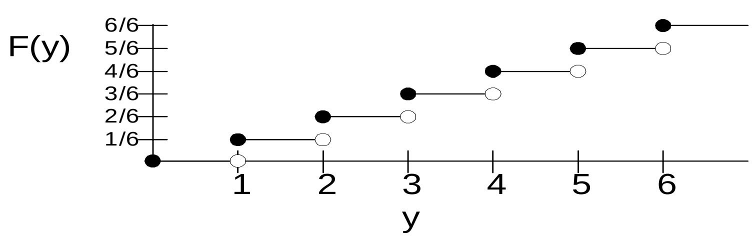
Example: Uniform Discrete Distrib.

Example: Tossing a single die.
Y = number showing
$p(y) = 1/6$ for $y = 1, 2, \cdots, 6$

$$E[Y] = \frac{1+2+3+4+5+6}{6} = 3.5$$

$$Var(Y) = \frac{(1-3.5)^2+\cdots+(6-3.5)^2}{6} = \frac{35}{12} = 2.9$$





Geometric Distribution

A geometric experiment has 3 properties:

- The experiment consists of repeated trials, each characterized as either a success (s) or a failure (f).
- The trials are identical & independent of each other; each has the same probability of success, p, with $0 < p < 1$.
- The random variable Y denotes the number of trials needed to obtain the first success.

Geometric Distribution continued

For $Y \sim \text{Geometric}(p)$

$$p(y) = P\left(Y = y\right) = (1-p)^{y-1}p \quad y = 1, 2, \ldots$$

$$F\left(Y\right) = P\left(Y \leq y\right) = \sum_{x=1}^{y}(1-p)^{x-1}p$$

$$E\left[Y\right] = \frac{1}{p}$$

$$Var\left(Y\right) = \frac{q}{p^2} \quad \text{where } q = 1-p$$

Binomial Distribution

A binomial experiment has 4 properties:

- The experiment consists of n repeated trials.
- Each trial has two outcomes, Success or Failure. The trials are independent.
- The probability of success for a trial is p, $0 < p < 1$, and is identical for all trials.
- A binomial random variable Y is the number of successes in n trials of a binomial experiment.

Binomial Distribution continued

For $Y \sim \text{Binomial}(n, p)$

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

$$F(Y) = \sum_{x=0}^{[y]} \binom{n}{x} p^x (1-p)^{n-x}$$

where $[y]$ means the greatest integer less than or equal to y.

$$E[Y] = np$$

$$Var(Y) = np(1-p)$$

Negative Binomial Distribution

- Same as a binomial experiment, but instead of counting the number of successes in n trials, we are interested in
- $Y = \#$ of trials needed to obtain r successes

$$\begin{aligned} P(Y = y) &= \text{prob. that the r}^{th}\text{ success occurs on} \\ &\quad \text{the } y^{th} \text{ trial, for } y \geq r, \\ &= \binom{y-1}{r-1} (1-p)^{y-r} p^r \end{aligned}$$

and r could be $1, 2, 3, \ldots$

$$E[Y] = \frac{r}{p}$$

$$Var(Y) = \frac{r(1-p)}{p^2}$$

Note: in the special case when $r = 1$, this reduces to the geometric distribution.

Hypergeometric Distribution

- A random sample of size n is chosen from N items (the population), $n \leq N$, sampled without replacement.
- Each item is either a success or a failure, and there are r successes in the population.
- Y is the number of succeses in the sample.

Hypergeometric Distribution cont.

$Y = \#$ of successes in the sample

$$P(Y=y) = \frac{\begin{array}{c}\text{number of ways to select } y\\ \text{successes and } n-y \text{ failures}\\ \text{from the population}\end{array}}{\begin{array}{c}\text{number of possible outcomes}\\ \text{in the experiment}\end{array}}$$

$$= \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}}$$

where $\max(0, n-(N-r)) \le y \le \min(n, r)$

$E[Y] = n\left(\frac{r}{N}\right)$

$Var(Y) = n\left(\frac{r}{N}\right)\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right)$

Poisson Distribution

The Poisson distribution is based on counting events that occur during a time interval where:

- The probability of 2 events occuring in a very short time interval is negligible.
- The probability of a single event occuring in a short time interval is proportional to the length of the interval.
- The number of events occuring in one time interval are independent of those occuring in a disjoint time interval.
- Y is the number of occurences of the event in a given time interval.

Poisson Distribution continued

Y = number of events in a time interval,
λ = expected # of events in an interval
(mean arrival rate) and $\lambda > 0$.

For $Y \sim$ Poisson(λ)

$$P(Y = y) = \frac{e^{-\lambda}\lambda^{y}}{y!} \qquad y = 0, 1, 2, \ldots$$

$$E[Y] = \lambda$$

$$Var(Y) = \lambda$$

Relationships

- Poisson Dist. is a limiting case of the Binomial Dist, as $n \to \infty$, $p \to 0$ in such a way that $\lambda = np$.
- Binomial Dist. is a special case of the Multinomial Dist.
- Geometric Dist. is a special case of the Negative Binomial Dist.
- Compare: Binomial with Hypergeometric (sampling with vs. without replacement)

Chapter 4:

Continuous Random Variables

Overview of Chapter 4

Definition & Properties:

- Continuous densities
- Expectation and distribution parameters

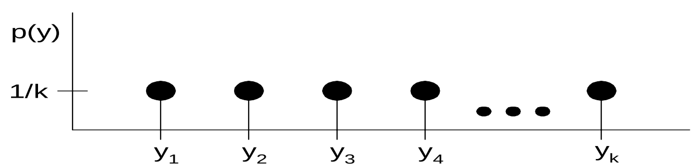
Continuous Probability Distributions:

- Uniform distribution
- Normal distribution
- Normal Probability Rule and Chebyshev's Inequality
- Approximations
- Gamma distribution
- Weibull distribution

Uniform Distribution: from discrete to continuous

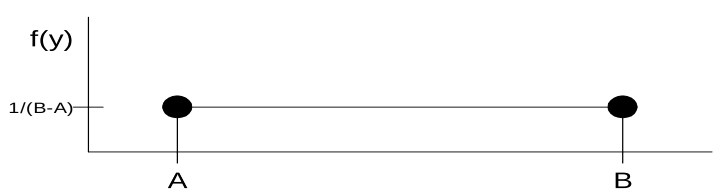
Discrete Uniform Distribution

Y assumes values $y_1, y_2, \ldots, y_k$
with equal probability
$P(Y = y) = \frac{1}{k}$ for $y = y_1, y_2, \ldots, y_k$



Continuous Uniform Distribution

Y assumes values in interval [A,B]
with equal probability.
$f(y) = \frac{1}{B-A}$ if $A \le y \le B$

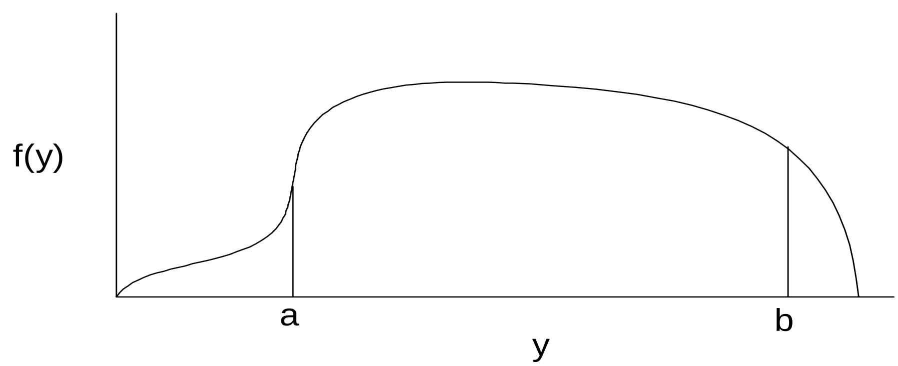


Continuous Random Variables

Let Y be a continuous random variable. A probability density function (pdf) of Y is a function $f(y)$ such that, for any two numbers a and b with $a \leq b$,

$$P(a \leq Y \leq b) = \int_a^b f(y)dy$$

Note: $\int_{-\infty}^{+\infty} f(y)dy = 1$ and $0 \leq f(y)$ for all y



Continuous Random Variables cont.

If Y is a continuous random variable, then $P(Y = c) = 0$ for any number c. Also, for any 2 numbers a, b with $a < b$, then

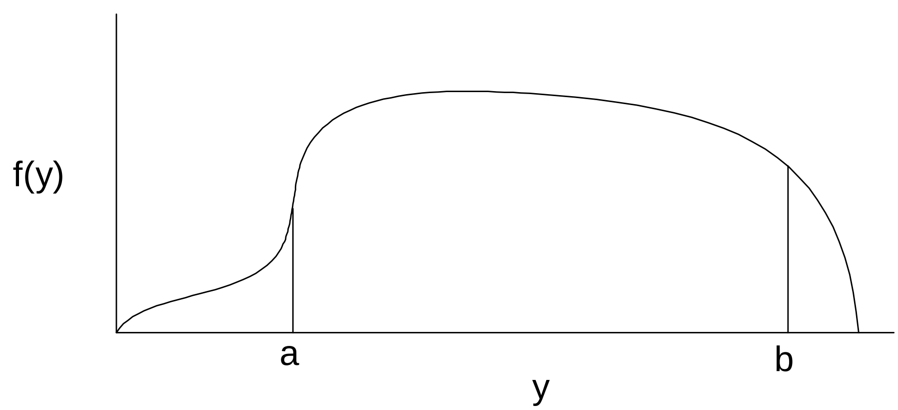
$$\begin{aligned} P(a \leq Y \leq b) &= P(a < Y \leq b) \\ &= P(a \leq Y < b) \\ &= P(a < Y < b) \end{aligned}$$

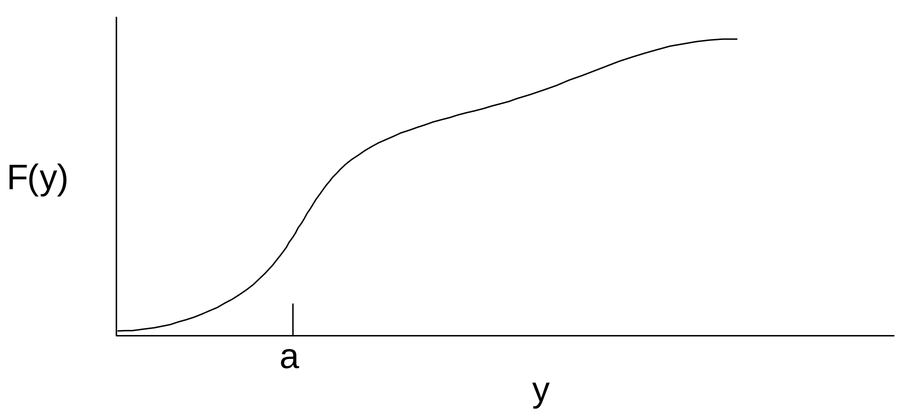
Notice this was not true for a discrete random variable.

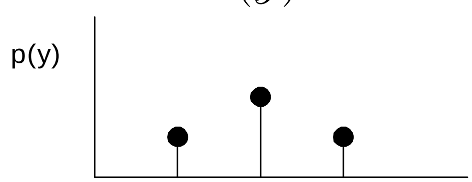
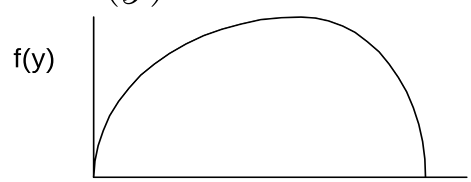
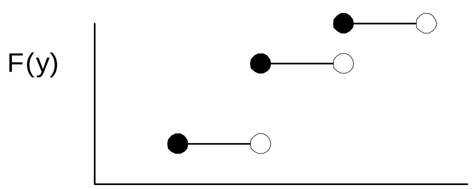
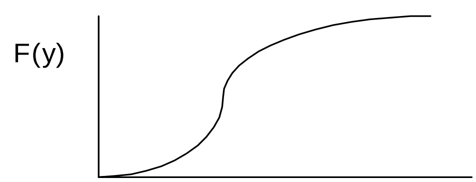
Cumulative Distribution Function

Let Y be a continuous random variable. The cumulative distribution function $F(y)$ is defined for every number y by

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





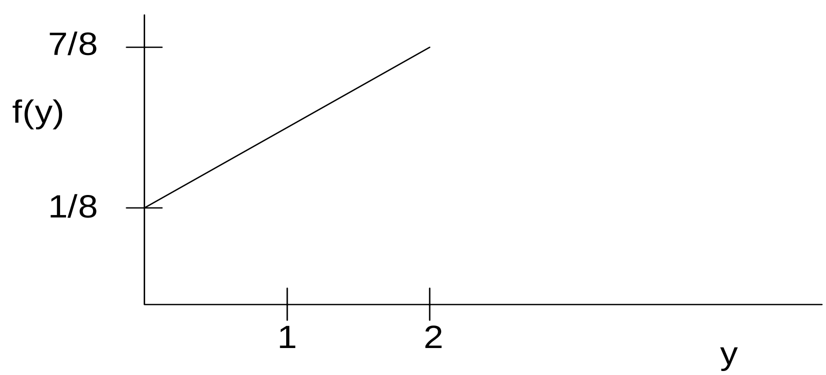
Discrete	Continuous
density function $p(y) = P(Y = y)$	density function $f(y)$
Cum. Dist. Fn. $F(y)$	Cum. Dist. Fn. $F(y)$
	
