STEVENS INSTITUTE of TECHNOLOGY

School of Business

Amir H Gandomi; PhD
Assistant Professor
Stevens Institute of Technology
a.h.gandomi@stevens.edu
https://web.stevens.edu/facultyprofile/?id=2273

Multivariate Data Analysis – BIA 652

Class 1 – Review of Probability

Course Information

- Text:
  - Practical Multivariate Analysis, 5th Ed., by A.A. Afifi, S.J. May, and V.A. Clark, Chapman Hall, New York: 2012
  - http://www.crcpress.com/product/isbn/9781439816806
- SAS, R, Python Bootcamps: Expected in the first weeks
- SAS tutorial should be available on CANVAS
- SAS Loading on Laptops: at Hanlon Lab

Course Information

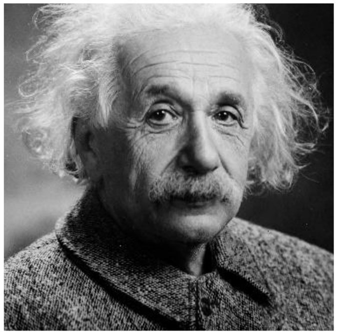
- Grading Assistant: TBD, On Canvas
- Office: Babbio 616
- Office Hours:
  - TBD
  - Or By Appointment
- Goal for next week: Download and examine the text datasets: Depression, Lung Function, and Chemical Companies Studies.
  - statistics.ats.ucla.edu/stat/examples/pma5/default.htm

Course Information

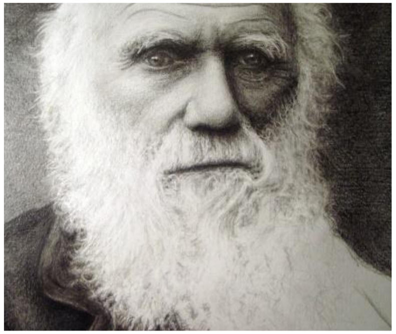
- Grades:
  - Team Written and Oral Project . (45%)
    - Written a Term Paper.
    - Orals Presented (15 min) + Q&A (5 min).
    - Groups of $\leq$ 3 Acceptable.
    - Submission of proposed teams by September 12th (format on next slide).
  - Final Exam – Probably 3rd from last class (30%)
  - Class Participation and Homework (25%)
    - Homework Due At Assigned Class (will be posted on Canvas)
    - Only Accepted one day after due date with 10% deduction

Team Pictures
Teams No More Than 3
(A picture of you, and your name)

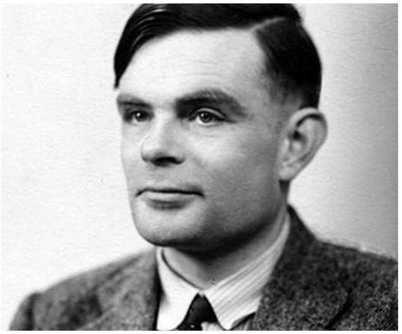
Albert Einstein



Charles Darwin



Alan Turing



Project Grading

- **Grades:**
  - **Presentation:**
    - **Knowledge**
    - **Depth**
    - **Clarity**
    - **Content**
  - **Slides**
  - **Paper**
    - **Abstract/Introduction/Objectives**
    - **Exploration of Data**
    - **Methods**
    - **Results**
    - **Conclusion**
    - **Reference**

Ethics Statement

The following statement is printed in the Stevens Graduate Catalog and applies to all students taking Stevens courses, on and off campus.

"Academic Improprieties

The term academic impropriety is meant to include, but is not limited to, cheating on homework, during in-class or take home examinations and plagiarism. The Institute has adopted a procedure to deal with such actions. An instructor of a graduate course may elect to formally charge a student with committing an academic impropriety to the Dean of Graduate Academics or to adjudicate the issue personally."

Consequences of academic impropriety are severe, ranging from receiving an "F" in a course, to a warning from the Dean of the Graduate School, which becomes a part of the permanent student record, to expulsion.

Reference: *https://www.stevens.edu/provost/graduate-academics/handbook/academic-standing.html#PDG*

Ethics Pledge

Consistent with the above statements, all homework exercises, tests and exams that are designated as individual assignments MUST contain the following signed statement before they can be accepted for grading.

____

I pledge on my honor that I have not given or received any unauthorized assistance on this assignment/examination. I further pledge that I have not copied any material from a book, article, the Internet or any other source except where I have expressly cited the source.