	
$F(y)$ is the sum of $p(y)$	$F(y)$ is the integral of $f(y)$
$p(y)$ is the difference between two $F(y)$ values	$f(y)$ is the derivative of $F(y)$; $f(y) = F'(y)$

Discrete	Continuous
$F(y) = \sum_{-\infty}^{y} p(x)$	$F(y) = \int_{-\infty}^{y} f(x)dx$
$E[Y] = \sum_{-\infty}^{+\infty} yp(y)$	$E[Y] = \int_{-\infty}^{+\infty} yf(y)dy$
$E[H(Y)] = \sum_{-\infty}^{+\infty} H(y)p(y)$	$E[H(Y)] = \int_{-\infty}^{+\infty} H(y)f(y)dy$
$Var(Y) = \sum_{-\infty}^{+\infty} (y-\mu)^2 p(y)$	$Var(Y) = \int_{-\infty}^{+\infty} (y-\mu)^2 f(y)dy$
$= E\left[(Y-\mu)^2\right]$	$= E\left[(Y-\mu)^2\right]$
$= E\left[Y^2\right] - \mu^2$	$= E\left[Y^2\right] - \mu^2$
$\sigma = \sqrt{Var(Y)}$	$\sigma = \sqrt{Var(Y)}$

Continuous Distributions Example

Suppose the density, $f(y)$, of the magnitude Y of a dynamic load on a bridge (in newtons) is given by

$$f(y) = \begin{cases} \frac{1}{8} + \frac{3}{8}y & \text{if } 0 \leq y \leq 2 \\ 0 & \text{otherwise} \end{cases}$$



Find the probability the load exceeds 1, and find the expected load.

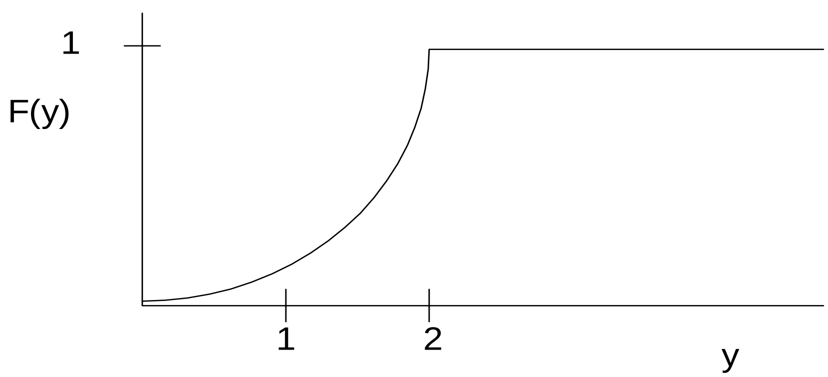
$$P(Y > 1) = 1 - P(Y \leq 1) = 1 - F(1)$$

$$E[Y] = \int_0^2 y f(y) dy$$

Continuous Dist. Example Cont.

Determine $F(y)$:

$$
\begin{aligned}
F(y) &= P(Y \le y) = \textstyle\int_{-\infty}^{y} f(x)dx \\
&= \textstyle\int_0^y (\frac{1}{8} + \frac{3}{8}x)dx = \frac{1}{8}x + \frac{3}{8 \cdot 2}x^2 \Big|_0^y \\
&= \frac{y}{8} + \frac{3}{16}y^2 \text{ for } 0 \le y \le 2
\end{aligned}
$$



Continuous Dist. Example Cont.

$$
\begin{aligned}
P(Y > 1) &= \text{prob. the load exceeds 1} \\
&= 1 - F(1) = 1 - (\frac{1}{8} + \frac{3}{16}) \\
&= \frac{11}{16} = 0.688
\end{aligned}
$$

$$
\begin{aligned}
E[Y] = \mu &= \int_0^2 y f(y) dy = \int_0^2 y \left(\frac{1}{8} + \frac{3}{8} y \right) dy \\
&= \frac{1}{8 \cdot 2} y^2 + \frac{3}{8 \cdot 3} y^3 \bigg|_0^2 \\
&= \frac{4}{16} + \frac{8}{8} = 1\frac{1}{4}
\end{aligned}
$$

Normal Distribution

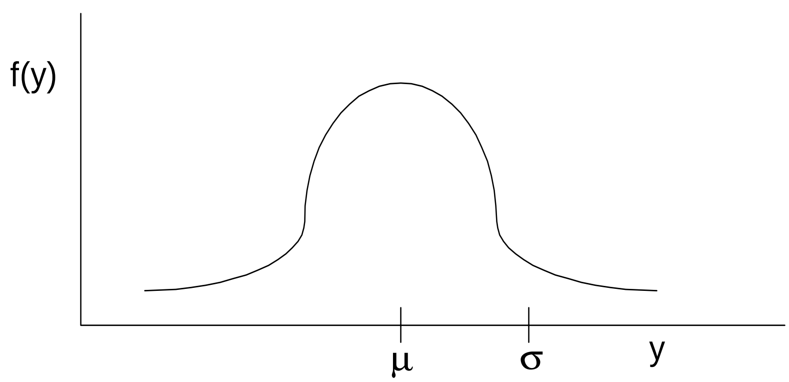
A continuous random variable Y has a normal distribution with parameters μ and σ (or σ^2), written $Y \sim N(\mu, \sigma)$, where $-\infty < \mu < +\infty$ and $0 < \sigma < +\infty$, if the probability density function is:

$$f(y; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-(y-\mu)^2/2\sigma^2} \quad -\infty < y < +\infty$$

Then,

$$E[Y] = \mu, \qquad Var(Y) = \sigma^2 \quad \text{and}$$

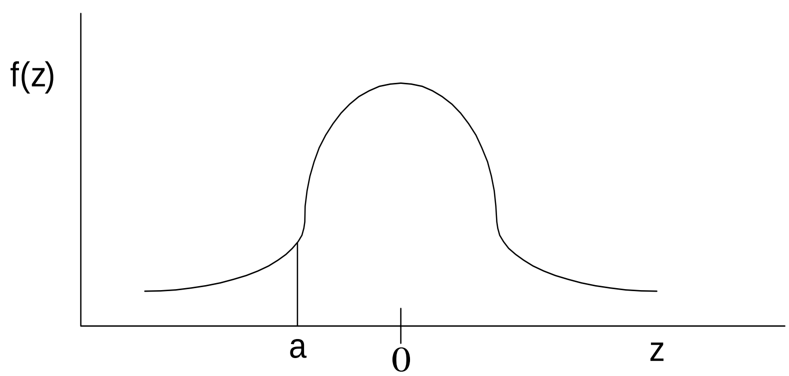
$F(y)$ is given in tables, or evaluated numerically.



Standard Normal Distribution

$Z \sim N(0,1)$ has a standard normal distribution ($\mu = 0$, $\sigma = 1$)

$$f(z) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-z^2/2} \quad -\infty < z < +\infty$$
$$F(z) = P(Z \le z) = \Phi(z)$$

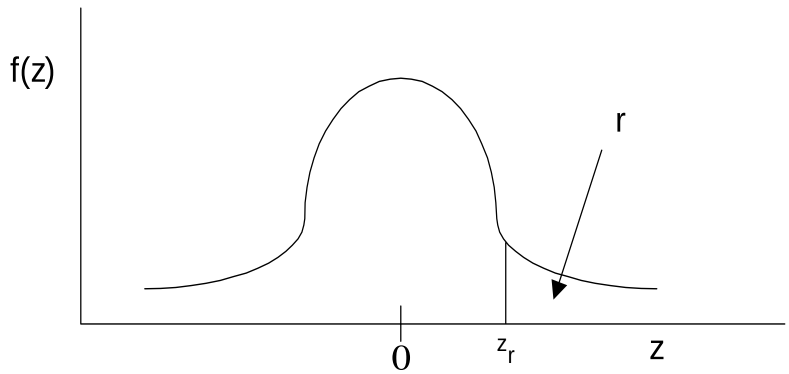


$\frac{Y-\mu}{\sigma}$ has a standard normal distribution; if $Y \sim N(\mu, \sigma)$, then $\frac{Y-\mu}{\sigma} \sim N(0,1)$

z_r Notation

The notation z_r denotes the value on the axis of a standard normal density for which

$$P\left(Z \geq z_r\right) = r$$



It is said that z_r is the $100(1-r)^{th}$ percentile of the standard normal distribution.

Evaluate $Y \sim N(\mu, \sigma)$ using $Z \sim N(0,1)$

If $Y \sim N(\mu, \sigma)$, a normal distribution, then $Z = \frac{Y-\mu}{\sigma}$ has a standard normal distribution.

Check $E[Z]$:
$$E[Z] = E\left[\frac{Y-\mu}{\sigma}\right] = \frac{E[Y]-\mu}{\sigma} = \frac{\mu-\mu}{\sigma} = 0$$

Check $Var(Z)$:
$$Var(Z) = Var\left(\frac{Y-\mu}{\sigma}\right) = Var\left(\frac{Y}{\sigma} - \frac{\mu}{\sigma}\right) = \frac{1}{\sigma^2}Var(Y) = \frac{\sigma^2}{\sigma^2} = 1$$

Check cumulative distribution function: see text

Normal Distribution Example

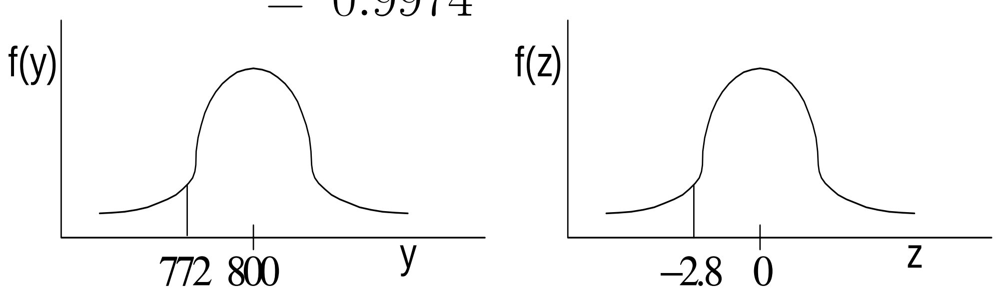
Suppose Y is the breaking strength (newtons) of a material, and

$$Y \sim N(\mu = 800, \sigma = 10).$$

We need this material to have a breaking strength of at least 772.

Find $P\,(Y \geq 772)$.
Transform Y to Z using $Z = \frac{Y-\mu}{\sigma}$

$$\begin{aligned} P\,(Y \geq 772) &= P\,(y - \mu \geq 722 - 800) \\ &= P\left(\frac{Y-\mu}{\sigma} \geq \frac{772-800}{10}\right) \\ &= P\left(Z \geq \frac{-28}{10}\right) = P\,(z \geq -2.8) \\ &= 1 - P\,(z \leq -2.8) = 1 - 0.0026 \\ &= 0.9974 \end{aligned}$$

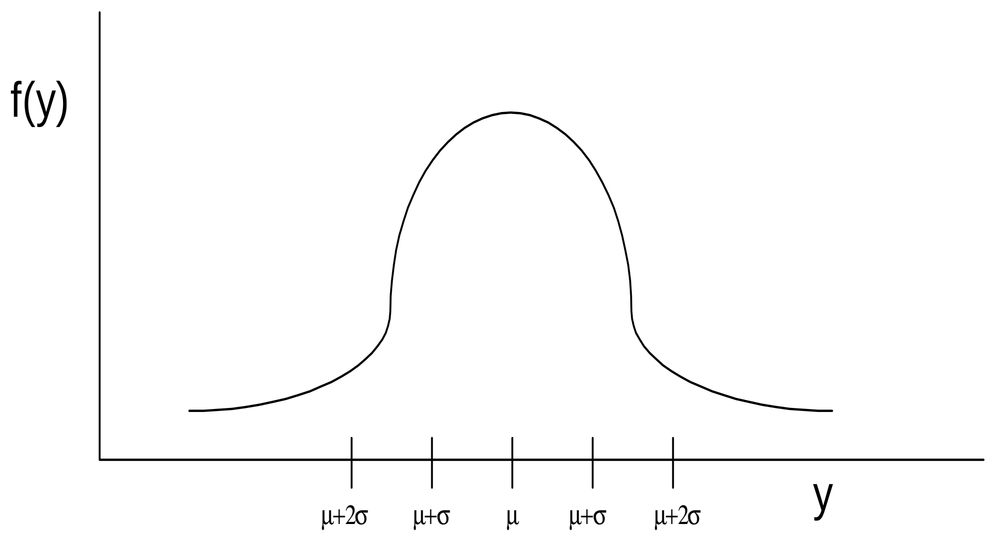


Relating σ to density

Normal Probability Rule:

If $Y \sim N(\mu, \sigma)$ then

$$\begin{aligned} P\left(-\sigma < Y - \mu < \sigma\right) &\approx 0.68 \\ P\left(-2\sigma < Y - \mu < 2\sigma\right) &\approx 0.95 \\ P\left(-3\sigma < Y - \mu < 3\sigma\right) &\approx 0.99 \end{aligned}$$



Relating σ to Density continued

Chebyshev's Inequality:

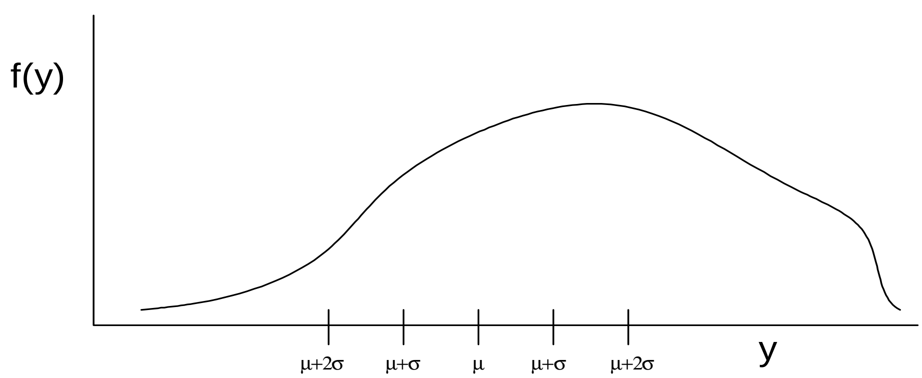
If Y is any random variable, with mean μ and standard deviation σ, then for $k > 0$,

$$P\left(|Y - \mu| < k\sigma\right) \geq 1 - \frac{1}{k^2}.$$

$$k = 1 \qquad P\left(-\sigma < Y - \mu < \sigma\right) \geq 1 - \frac{1}{1} = 0$$

$$k = 2 \qquad P\left(-2\sigma < Y - \mu < 2\sigma\right) \geq 1 - \frac{1}{4} = \frac{3}{4}$$

$$k = 3 \qquad P\left(-3\sigma < Y - \mu < 3\sigma\right) \geq 1 - \frac{1}{9} = \frac{8}{9}$$

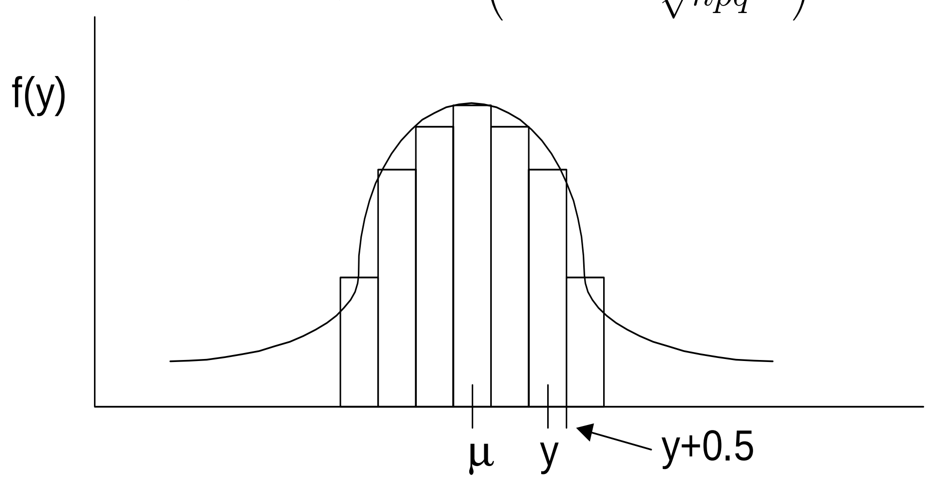


Using Normal Distribution To Approximate Binomial Distribution

$$\text{Suppose } Y \sim Binomial(n,p)$$
$$\mu = np$$
$$\sigma^2 = npq$$

and suppose the binomial distribution is fairly symmetric ($np \geq 5$ and $nq \geq 5$).

$$\text{Then } P(Y \leq y) \approx P\left(Z \leq \frac{y+0.5-np}{\sqrt{npq}}\right)$$



Family of Gamma Distributions

$$\underline{Y \sim \text{Gamma}(\alpha, \beta)}$$

$$\begin{aligned} f(y;\alpha,\beta) &= \frac{1}{\beta^{\alpha}\Gamma(\alpha)} y^{\alpha-1} e^{-y/\beta} \quad \text{if } y \geq 0 \\ \mu &= \alpha\beta \\ \sigma^2 &= \alpha\beta^2 \end{aligned}$$

$$\underline{Y \sim \text{Exponential}(\lambda)}$$

$$f(y;\lambda) = \lambda e^{-\lambda y}$$

$$\mu = \frac{1}{\lambda}$$

$$\sigma^2 = \frac{1}{\lambda^2}$$

Special case of

$\text{Gamma}(\alpha = 1, \beta = \frac{1}{\lambda})$

Used in reliability,

quality & queueing

$$\underline{Y \sim \text{Chi-Squared}(\nu)}$$

$$f(y;\nu) = \frac{1}{2^{\nu/2}\Gamma(\frac{\nu}{2})} y^{\frac{\nu}{2}-1} e^{-y/2}$$

$$\mu = \nu$$

$$\sigma^2 = 2\nu$$

Special case of

$\text{Gamma}(\alpha = \frac{\nu}{2}, \beta = 2)$

Used in hypothesis testing

(ν degrees of freedom)

Define Gamma Function, $\Gamma(\alpha)$

Definition:
For $\alpha > 0$, the gamma function is defined by

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy$$

Properties:

- For any $\alpha > 1$, $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$
- For any positive integer n,

$$\Gamma(n) = (n-1)!$$

- $\Gamma(\frac{1}{2}) = \sqrt{\pi}$
- $\Gamma(1) = 1$

Gamma Function Example

Example:

$$\begin{aligned} Y &= \text{survival time (weeks) of a mouse} \\ &\quad\ \text{exposed to gamma radiation} \\ Y &\sim \text{Gamma}(\alpha = 8, \beta = 15) \end{aligned}$$

$$\begin{aligned} \text{Mean survival time} &= E\left[Y\right] = \alpha\beta \\ &= (8)(15) = 120 \text{ weeks} \\ \text{Variance} &= Var\left(y\right) = \alpha\beta^2 = 1800 \\ \text{Standard Deviation} &= \sigma = \sqrt{1800} = 42.43 \text{ weeks} \end{aligned}$$

Prob. a mouse survives at least 30 weeks

$$\begin{aligned} = P\left(Y \geq 30\right) &= 1 - P\left(Y \leq 30\right) \\ &\vdots \ \text{evaluate numerically} \\ &\approx 1 - 0.001 - 0.999 \end{aligned}$$

Exponential Distribution

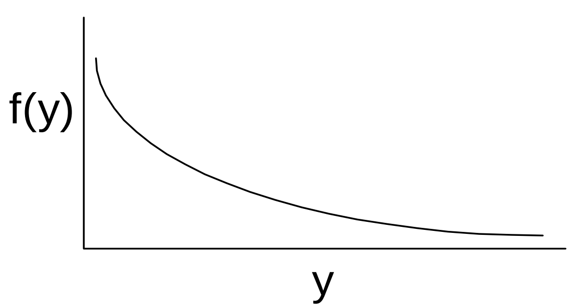
$$Y \sim \text{Exponential}(\lambda) \qquad (\text{Gamma}(\alpha = 1, \beta = \frac{1}{\lambda}))$$

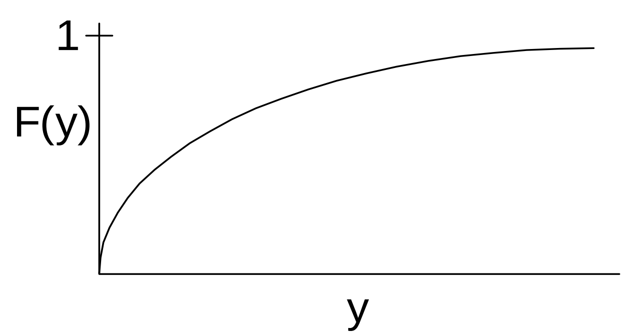
$$f(y;\lambda) = \begin{cases} \lambda e^{-\lambda y} & \text{if } y \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\mu = \frac{1}{\lambda}$$
$$\sigma^2 = \frac{1}{\lambda^2}$$

$$F(y;\lambda) = \begin{cases} 0 & y \leq 0 \\ 1 - e^{-\lambda y} & y \geq 0 \end{cases}$$

Useful in statistical quality control, reliability and queueing.





Chi-Squared Distribution

$$Y \sim \text{Chi-Squared}(\nu) \qquad (\text{Gamma}(\alpha = \frac{\nu}{2}, \beta = 2))$$

often denoted χ^2, ν is called
degrees of freedom

$$f(y;\nu) = \begin{cases} \frac{1}{2^{\nu/2}\Gamma(\nu/2)} y^{(\nu/2)-1} e^{-y/2} & y \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned} \mu &= \nu \\ \sigma^2 &= 2\nu \end{aligned}$$

Useful in hypothesis testing.

Weibull Distribution

$$Y \sim \text{Weibull}(\alpha, \beta)$$

$$f(y; \alpha, \beta) = \begin{cases} \frac{\alpha}{\beta^\alpha} y^{\alpha-1} e^{-(y/\beta)^\alpha} & y \geq 0 \\ 0 & y < 0 \end{cases}$$

$$\begin{aligned} \mu &= \beta\Gamma(1 + 1/\alpha) \\ \sigma^2 &= \beta^2\Gamma(1 + 2/\alpha) - \beta^2(\Gamma(1 + 1/\alpha))^2 \end{aligned}$$

$$F(y; \alpha, \beta) = \begin{cases} 0 & y < 0 \\ 1 - e^{-(y/\beta)^\alpha} & y \geq 0 \end{cases}$$

Notice: The exponential distribution (λ) is also a special case of the Weibull distribution with ($\alpha = 1, \beta = 1/\lambda$).

Useful in reliability.

Chapter 5

Joint Probability Distributions

Overview of Chapter 5

- Joint Densities and Independence
- Expectation and Covariance
- Correlation
- Conditional Densities

Joint Probability Distributions for Discrete Random Variables

Let X and Y be discrete random variables. The joint density function for X and Y is,

$$p_{XY}(x, y) = P(X = x \text{ and } Y = y).$$

- Two Conditions:

$$\begin{array}{ll} (1) & p_{XY}(x, y) \geq 0 \\ (2) & \sum_{all\ x} \sum_{all\ y} p_{XY}(x, y) = 1 \end{array}$$

- The discrete marginal density for X is,

$$p_X(x) = \sum_{all\ y} p_{XY}(x, y)$$

- The discrete marginal density for Y is,

$$p_Y(y) = \sum_{all\ x} p_{XY}(x, y)$$

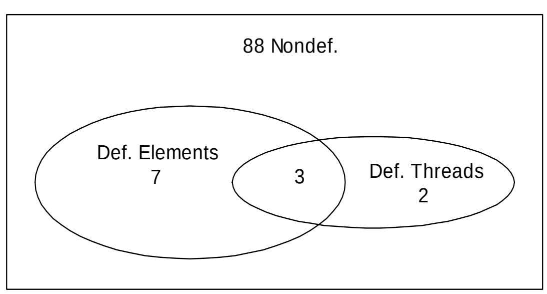
Example: Suppose we have 100 light bulbs, we choose 2 randomly and 2 attributes are tested:

(1) Does the light work? (element)

(2) Does it fit the socket? (thread)

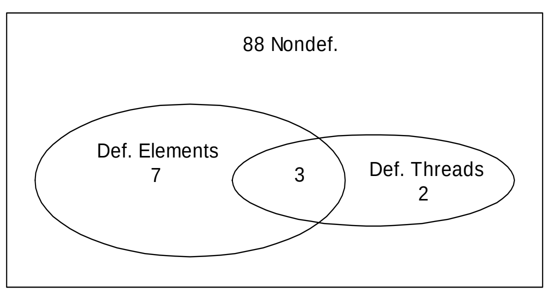
X = number of light bulbs with defective elements
Y = number of light bulbs with defective threads

Suppose 10 lightbulbs have defective elements
5 lighbulbs have defective threads,
3 lightbulbs have defective elements and threads.



What is the joint probability distribution?

P(X=x,Y=y)	y=0	y=1	y=2
x=0			
x=1			
x=2			



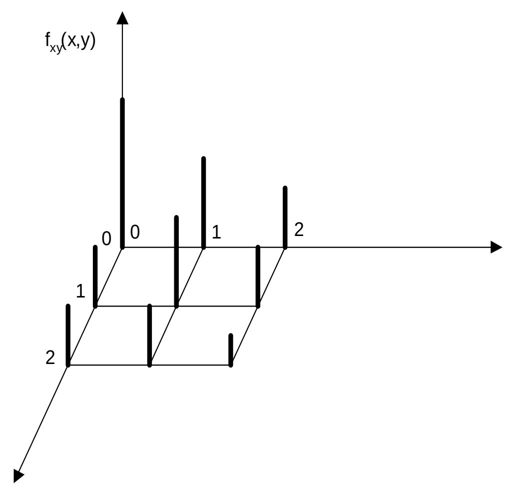
Example continued:

P(X=x,Y=y)	y=0	y=1	y=2	$p_X(x)$
x=0	0.7733	0.0356	0.0002	0.8091
x=1	0.1244	0.0562	0.0012	0.1818
x=2	0.0042	0.0042	0.0006	0.0090
$p_Y(y)$	0.9019	0.0960	0.0020	1.0000

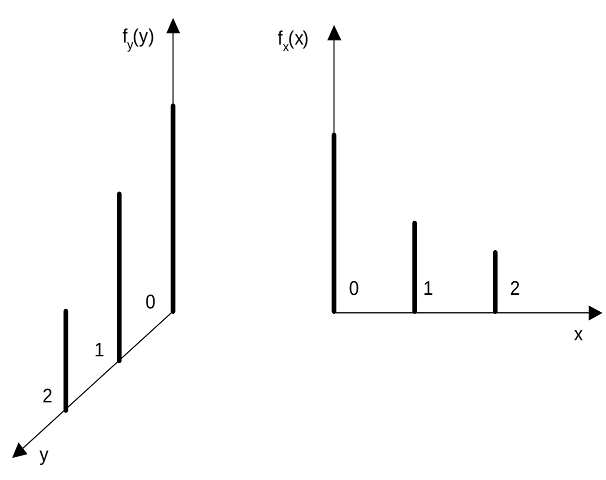
Notice that $\sum_{x=0}^{x=2} \sum_{y=0}^{y=2} p(x, y) = 1.0$

Density Functions

- Joint Density Function



- Marginal Density Functions



Joint Probability Distribution

for Continuous Random Variables

Let X and Y be continuous random variables. The joint density function for X and Y is,

$$f_{XY}(x, y).$$

- Two conditions:

$$(1)\ f_{XY}(x, y) \geq 0$$
$$(2)\ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{XY}(x, y) dx dy = 1$$

- The continuous marginal density for X:

$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x, y) dy$$

- The continuous marginal density for Y:

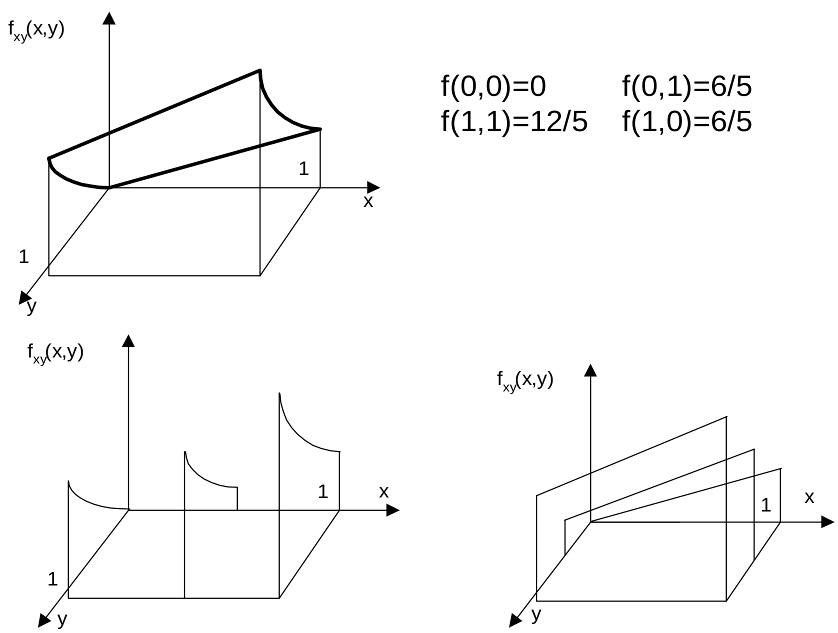
$$f_Y(x) = \int_{-\infty}^{+\infty} f_{XY}(x, y) dx$$

Example: A bank operates a drive-in facility and a walk-up window.

Let X = proportion of time the drive-in is used
$\quad Y$ = proportion of time the walk-up is used

$$f_{XY}(x,y) = \begin{cases} \frac{6}{5}(x+y^2) & \text{for } 0 \le x \le 1, 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Note: $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \dfrac{6}{5}(x+y^2)dxdy = 1$

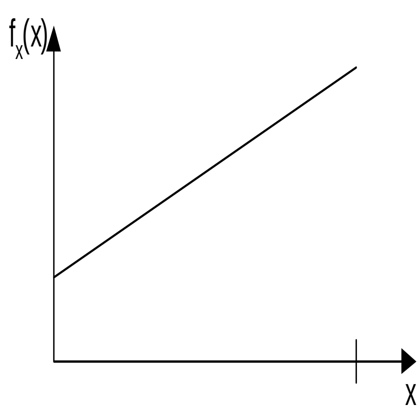


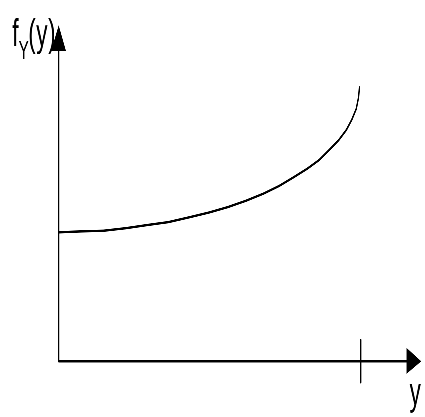
Example continued:

Marginal density functions:

$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x,y)dy = \int_0^1 \frac{6}{5}(x+y^2)dy$$
$$= \begin{cases} \frac{6}{5}(x+\frac{2}{5}) & \text{for } 0 \le x \le 1, 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{XY}(x,y)dx = \int_0^1 \frac{6}{5}(x+y^2)dx$$
$$= \begin{cases} \frac{6}{5}(y^2+\frac{3}{5}) & \text{for } 0 \le x \le 1, 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$





Independence

Let X and Y be random variables with joint density $f_{XY}(x, y)$ and marginal densities $f_X(x)$ and $f_Y(y)$ respectively.

The two random variables X and Y are independent if and only if

$$f_{XY}(x, y) = f_X(x)f_Y(y)$$

for all x and y.

Expected Value

Let X and Y be 2 random variables. Then the expected value of a function $H(X,Y)$ denoted as $E[H(X,Y)]$ is:

$E[H(X,Y)]$

$$= \begin{cases} \sum_{all\ x} \sum_{all\ y} H(x,y) p_{XY}(x,y) & \text{if } X,Y \\ & \text{discrete} \\ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} H(x,y) f_{XY}(x,y) dy dx & \text{if } X,Y \\ & \text{continuous} \end{cases}$$

Examples:

$$\begin{aligned} H(X,Y) &= X+Y \\ H(X,Y) &= |X-Y| \\ H(X,Y) &= X/Y \\ H(X,Y) &= XY \end{aligned}$$

<u>Covariance</u>

A measure of how strongly X and Y are related to one another.

$$
\begin{aligned}
Cov(X,Y) = \sigma_{XY} &= E[(X-\mu_X)(Y-\mu_Y)] \\
&= E[XY] - \mu_X\mu_Y
\end{aligned}
$$

$$
= \begin{cases}
\sum\limits_{all\ x}\sum\limits_{all\ y}(x\text{-}\mu_x)(y\text{-}\mu_y)p_{XY}(x,y) & \text{if } X,Y \text{ discrete} \\
\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x\text{-}\mu_x)(y\text{-}\mu_y)f_{XY}(x,y)dxdy & \text{if } X,Y \text{ continuous}
\end{cases}
$$

If X and Y are <u>independent</u>, then

$$Cov(X,Y) = 0.$$

But $Cov(X,Y)$ can $= 0$, when X and Y are dependent.

Correlation

The correlation coefficient of X and Y denoted $Corr(X,Y)$, or $\rho_{X,Y}$ is:

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

- If a, c are either both positive or both negative, $Corr(aX+b, cY+d) = Corr(X,Y)$
- For any X, Y, then

$$-1 \leq Corr(X,Y) \leq 1$$

- If X, Y are independent, then $\rho_{X,Y} = 0$

Note: $\rho_{XY} = 0$ does not imply X, Y are independent

Conditional Probability

Let X and Y be two random variables with joint density f_{XY} and marginal densities f_X and f_Y.

The conditional density for X given $Y = y$ is,

$$f_{X|Y}(x) = \frac{f_{XY}(x, y)}{f_Y(y)}$$

The conditional density for Y given $X = x$ is,

$$f_{Y|X}(y) = \frac{f_{XY}(x, y)}{f_X(x)}$$

Chapters 6 and 7

Point and Interval Estimation

Overview of Chapters 6 and 7

- Point Estimation
- Unbiased Estimators
- Interval Estimators (Confidence Intervals)
- Estimation of a Population Mean
- Estimation of a Population Variance
- Estimation of a Population Proportion

Point Estimation

A point estimate of a parameter θ is obtained by selecting an appropriate statistic, and computing its value using the sample data. The selected statistic is called the point estimator.

Example:

Parameter of interest:	μ, the pop. mean
Statistic or estimator:	$\bar{Y} = \sum_{i=1}^{n} Y_i/n = \hat{\mu}$
Point estimate:	5.77

Example:

Parameter of interest:	p, prob. of success
Statistic or estimator:	$Y/n = \hat{p}$
Point estimate:	0.65

Example:

Parameter of interest:	σ^2, variance
Statistic or estimator:	$s^2 = \sum_{i=1}^{n} \frac{(Y_i - \bar{Y})^2}{n-1}$
Point estimate:	1.0

Unbiased Estimators

An estimator $\hat{\theta}$ is an unbiased estimator for a parameter θ if and only if

$$E[\hat{\theta}] = \theta.$$

- Example:
  - If $Y \sim \text{Binomial}(n, p)$, then the sample proportion, Y/n is an unbiased estimator of p, probability of success
  - To check:

$$E[Y/n] = \frac{1}{n}E[Y] = \frac{np}{n} = p$$

- Example:
  - If $Y_1, Y_2, \ldots, Y_n$ is a random sample, then $\bar{Y} = \sum_{i=1}^{n} Y_i/n$ is an unbiased estimator of μ.
  - To check:

$$
\begin{aligned}
E[\bar{Y}] &= E[\sum_{i=1}^{n} Y_i/n] \\
&= \tfrac{1}{n}(E[Y_1] + E[Y_2] + \ldots + E[Y_n]) \\
&= \tfrac{1}{n}(\mu + \mu + \ldots + \mu) = \tfrac{n\mu}{n} \\
&= \mu
\end{aligned}
$$

- Let s^2 be the sample variance based on a random sample of size n, from a distribution with mean μ and variance σ^2.
- s^2 is an <u>unbiased</u> estimator for σ^2, but s is a <u>biased</u> estimator for σ. (as long as n is large, the bias is negligible)
- To check:

$$
\begin{aligned}
E[s^2] &= \tfrac{1}{n-1} E[\sum_{i=1}^{n} (Y_i - \bar{Y})^2] \\
&= \quad \vdots \\
&= \tfrac{1}{n-1}(n\sigma^2 - \sigma^2) \\
&= \sigma^2
\end{aligned}
$$

- There may be several unbiased estimators for a parameter.
- For example,

$$\bar{Y}, \tilde{Y}, \frac{\min Y_i + \max Y_i}{2}, \bar{Y}_{\text{trimmed}}$$

  are all unbiased estimators for μ.
- How to choose among them?
  - Method of Moments
  - Maximum Likelihood Estimators
  - MVUE Minimum Variance Unbiased Estimator

Confidence Intervals

- Instead of a point estimator, an interval estimator (or confidence intervals) gives more information.
- A $100(1-\alpha)\%$ confidence interval for a parameter θ is a random interval $[L_1, L_2]$, such that

$$P(L_1 \leq \theta \leq L_2) \approx 1 - \alpha.$$

- To find L_1 and L_2, we need to know the distribution of a random variable involving θ.
- For example, $\bar{Y} - \hat{\mu}$, and we need distribution of $\bar{Y}$.
- Let $Y_1, Y_2, \ldots, Y_n$ be a random sample of size n from a normal distribution with mean μ and variance σ^2. Then

$$\bar{Y} \sim \text{Normal}(\mu_{\bar{Y}} = \mu, \sigma^2_{\bar{Y}} = \sigma^2/n)$$

Sums of Normal R.V.'s

- If $Y_1, Y_2, \ldots, Y_n$ are independent random variables with $Y_i \sim \mathrm{N}(\mu_i, \sigma_i)$ then

$$Y = a_1Y_1 + a_2Y_2 + \ldots + a_nY_n$$

$$Y \sim \mathrm{N}(\mu_Y, \sigma_Y)$$

- where

$$\mu_Y = \mu_1 + \mu_2 + \ldots + \mu_n$$

$$\sigma_Y^2 = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \ldots + a_n^2\sigma_n^2$$

The Central Limit Theorem

- Let $Y_1, Y_2, \ldots, Y_n$ be a random sample with mean μ and variance σ^2.
- For sufficiently large n,

$$\bar{Y} \sim \mathrm{N}(\mu_{\bar{Y}} = \mu, \sigma^2_{\bar{Y}} = \sigma^2/n)$$

- and

$$T_o \sim \mathrm{N}(\mu_{T_o} = n\mu, \sigma^2_{T_o} = n\sigma^2)$$

- The larger n, the better the approximation. (if $n > 30$, CLT can be used)

Example:

40 samples of one-gallon cans of paint were tested to see how many square feet could be covered with 1 can. The results of the experiment are given below. Construct a confidence interval for average square footage covered by 1 can of paint.