Signature ________________________ Date: _____________

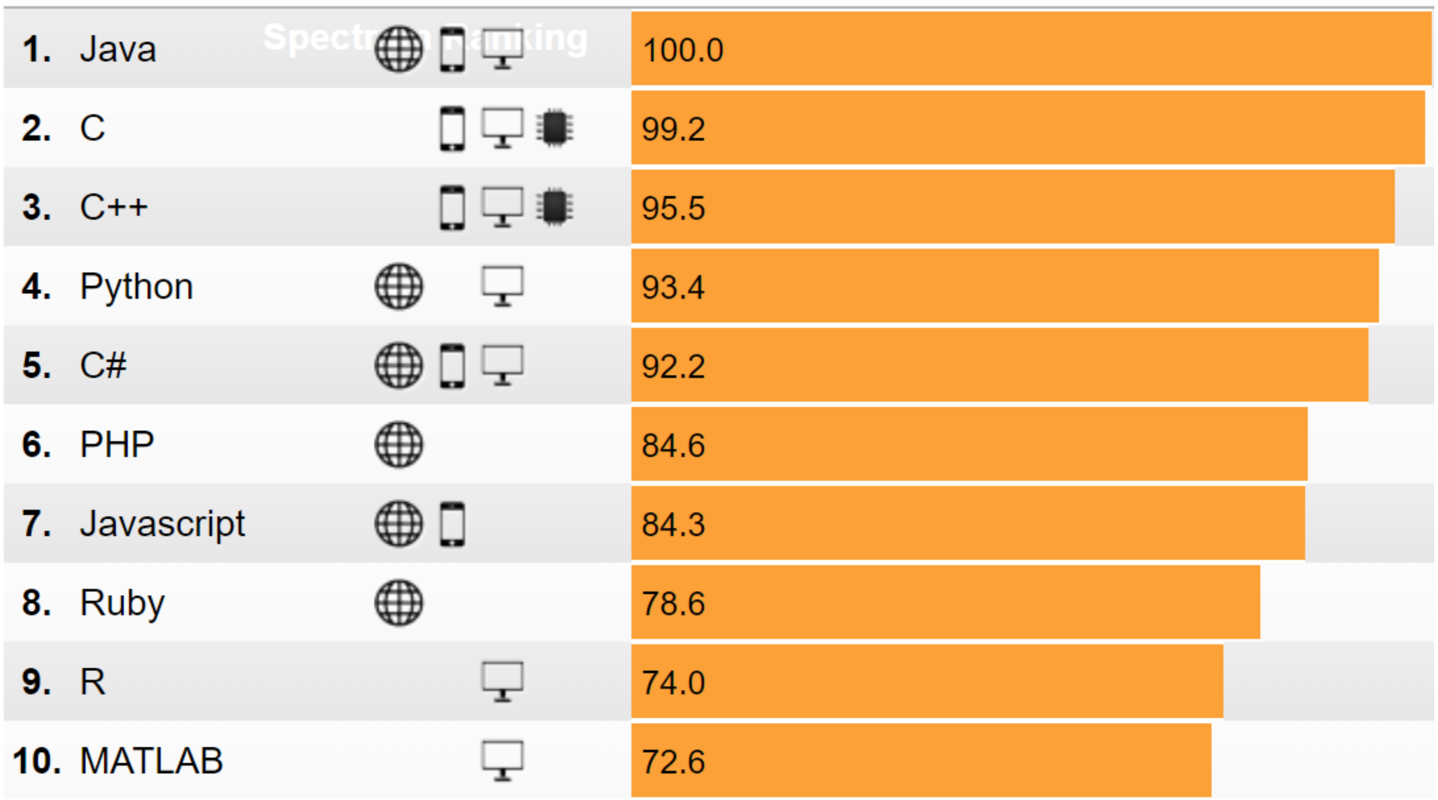
Please note that assignments in this class may be submitted to www.turnitin.com, a web-based anti-plagiarism system, for an evaluation of their originality.

____

Class 1
Intro + Review

- Intro to the purpose of Multivariate Analysis, with examples
- Review of Univariate Statistics.
- Homework for next meeting:
  – Get SAS, R, or Python loaded on your computer
  – Go to Bootcamp for SAS, R, and Python (poll class)
  – I am going to provide you free access to DataCamp
  – Look at primary datasets