508	486	551	536	482
534	472	489	529	452
477	507	508	459	528
530	490	541	504	553
507	540	508	549	492
544	515	483	516	533
531	516	478	564	581
546	472	555	465	501

95% **Confidence Interval for** μ

- Suppose we have a random sample $Y_1, Y_2, \ldots, Y_n$ where $Y_i \sim \mathrm{N}(\mu, \sigma^2)$ and μ is unknown and σ is known. Then

  $$P(\bar{Y} - 1.96\sigma/\sqrt{n} \leq \mu \leq \bar{Y} + 1.96\sigma/\sqrt{n}) = .95$$

- Now suppose we observe $Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n$ and we compute $\bar{y}$ and substitute $\bar{y}$ for $\bar{Y}$. Then

  $$(\bar{y} - 1.96\sigma/\sqrt{n}, \bar{y} + 1.96\sigma/\sqrt{n})$$

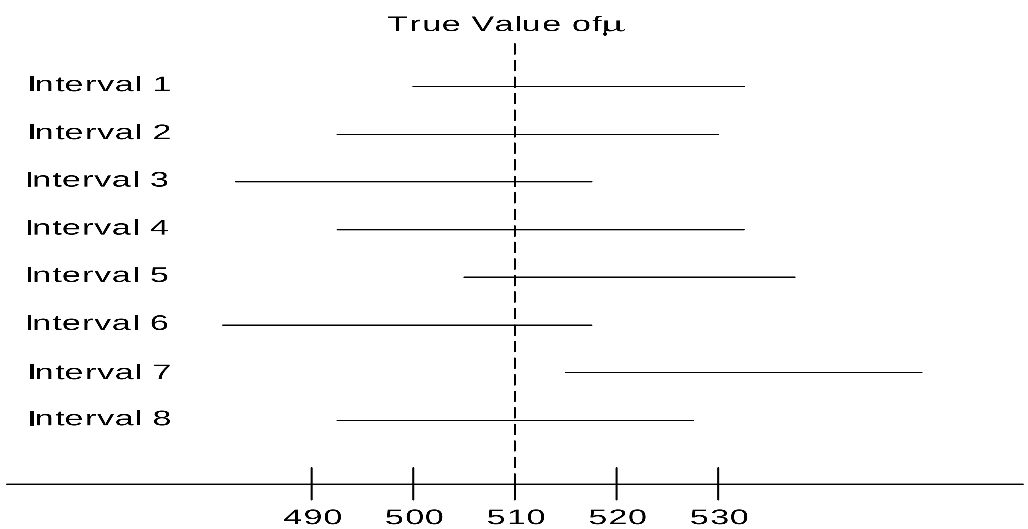
  is called a 95% confidence interval for μ or

  $$\bar{Y} - 1.96\sigma/\sqrt{n} \leq \mu \leq \bar{Y} + 1.96\sigma/\sqrt{n}$$

  with 95% confidence

Interpreting Confidence Intervals

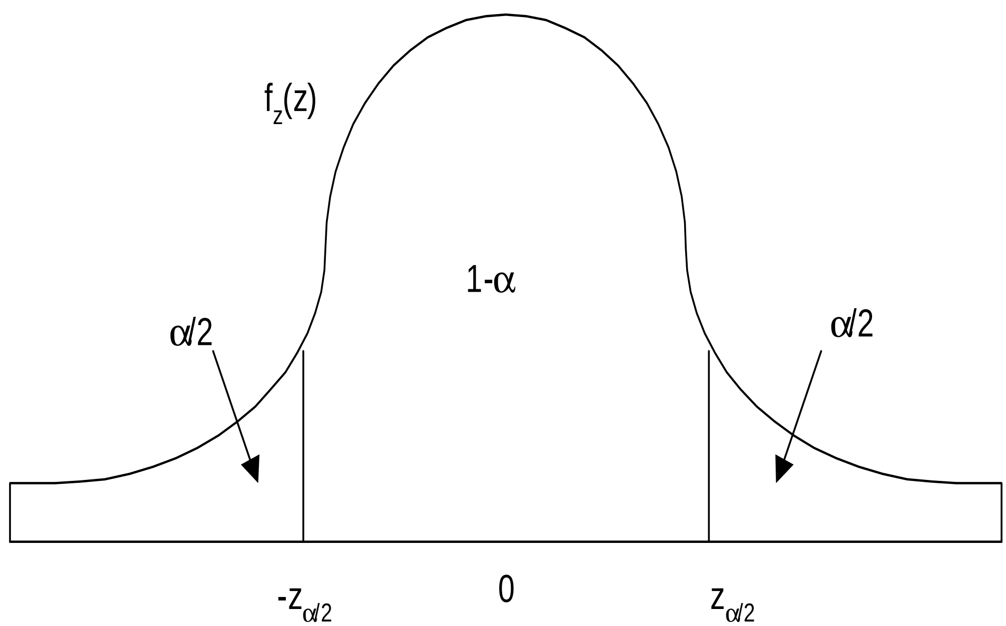
Suppose we sampled another 40 paint cans and constructed a confidence interval. And if we did this again and again, approximately 95% of the constructed intervals would contain μ, and 5% of the constructed intervals would not contain μ.



Confidence Interval for μ

A 100$(1-\alpha)$% confidence interval, $0 < \alpha < 1$, for the mean μ of a Normal population (when σ is known) is given by:

$$(\bar{Y} - Z_{\alpha/2}\sigma/\sqrt{n} \leq \mu \leq \bar{Y} + Z_{\alpha/2}\sigma/\sqrt{n})$$



Deriving a Confidence Interval

- We can derive a confidence interval for any statistic
- Suppose $Y_1, Y_2, \ldots, Y_n$ is a random sample and we want to estimate θ
- Find an estimator $\hat{\theta}$ and its distribution
- This suggests a function $h(Y_1, Y_2, \ldots, Y_n; \theta)$ with

$$P(a \leq h(Y_1, Y_2, \ldots, Y_n; \theta) \leq b) = 1 - \alpha$$

- Isolate θ;

$$P(l_1(Y_1, Y_2, \ldots, Y_n) \leq \theta \leq l_2(Y_1, Y_2, \ldots, Y_n)) = 1 - \alpha$$

- The lower and upper confidence limits for a $100(1-\alpha)\%$ confidence interval are given by

$$l_1(Y_1, Y_2, \ldots, Y_n) \quad \text{and} \quad l_2(Y_1, Y_2, \ldots, Y_n)$$

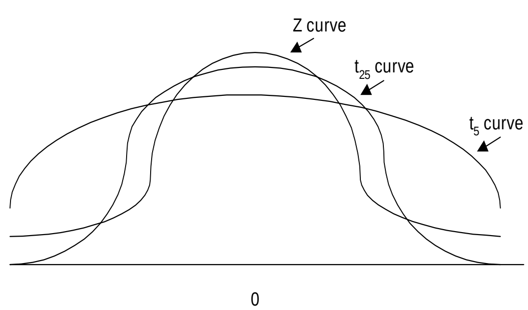
Estimating the Mean

- $\bar{Y}$ is our estimator for the population μ
- If the population σ^2 is <u>known</u>, and the sample comes from a normal distribution, then $Z = \frac{\bar{Y}-\mu}{\sigma/\sqrt{n}}$ has a standard normal distribution
- If the population σ^2 is <u>unknown</u> but the sample comes from a normal distribution, then $T = \frac{\bar{Y}-\mu}{S/\sqrt{n}}$ has a t-distribution with $\gamma = n-1$ degrees of freedom

t-distribution

The t-distribution (T), is similar to the standard normal distribution (Z)

- both means equal zero ($\mu_Z = \mu_{t_\gamma} = 0$)
- both are bell-shaped
- the normal distribution has two parameters (μ, σ), but the t-distribution has only one parameter ($\gamma = \text{DOF}$)
- the density t_γ is spread out more than the standard normal, $\sigma_{t_\gamma} > \sigma_Z = 1$



- as $\gamma \to +\infty$, $t_\gamma \to$ standard normal

Estimating the Mean continued

If the population σ^2 is **known**, and the sample comes from a normal distribution, we had:

$$P(-z_{\alpha/2} \leq Z = \frac{\bar{Y}-\mu}{\sigma/\sqrt{n}} \leq +z_{\alpha/2}) = 1-\alpha$$

gives

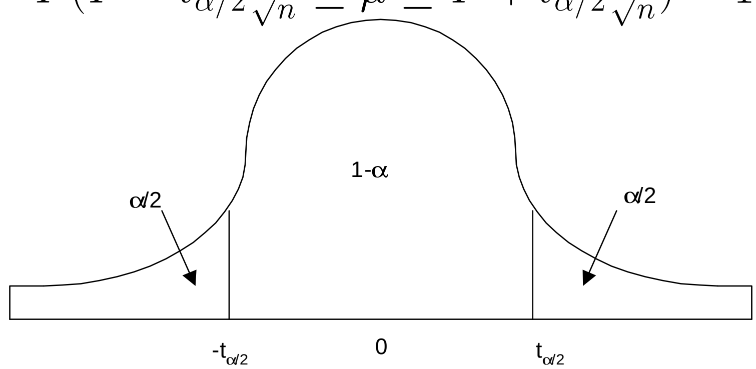
$$P(\bar{Y} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{Y} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = 1-\alpha$$

If the population σ^2 is **unknown** but the sample comes from a normal distribution, we have:

$$P(-t_{\alpha/2} \leq T = \frac{\bar{Y}-\mu}{S/\sqrt{n}} \leq +t_{\alpha/2}) = 1-\alpha$$

gives

$$P(\bar{Y} - t_{\alpha/2}\frac{S}{\sqrt{n}} \leq \mu \leq \bar{Y} + t_{\alpha/2}\frac{S}{\sqrt{n}}) = 1-\alpha$$



Estimating the Mean summarized

100(1-α)% CI on μ when σ^2 **is known:**

Let $Y_1, Y_2, \ldots, Y_N$ be a random sample from a normal distribution with mean μ and variance σ^2. A $100(1-\alpha)\%$ confidence interval on μ is given by

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

100(1-α)% CI on μ when σ^2 **is unknown:**

Let $Y_1, Y_2, \ldots, Y_N$ be a random sample from a normal distribution with mean μ and variance σ^2. A $100(1-\alpha)\%$ confidence interval on μ is given by

$$\bar{Y} \pm t_{\alpha/2} \frac{S}{\sqrt{n}}$$

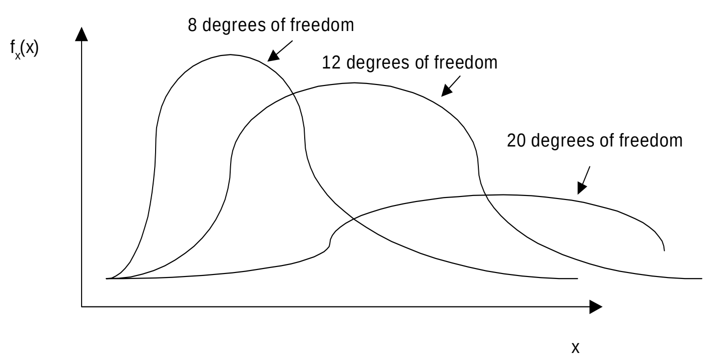
Estimating the Variance

- s^2 is our estimator for the population σ^2
- If the sample comes from a normal distribution, then

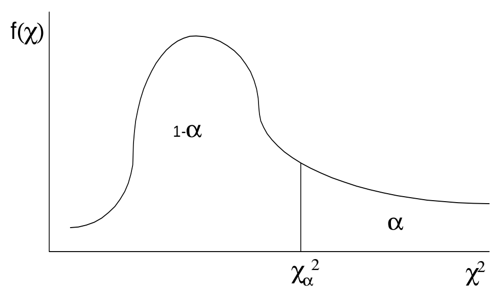
$$\frac{(n-1)s^2}{\sigma^2} = \sum_{i=1}^{n} \frac{(Y_i - \bar{Y})^2}{\sigma^2}$$

  has a chi-squared distribution with $n-1$ degrees of freedom

Chi-squared (χ^2) distribution:



Estimating the Variance continued



Estimating the Variance concluded

- If the sample comes from a normal distribution, we have:

$$P(\chi^2_{1-\alpha/2} \leq \frac{(n-1)s^2}{\sigma^2} \leq \chi^2_{\alpha/2}) = 1 - \alpha$$

  and we isolate σ^2 using: $\sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}$ and $\frac{(n-1)s^2}{\chi^2_{\alpha/2}} \leq \sigma^2$ obtaining a

- $100(1-\alpha)\%$ Confidence Interval on σ^2

$$P\left(\frac{(n-1)s^2}{\chi^2_{\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}\right) = 1 - \alpha$$

Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from a normal distribution with mean μ and variance σ^2. A $100(1-\alpha)\%$ confidence interval on σ^2 is given by

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}}, \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}$$

Estimating Proportions

- Let p be the proportion of “successes” in a population.
- Let Y be the number of successes in a sample of size n. The sample proportion, $\hat{p} = \frac{Y}{n}$ is our estimator for p. It is an unbiased estimator because

$$\begin{aligned} E[\hat{p}] &= E\left[\tfrac{Y}{n}\right] = \frac{E[Y]}{n} \\ &= \frac{np}{n} = p \\ Var(\hat{p}) &= Var\left(\tfrac{Y}{n}\right) \\ &= \frac{1}{n^2} Var(Y) \\ &= \frac{npq}{n^2} = \frac{p(1-p)}{n} \end{aligned}$$

Estimating Proportions Continued

- By the Central Limit Theorem,

$$\hat{p} = \frac{Y}{n} \sim \left\{ \begin{array}{l} \text{approx. Normal distribution} \\ \text{with mean } p \text{ \& variance } \frac{p(1-p)}{n}. \end{array} \right\}$$

so $$Z = \frac{\frac{Y}{n} - p}{\sqrt{\frac{p(1-p)}{n}}} \text{is approx. Standard Normal}$$

- When n is large enough,

$$P\left(-z_{\alpha/2} \leq \frac{\frac{Y}{n}-p}{\sqrt{p(1-p)/n}} \leq +z_{\alpha/2}\right) = 1 - \alpha$$

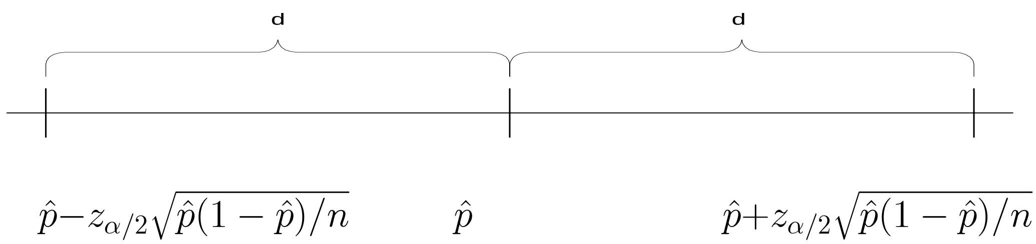
$$P\left(\frac{Y}{n} - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \leq p \leq \frac{Y}{n} + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right)$$
$$= 1 - \alpha$$

- It is so difficult to completely isolate p that we replace p by its unbiased estimator $\hat{p} = Y/n$ to obtain the

Confidence Interval on p:

$$\hat{p} \pm z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$$

Sample Size for Estimating p



We are $100(1-\alpha)\%$ confident that p is in the above interval, of length $2d$, where

$$d = z_{\alpha/2} \cdot \sqrt{\hat{p}(1-\hat{p})/n}$$

Solve for n:

$$n \approx \frac{z_{\alpha/2}^2 \cdot \hat{p}(1-\hat{p})}{d^2} \leq \frac{z_{\alpha/2}^2}{4d^2}$$

(since $\hat{p}(1-\hat{p}) \leq 1/4$).

Estimating Proportions Example

A certain treatment was effective in 32 out of 50 cases in which it was tried. Find a 95% confidence interval for the probability of effectiveness in a single treatment.

$$\hat{p} = \frac{Y}{n} = \frac{32}{50}$$

Since 50 is large, a 95% confidence interval is given by:

$$\left(\hat{p} - z_{\alpha/2} \cdot \sqrt{\hat{p}(1-\hat{p})/n} \;\;,\;\; \hat{p} + z_{\alpha/2} \cdot \sqrt{\hat{p}(1-\hat{p})/n}\right)$$

$$\left(\frac{32}{50} - 1.96\sqrt{\frac{32}{50} \cdot \frac{18}{50}/50} \;\;,\;\; \frac{32}{50} + 1.96\sqrt{\frac{32}{50} \cdot \frac{18}{50}/50}\right)$$

$$(0.64 - 0.133 \;\;,\;\; 0.64 + 0.133)$$

$$(0.507 \;\;,\;\; 0.773)$$

Chapter 8

Test of Hypotheses

Overview of Chapter 8

- Elements of a Statistical Test
- Choosing the Null and Alternative Hypotheses
- Testing a Population Mean
- Testing a Population Proportion
- Comparing Two Population Means
- Comparing Two Population Proportions
- Comparing Two Population Variances
- Comparing Means with Paired Data

Hypothesis Testing

1. State the null hypothesis, H_0 and the alternative hypothesis H_A.

Ex: Two-tailed test	One-tailed test
H_0: $\mu = \mu_0$	H_0: $\mu = \mu_0$ $(\mu \leq \mu_0)$
H_A: $\mu \neq \mu_0$	H_A: $\mu \geq \mu_0$

2. Select the appropriate test statistic
Ex: Use $\overline{Y}$ to estimate μ, and use $T = \frac{\overline{Y}-\mu_0}{s/\sqrt{n}}$

3. Specify the rest of the experiment, such as:

3i. Choose the critical region for specific α and n, or

3ii. Choose n for specific α and β.

Hypothesis Testing Continued

4. Compute the value of the test statistic from the data
Ex: Compute $\overline{y}$ and t.

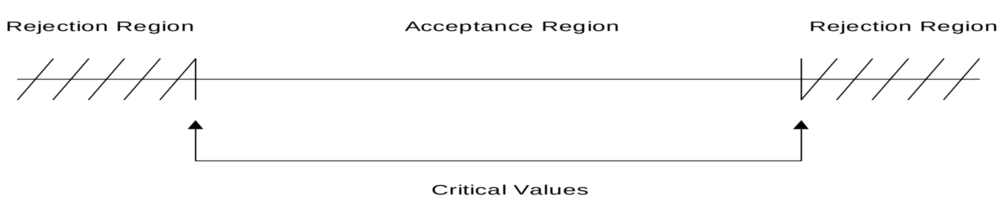
5. Decision:

- if the value of the test statistic is in the critical region, reject H_0 with α level of significance, or
- find the P-value, i.e. the smallest level of significance at which H_0 would be rejected, or
- conclude there is insufficient evidence to reject H_0

2 Types of Errors

	H_0 is True	H_0 is False
Accept H_0	No Error	Type II Error
Reject H_0	Type I Error	No Error

$$\begin{aligned} \alpha &= P\,(\text{type I error}) \leftarrow \text{level of significance} \\ &= P\,(\text{reject} H_0 | H_0 \text{ is true}) \\ \beta &= P\,(\text{type II error}) \leftarrow \text{power of the test} \\ &= P\,(\text{accept } H_0 | H_0 \text{ is false}) \end{aligned}$$



Stating the Null Hypothesis

In statistics, we can only **reject** hypotheses. We can **never** prove a hypothesis, all we can say is, we had insufficient evidence to reject the hypothesis. The strong conclusion is to provide sufficient evidence to reject the hypothesis.

State the null hypothesis, hoping we can reject it.

Example:
A beer company is going to contract a glass company to supply bottles. The mean bursting strength should exceed 100 psi. They will do an experiment & decide whether to sign the contract.

1: H_0: $\mu \leq 100$psi H_A: $\mu > 100$psi If the data allows us to reject H_0, we conclude the bottles are strong enough	2: H_0: $\mu \geq 100$psi H_A: $\mu < 100$psi If the data allows us to reject H_0, we conclude the bottles are too weak
This formulation forces the glass co. to demonstrate the bottles are strong enough.	This formulation forces a demonstration of an unusual number of weak bottles.

Diagram of α For Tests on Pop. Mean

$\alpha = P\,(\text{type I error}) = P\,(\text{reject } H_0 | H_0 \text{ is true})$
using $\overline{Y}$ to estimate μ,
$\alpha = P\,(\overline{Y} \text{ is in rejection region} | \mu = \mu_0)$

- Upper-tailed test:
  $H_0: \quad \mu = \mu_0 \quad (\mu \leq \mu_0) \qquad H_A: \quad \mu > \mu_0$

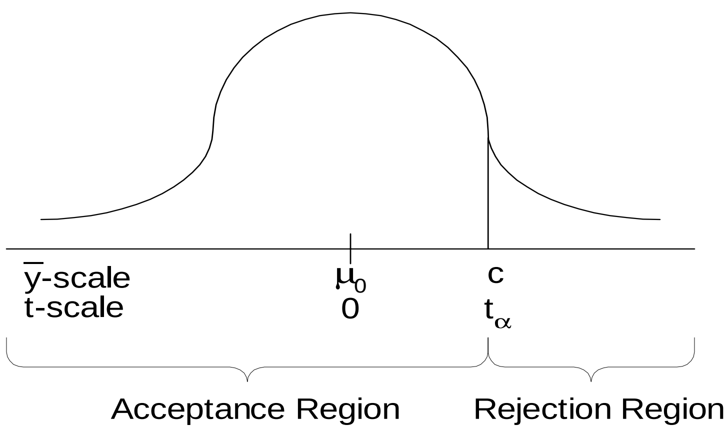
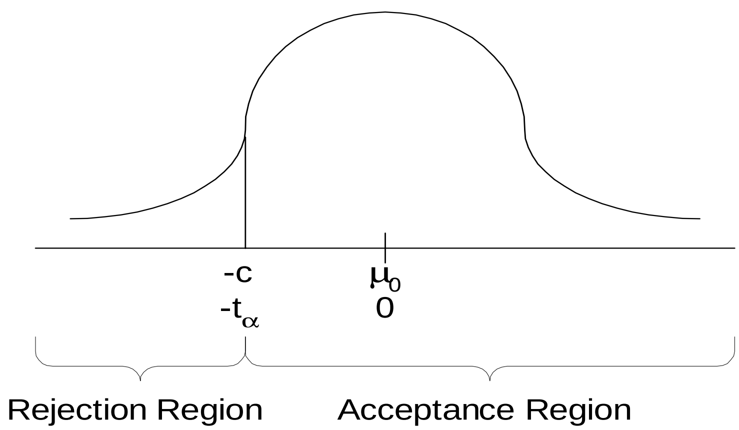
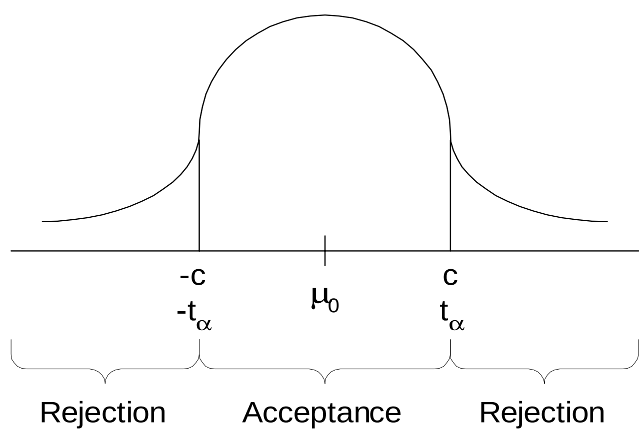


Diagram of α for Pop. Mean Cont.

- Lower-tailed test:
  $H_0: \quad \mu = \mu_0 \quad (\mu \leq \mu_0) \qquad H_A: \quad \mu < \mu_0$



- Two-tailed test:
  $H_0: \quad \mu = \mu_0 \qquad\qquad H_A: \quad \mu \neq \mu_0$



Example of New Design Testing

A company produces bias-ply tires & is considering a change in the tread design. The tire life should be able to exceed 20,000 miles.

In order to convince management that the new design significantly changes the average tire life to exeed 20,000 miles, we formulate the following null hypothesis.

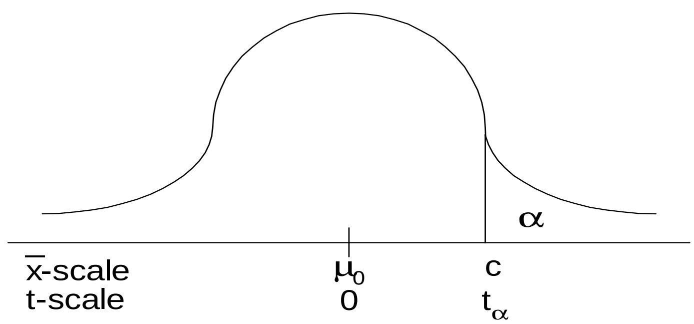
1. H_0: $\mu = 20,000 \qquad (\mu \leq 20,000)$
H_A: $\mu > 20,000$

2. Choose the test statistic & identify distribution;
use $\overline{Y}$ to estimate μ,
since our population is bell-shaped, we will use
$T = \dfrac{\overline{Y} - \mu}{s/\sqrt{n}}$ with $n - 1$ degrees of freedom.