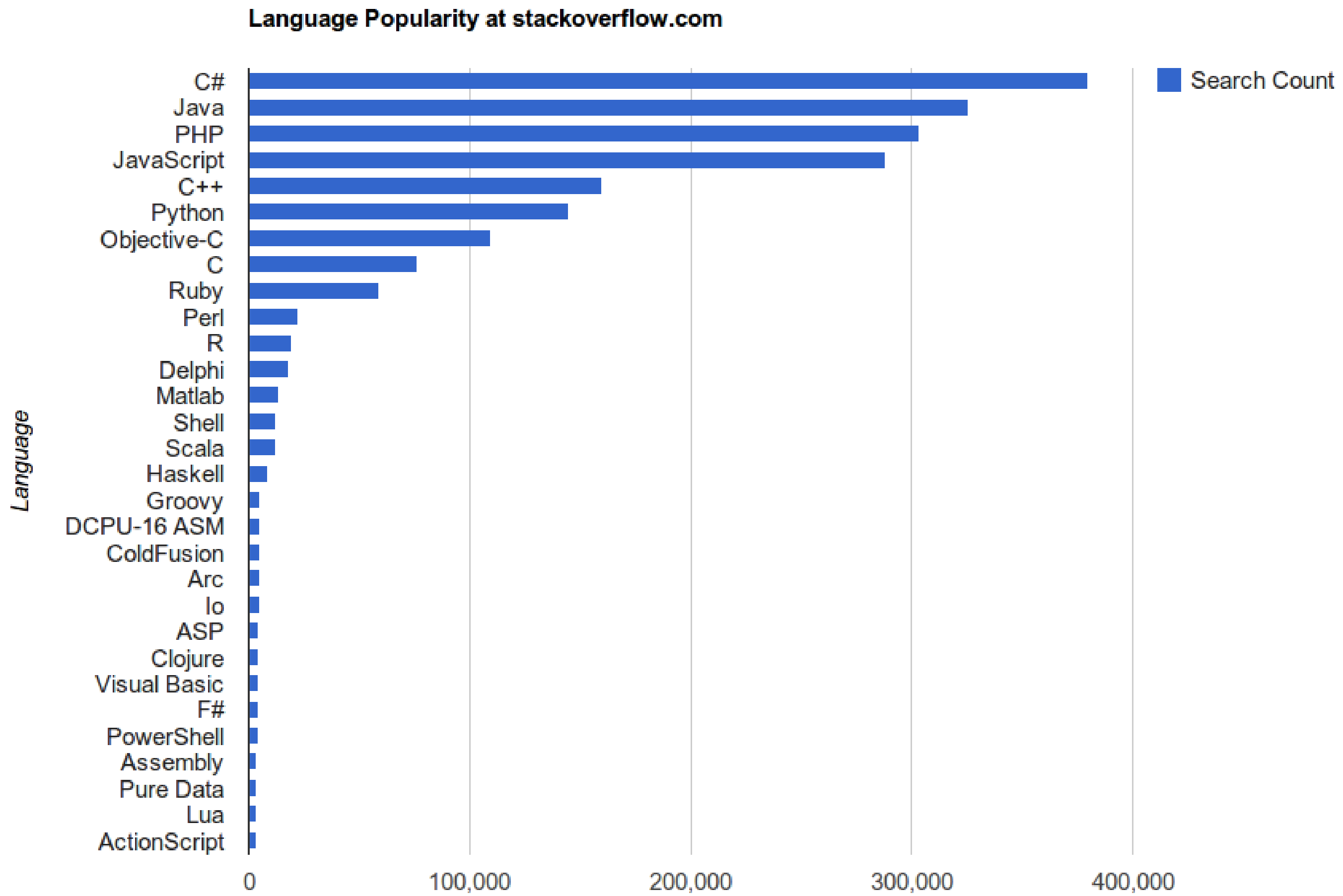
IEEE Top 10 Programming Language



http://spectrum.ieee.org/static/interactive-the-top-programming-languages#index/1/1/1/1/1/25/1/25/1/50/1/25/1/25/1/50/1/25/1/25/1/100/1/100/1/25/1/40/

Language Rank Types Jobs Ranking

1. Java 100.0
2. C 99.8
3. C++ 92.8
4. Python 91.8
5. C# 91.0
6. Javascript 90.6
7. PHP 86.1
8. SQL 84.2
9. Ruby 82.9
10. HTML 79.1
11. Shell 78.4
12. PERL 77.9
13. Objective-C 72.4
14. Visual Basic 71.3
15. MATLAB 67.5
16. Assembly 64.8
17. ASP.NET 62.2
18. R 61.1
19. Scala 60.3
20. SAS 54.5



Some Multivariate Datasets

- https://archive.ics.uci.edu/ml/datasets.html
- http://kaggle.com
- http://www.rdatamining.com/resources/data
- http://statistics.ats.ucla.edu/stat/examples/pma5/default.htm
- https://opendata.socrata.com/
- http://hadoopilluminated.com/hadoop_illuminated/hadoop-illuminated.pdf (Chapter 16)
- https://www.linkedin.com/pulse/ten-sources-free-big-data-internet-alan-brown