New Design Testing Cont.

3. Specify the rest of the experiment, n, critical region, $\alpha, \beta, \ldots$

3i. First, fix $\alpha = 0.01$ level of significance, and find the critical region for a fixed n.



$$
\begin{aligned}
P\left(\overline{Y} > c | \mu = \mu_0\right) &= P\left(T = \frac{\overline{Y} - \mu_0}{s/\sqrt{n}} > t_\alpha\right) \\
&= P\left(\overline{Y} > \mu_0 + t_\alpha(s/\sqrt{n}\right) = \alpha
\end{aligned}
$$

New Design Testing Cont.

For example; if $n = 16, s = 1500$, and $\mu_0 = 20,000, \alpha = 0.01$,

$t_\alpha = t_{0.01} = 2.602$
(check table, 15 deg. of freedom)
and then
$c = 20,000 + 2.602(\frac{1500}{\sqrt{16}}) = 20,976$

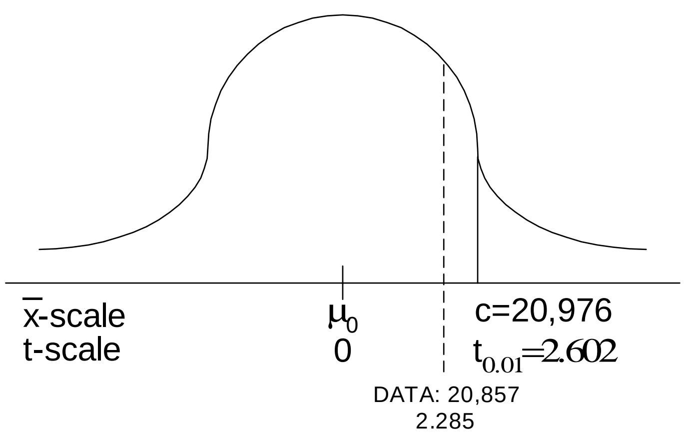
New Design Testing Cont.

4. Calculate test statistic from the data

Example: $\overline{y} = 20,857$, $s = 1500$ and $n = 16$.

Calculate t:

$$t = \frac{\overline{y} - \mu_0}{s/\sqrt{n}} = \frac{20,857 - 20,000}{1500/\sqrt{16}} = 2.285$$



New Design Testing Cont.

5. Decision:

There is **insufficient evidence** to reject the null hypothesis at a 0.01 level of significance.

However, we would be able to reject the null hypothesis at a 0.05 level of significance.

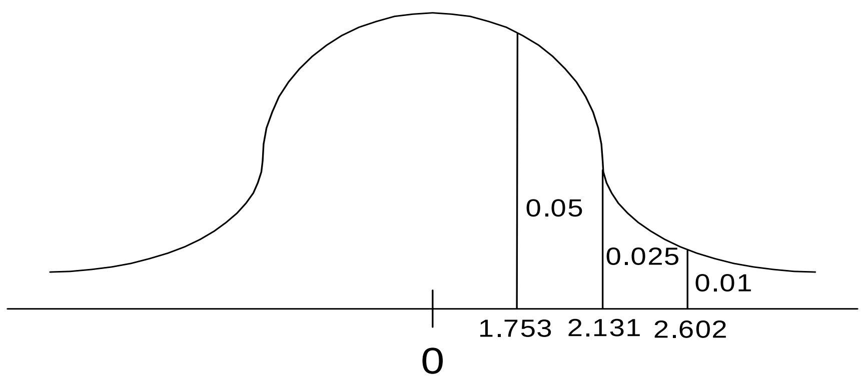
(To check this, use table to get $t_{0.05} = 1.753$, with 15 degrees of freedom.)

Find the P-value; the value at which we could just reject the null hypothesis.

The exact P-value is hard to calculate. Some computer packages numerically estimate P-value quite well.

The table in the text only gives certain values for α:

$\gamma \setminus$ F	0.90	0.95	0.975	0.99
15	1.341	1.753	2.131	2.602



Since our calculated t-value is 2.285, which is inbetween $t_{0.025} = 2.131$ and $t_{0.01} = 2.602$, then we conclude that we can reject the null hypothesis at a 0.025 level of significance (our P-value is 0.025).

Tossing a Coin Example: Estimating a Proportion p

I lost a gambling game with coin tosses and I suspect that the coin was weighted. I want to conduct an experiment to demonstrate the coin is unfair.

1. H_0: $p = 1/2 = p_0$
H_A: $p \neq 1/2$

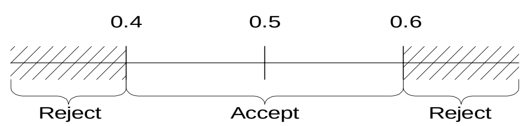
2. Select the appropriate test statistic:
use $\hat{p} = Y/n$ to estimate p.

if n is large:
$\hat{p} \sim \text{Normal}(\mu_{\hat{p}} = p_0, \sigma_{\hat{p}} = \sqrt{p_0(1-p_0)/n})$
(and $np_0 \geq 5$ and $n(1-p_0) \geq 5$)

if n is small:
$Y \sim \text{Binomial}(\mu_X = np, \sigma_{\overline{X}} = \sqrt{npq})$
(or $np_0 \ngeq 5$ or $n(1-p_0) \ngeq 5$)

Tossing a Coin Example, Cont.

3. Specify the rest of the experiment. For example, suppose the critical values are 0.4 and 0.6.



and suppose $n = 10$. Find α and $\beta(0.8)$

$$
\begin{aligned}
\alpha &= P\left(\text{reject } H_0 | H_0 \text{ is true}\right) \\
&= P\left(0.6 < \tfrac{Y}{n} \text{ or } \tfrac{Y}{n} < 0.4 | p = 0.5\right) \\
&= P\left(6 < Y \text{ or } Y > 4 | p = 0.5\right) \\
&= 1 - \sum_{y=4}^{6} \binom{10}{y} (0.5)^y (0.5)^{10-y} \\
&= 1 - \left(\binom{10}{4}(.5)^{10} + \binom{10}{5}(.5)^{10} + \binom{10}{6}(.5)^{10}\right) \\
&= 1 - (210 + 252 + 210)(0.5^{10}) \\
&= 1 - 0.656 \cong 0.344
\end{aligned}
$$

Tossing a Coin Example, Cont.

$$\begin{aligned}\beta(0.8) &= P\left(\text{accept } H_0 | H_0 \text{ is false}\right) \\ &= P\left(4 \leq Y \leq 6 | p = 0.8\right) \\ &= \sum_{y=4}^{6} \binom{10}{y} (0.8)^y (0.2)^{10-y} \simeq 0.12\end{aligned}$$

Since tossing a coin is an inexpensive experiment, we may wish to have n be very large, and have a narrow critical region.

If n is large, then

$$\hat{p} = \frac{Y}{n}$$

has an approx. Normal distribution,with
$\mu_{\hat{p}} = p$ and $\sigma_{\hat{p}} = \sqrt{p(1-p)/n}$
(also used for conf. interval) or

$$Z = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}}$$

Tossing a Coin Example, Cont.

Now, find n so that $\alpha = 0.02$ and we want a tight interval of ± 0.1.
As before:

1. H_0: $p = 1/2$
H_A: $p \neq 1/2$

2. $\hat{p}$ is estimator for p, and use Z-statistic.

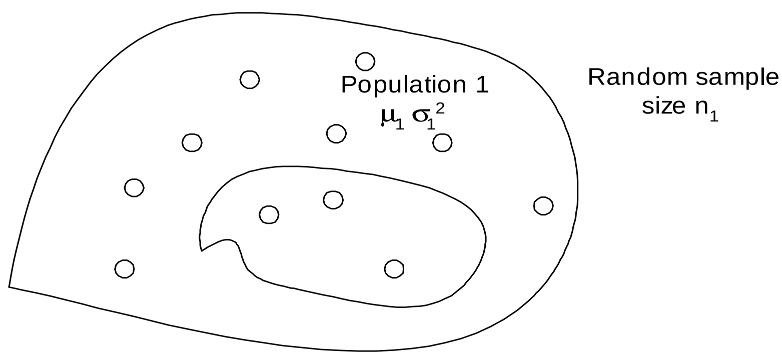
3. $\alpha = 0.02$, $-z_{\alpha/2} = -2.33$
because $P\left(Z \leq -2.33\right) = 0.01$
Use the expression $n \cong \dfrac{z^2_{\alpha/2}}{4d^2}$ with $z_{\alpha/2} = 2.33$ and $d = 0.1$
$n \cong \dfrac{(2.33)^2}{4(0.1)^2} = 135.7$
Use $n = 136$

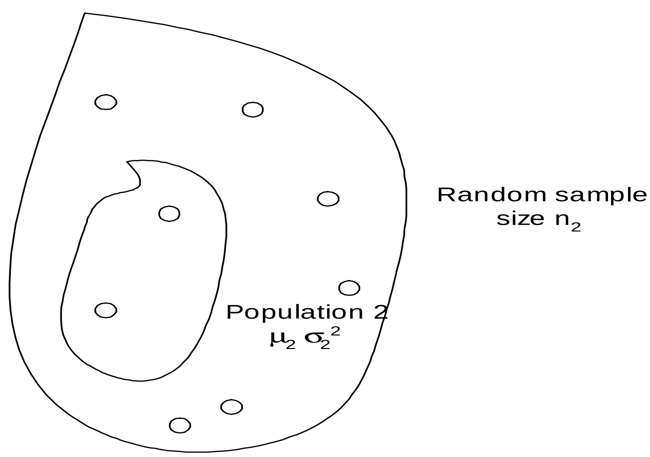
4. Conduct the experiment with $n = 136$, determine $\hat{p} = Y/n$.

5. Use P-value to reject the hypothesis with a specific level of significance.

Comparing Two Means, Proportions, or Variances

This section shows how to draw statistical inferences when comparing 2 random variables. The basic steps for developing confidence intervals & hypothesis testing remain the same.



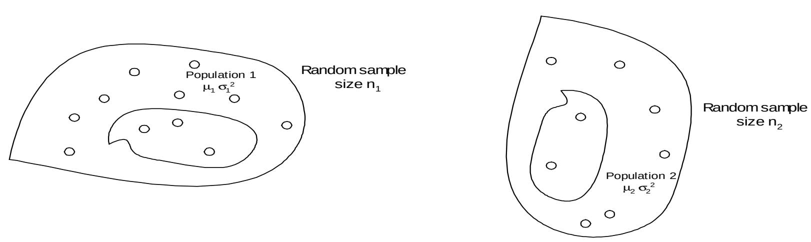


- What are you trying to estimate or what is your hypothesis?
  - Difference between means, $\mu_1 - \mu_2$.
  - Ratio of variances, σ_1^2/σ_2^2.
  - Difference between proportions, $p_1 - p_2$.
- What is your test statistic and what is its distribution?
  - Distribution of $\overline{Y_1} - \overline{Y_2}$, S_1^2/S_2^2, $\hat{p}_1 - \hat{p}_2$, depend on assumptions.

Difference Between 2 Pop. Means

Example: Compare the strength of cold-rolled steel (μ_1) with galvanized steel (μ_2), where we suspect that galvanized steel is stronger.

Population 1:	Population 2:
Cold Rolled Steel	Galvanized Steel
μ_1, σ_1^2	μ_2, σ_2^2
random sample of size n_1	random sample of size n_2



- We want to estimate $\mu_1 - \mu_2$
- Test statistic: $\overline{Y_1} - \overline{Y_2}$
  What is the distribution of $\overline{Y_1} - \overline{Y_2}$?

Distribution of $\overline{Y_1} - \overline{Y_2}$

If the sample means $\overline{Y_1}$ and $\overline{Y_2}$ are from random samples that are **independent**, and sampled from **normal** distributions with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2, respectively, and sample sizes n_1 and n_2, then

$$\overline{Y_1} - \overline{Y_2} \sim \text{Normal}\begin{pmatrix} \text{mean} = \mu_1 - \mu_2 \\ \text{variance} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \end{pmatrix}$$

Difference of 2 Means

Independent samples, population has a bell-shaped distribution.

$$\frac{(\overline{Y_1} - \overline{Y_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \left\{ \begin{array}{c} \text{Standard Normal} \\ (Z) \text{ Distribution} \end{array} \right\}$$

Using S_1^2 & S_2^2 to estimate σ_1^2 & σ_2^2:

$$\frac{(\overline{Y_1} - \overline{Y_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim \left\{ \begin{array}{c} \text{T-distribution} \\ \text{with } \gamma \text{ deg. of freedom} \end{array} \right\}$$

$$\gamma \cong \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2 - 1}}$$

Difference of 2 Means: Example

Cold-rolled steel:

A random sample of $n_1 = 38$ gives an average strength of $\overline{y_1} = 29.8$ksi, with $s_1^2 = 16$

Galvanized steel:

A random sample of $n_2 = 32$ gives an average strength of $\overline{y_2} = 34.7$ksi, with $s_2^2 = 25$

1. Null Hypothesis H_0:

$\mu_1 - \mu_2 = 0 \quad (\mu_1 \geq \mu_2)$

Alternative Hypothesis H_A:

$\mu_1 - \mu_2 < 0 \quad (\mu_1 < \mu_2)$

Difference of 2 Means Ex., Cont.

2. Select the test statistic:

$$\frac{\overline{Y_1} - \overline{Y_2} - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim \text{t-distribution}$$

with

$$\begin{aligned}
\gamma &= \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1-1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2-1}} \\
&= \frac{\left(\frac{16}{38} + \frac{25}{32}\right)^2}{\frac{\left(\frac{16}{38}\right)^2}{37} + \frac{\left(\frac{25}{32}\right)^2}{31}} \\
&\approx 59 \;\; \text{(round down)}
\end{aligned}$$

Difference of 2 Means Ex., Cont.

4. Calculate the value of the test statistic:

$$\frac{(\overline{Y_1} - \overline{Y_2}) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

$$= \frac{(29.8 - 34.7) - 0}{\sqrt{\frac{16}{38} + \frac{25}{32}}}$$

(null hypothesis is $\mu_1 - \mu_2 = 0$)

$$= -4.4688$$

Difference of 2 Means Ex., Cont.

5. Decision:

F	$\cdots$	0.90	0.95	$\cdots$	0.995	0.9995
γ						
40		1.303	1.684		2.704	3.551
60		1.296	1.671		2.660	3.460
$\vdots$						
∞		1.282	1.645		2.576	3.291

We can reject the null hypothesis at a **very** high level of significance (P-value less than 0.0005). Therefore we conclude that with a very high probability, galvanized steel is stronger than cold-rolled steel.

Confidence Interval for $\mu_1 - \mu_2$

$$\text{Assuming} \quad \frac{(\overline{Y_1} - \overline{Y_2}) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_1^2}{n_1}}}$$

has a t-distribution, upper & lower confidence limits for $100(1-\alpha)\%$ confidence, are given as:

$$\overline{y_1} - \overline{y_2} - t_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \quad , \quad \overline{y_1} - \overline{y_2} + t_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

Example: For steel example, a 99% confidence interval would give $\alpha = 0.01$, so $t_{\alpha/2} = t_{0.005} = 2.704$. (From before, $\gamma = 59$, but I used $\gamma = 40$ in the tables to be conservative.)

$$\begin{aligned} \overline{y_1} - \overline{y_2} \pm (2.704)\sqrt{\frac{16}{38} + \frac{25}{32}} &= (29.8 - 34.7) \pm 2.96 \\ &= -4.9 \pm 2.96 \end{aligned}$$

$-7.86 \leq \mu_1 - \mu_2 \leq -1.94$ with 99% confidence.

Pooled t-test

To compare $\mu_1 - \mu_2$, assuming:
independent samples
normal populations
equal variances, $\sigma_1^2 = \sigma_2^2$

$$\frac{\overline{Y_1} - \overline{Y_2} - (\mu_1 - \mu_2)}{\sqrt{S_p^2(\frac{1}{n_1} + \frac{1}{n_2})}} \sim \left\{ \begin{array}{c} t\text{-distribution with} \\ n_1 + n_2 - 2 \\ \text{degrees of freedom} \end{array} \right\}$$

$$\text{where } S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

Comparing Two Proportions

Let:
$\hat{p_1}$ be the proportion from pop. 1, and
$\hat{p_2}$ be the proportion from pop. 2.

For large sample sizes (n_1 & n_2):

$\hat{p_1} = \dfrac{Y_1}{n_1}$ and $\hat{p_2} = \dfrac{Y_2}{n_2}$ and

$\widehat{p_1 - p_2} = \hat{p_1} - \hat{p_2} = \dfrac{Y_1}{n_1} - \dfrac{Y_2}{n_2}$

For large sample sizes, the estimator $\hat{p_1} - \hat{p_2}$ is approx. normal with mean $p_1 - p_2$ and variance $\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}$.

The $100(1-\alpha)\%$ C.I. for $\hat{p_1} - \hat{p_2}$ is:

$$(\hat{p_1} - \hat{p_2}) \pm z_{\alpha/2}\sqrt{\frac{\hat{p_1}(1-\hat{p_1})}{n_1} + \frac{\hat{p_2}(1-\hat{p_2})}{n_2}}$$

Comparing Two Proportions: Hypothesis Testing

Three Forms:

1. H_0: $p_1 - p_2 = (p_1 - p_2)_0$ right-tail test
   H_1: $p_1 - p_2 > (p_1 - p_2)_0$
2. H_0: $p_1 - p_2 = (p_1 - p_2)_0$ left-tail test
   H_2: $p_1 - p_2 < (p_1 - p_2)_0$
3. H_0: $p_1 - p_2 = (p_1 - p_2)_0$ two-tail test
   H_3: $p_1 - p_2 \neq (p_1 - p_2)_0$

We use $\dfrac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)_0}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}$

as an approx. standard normal.

Comparing Variances

Hypothesis tests on two pop. variances:

Right-tailed: H_0: $\sigma_1^2 = \sigma_2^2$
H_A: $\sigma_1^2 > \sigma_2^2$
Left-tailed: H_0: $\sigma_1^2 = \sigma_2^2$
H_A: $\sigma_1^2 < \sigma_2^2$
Two-tailed: H_0: $\sigma_1^2 = \sigma_2^2$
H_A: $\sigma_1^2 \neq \sigma_2^2$

Test Statistic: S_1^2/S_2^2

- If $\sigma_1^2 = \sigma_2^2$, then S_1^2/S_2^2 is close to 1.
- If $\sigma_1^2 < \sigma_2^2$, then S_1^2/S_2^2 is close to 0.
- If $\sigma_1^2 > \sigma_2^2$, then S_1^2/S_2^2 is much larger than 1.

$$S_1^2/S_2^2 \sim \left\{ \begin{array}{c} F\text{-distribution} \\ \text{(ratio of 2 indep. } \chi\text{-squared r.v.s)} \\ \text{with } \gamma_1 = n_1 - 1 \text{ and } \gamma_2 = n_2 - 1 \\ \text{assuming null hyp. is true, } \sigma_1^2 = \sigma_2^2 \end{array} \right\}$$

Comparing Variances

- The test is not very powerful (often fail to reject the null Hypothesis $\sigma_1^2 = \sigma_2^2$, when indeed the variances are different).
- The test performs best when sample sizes are equal ($n_1 = n_2$) & large.
- The test is very sensitive to the normality assumption. If a histogram is not bell-shaped, do not use the test.

Comparing Variances Example

We wish to determine whether there is less variability in the silver plating done by Company 1 than Company 2. Independent samples of work done by the two companies yield:

Company 1: $n_1 = 16$, $s_1 = 0.035$mil
Company 2: $n_2 = 10$, $s_2 = 0.062$mil

Test the hypothesis at 0.05 sig level.

H_0: $\sigma_1^2 = \sigma_2^2$ $\qquad$ $(\sigma_1^2 = \sigma_2^2)$
H_A: $\sigma_1^2 < \sigma_2^2$

Determine test stat. & dist.

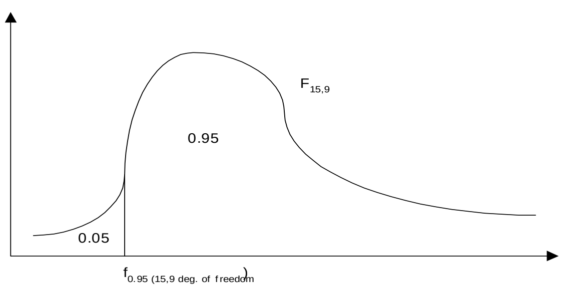
S_1^2/S_2^2 $\qquad$ F-distribution
$\gamma_1 = n_1 - 1 = 15$ $\qquad$ $\gamma_2 = n_2 - 1 = 9$

Comparing Variances Example Cont.

Calculate test statistic from data;

$$s_1^2/s_2^2 = 0.035^2/0.062^2 = 0.319$$

Rejection region:



Reciprocal relationship:

$$f_{0.95}(15, 9 \text{ df}) = \frac{1}{f_{0.5}(9, 15 \text{ df})}$$

and $f_{0.5}(9, 15 \text{ df}) = 2.59$

so $f_{0.95}(15, 9 \text{ df}) = \dfrac{1}{2.59} = 0.3861$

Comparing Variances Example Cont.

Decision: The test statistic 0.319 is less than 0.3861, so we **reject** the null hypothesis. The data support the conclusion that the silver plating done by Company 1 is less variable than Company 2.

Comparing Means: Paired Data

Before, we had a random sample of size n_1 from population 1, and an **independent** random sample of size n_2 from population 2.

Now, we make n observations, and collect two types of data per observation.

Example: Sample 6 river locations & measure zinc concentration in bottom water and surface water. Does the data suggest that the average concentration in bottom water exceeds that of surface water?

Location	1	2	3	4	5	6
Zinc bot, y_1	0.430	0.266	0.567	0.531	0.707	0.716
Zinc surf, y_2	0.415	0.238	0.390	0.410	0.605	0.609

Comparing Means: Paired Data Ex.

Two different methods (finite element method, and a new approximation method) are being compared to predict buckling load of a certain structure under 10 different conditions. Is the average difference equal to zero? Are the 2 methods consistent?

The Paired t-test

The data consist of n pairs of observations;

$$(X_1, Y_1), (X_2, Y_2), \cdots, (X_n, Y_n)$$

Let $D_i = \text{difference} = X_i - Y_i$ for $i = 1, 2, \ldots, n$
We can get $D_1, D_2, \cdots, D_n$ and
$\overline{D} = \Sigma_{i=1}^{n} D_i / n$

$$\text{and } \frac{\overline{D} - 0}{S_d / \sqrt{n}} \sim \left\{ \begin{array}{c} t\text{-distribution} \\ \text{with n-1 degrees of freedom} \end{array} \right\}$$

$$\begin{aligned} S_d &= \text{sample standard deviation} \\ &= \sqrt{S_d^2} = \sqrt{\frac{\Sigma_{i=1}^{n} (D_i - \overline{D})^2}{n - 1}} \end{aligned}$$

$100(1 - \alpha)\%$ Confidence limits on $\mu_X - \mu_Y$ for paired data:

$$\overline{D} \pm t_{\alpha/2} \left(\frac{S_D}{\sqrt{n}} \right)$$

Paired t-test Example

$\mu_D = \mu_{\text{Bottom}} - \mu_{\text{Surface}}$
H_0: $\mu_D = 0$
H_A: $\mu_D > 0$
(i.e., more zinc conc. in bottom water than surface)

Test Statistic: $$\frac{\overline{d}}{S_d/\sqrt{n}}$$

Data: $\overline{d} = 0.0917$, $s_d^2 = 0.003683$, $n = 6$

$$\frac{\overline{d}}{s_d/\sqrt{n}} = 3.70$$

Paired t-test Example Cont

For t-test, using $5 = n$ - 1 degrees of freedom, $t_{0.01} = 3.365$. Since $3.70 > 3.365$, we reject the null hypothesis at a 0.01 level of significance. We conclude that zinc concentration in bottom water does exceed zinc concentration in surface water.

$p_1 - p_2$ Hypothesis Test

In the 1954 Salk polio vaccine experiment, one group had a placebo & one group had the new vaccine.

Let:
p_1 = probability of getting paralytic polio for control group
p_2 = probability of getting paralytic polio for vaccinated group

H_0: $p_1 - p_2 = 0$
H_A: $p_1 - p_2 > 0$

(i.e., a vaccinated child is less likely to contract polio than an unvaccinated child.)

$p_1 - p_2$ Hypothesis Test Cont

Find sample size if $\alpha = 0.05$ and $\beta = 0.1$, when $p_1 = 0.0003$ and $p_2 = 0.00015$

$$\begin{aligned} n &= \frac{\left(z_\alpha \sqrt{(p_1 + p_2)(q_1 + q_2)/2} + z_\beta \sqrt{p_1 q_1 + p_2 q_2}\right)^2}{d^2} \\ &\approx 171,400 \end{aligned}$$

where $d = p_1 - p_2 = 0.0003 - 0.00015$

$z_\alpha = z_{0.05} = 1.645$
$z_\beta = z_{0.1} = 1.28$

$p_1 - p_2$ **Hypothesis Test Cont**

Actual data: Placebo, $n_1 = 201,229$

$y_1 = 110$ polio cases

Vaccine, $n_2 = 200,745$

$y_2 = 33$ polio cases

Calculate $z = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}(\frac{1}{n_1} + \frac{1}{n_2})}} \approx 6.47$

where $\hat{p}_1 = \dfrac{110}{201,229}$, $\hat{p}_2 = \dfrac{33}{200,745}$

$$\hat{p} = \frac{y_1 + y_2}{n_1 + n_2} = \frac{110 + 33}{201,229 + 200,745}$$

The p-value is less than 0.0003, so we reject the null hypothesis ($p_1 - p_2 = 0$), and conclude that the probability of contracting polio with the vaccine is different from the control group.