Course Topics

Topics	Text Reference
Mulitvariate Overview/Univariate Review	Slides
Univariate Review/Introduction	Slides/Chapters 1 – 5, +
Regression & Correlation	Chapter 6
Multiple Regression & Correlation	Chapter 7, +
PCA & Factor Analysis	Chapter 14, 15
Cluster Analysis	Chapter 16
Discriminant Analysis	Chapter 11
Logistic Regression	Chapter 12
Naïve Bayes, ANOVA, MANOVA, Multi Dimensional Scaling (as time permits)	Outside References

Multivariate Analytic Techniques
This Course

- Regression
  - Multiple Linear Regression
- Classification
  - Decision Trees
  - Support Vector Machines
  - Naïve Bayes
  - Logistic Regression
  - K Nearest Neighbor
  - Ensemble – e.g. Adaboost
- Dimension Reduction
  - Principal Component Analysis
  - Factor Analysis
- Clustering
  - K-Means
  - Hierarchical
  - Density

Example of Multivariate Data Simple

Inputs (Independent Variables)					Outputs (Dependent Variables)
Individual	Age	Gender	Height	Weight	Health Code

Regression:
- Continues Outcome
- Predicting Response

Classification:
- Class Labels
- Predicting Membership

Example of Multivariate Data Wrong-Way Driving Crashes

Season	Time of Day	Dry Roadway Surface	Dark Not Lit	Clear Weather	DUI Driver	Seat Belt Used	Injury Severity
Winter	Night	No	No	No	N/A	Yes	Severe Injury
Summer	Evening	Yes	Yes	Yes	Yes	Yes	Minor Injury
Winter	Afternoon	Yes	No	Yes	No	No	No Injury

Multivariate analysis

- Concerned with datasets that have more than one response variables for each observational unit
- N rows (cases) and P columns (variables)
  - Relationships among cases
  - Relationships among variables
- First, visualize
  - Pairs plot – plot scatter plot matrix
  - pairs plots can easily miss interesting structure
  - multivariate methods explore the data in a less coordinate-dependent way

+

The number of datasets to analyze: one or two

- One dataset
  - Find relationships among variables or cases
    - Principal component analysis (PCA) – continuous variables
    - Correspondence analysis (CA) – categorical variables
    - Multidimensional scaling (MDS) and cluster analysis - proximity
- Two datasets
  - Find relationships between independent variable set and dependent variable set, or between two dependent variable sets
    - Multiple linear regression – not really multivariate, just one y
    - Discriminant analysis (DA) and logistic regression – categorical DV
    - Multivariate analysis of variance (MANOVA) – math same as DA
    - Confirmatory factor analysis – create models (SEM)
    - Exploratory factor analysis – no constraints on models
    - Canonical correlation analysis (CC) – extension of multiple linear regression
    - Multiple factor analysis – perform PCA on each data table

Principal component analysis (PCA)

- Interval or ratio level of measurement
  - Nominal scale: No ordering of values (e.g., Male or female)
  - Ordinal scale: Can infer ordering of values (e.g., low, medium, and high self-esteem)
  - Interval scale: Can infer ordering of values, the values are evenly spaced (e.g., Celcius scale, intelligence tests)
  - Ratio scale: Same as interval scale but zero is special, can make ratio (e.g., Weight, Kelvin scale)

- The goal of PCA is to decompose a data table with correlated measurements into a new set of uncorrelated (i.e., orthogonal) variables.
  - Step 1: Subtract the mean
  - Step 2: Calculate the covariance matrix (or correlation)
  - Step 3: Calculate the eigenvectors and eigenvalues
  - Step 4: Choose components and form a feature vector
  - Step 5: Derive the new data

- The importance of each component is expressed by the variance (i.e., eigenvalue) of its projections or by the proportion of the variance explained.

- PCA is useful when you want to develop an index to summarizes complex data - rank students by examination scores, rank cities by cost of living

- PCA is useful for visualization when there are too many explanatory variables relative to the number of observations and when the explanatory variables are highly correlated.

Multidimensional scaling (MDS), additive tree, cluster analysis

- Similarity or distance
- Applied when the rows and the columns of the data table represent the same units and when the measure is a distance or a similarity.
- The goal of the analysis is to represent an observed proximity matrix geometrically.
- MDS is used to represent the units as points on a map such that their Euclidean distances on the map approximate the original similarities
  - Classic MDS, which is equivalent to PCA, is used for distances
  - Non-metric MDS for similarities
- Additive tree analysis and cluster analysis are used to represent the units as "leaves" of a tree with the distance "on the tree" approximating the original distance or similarity.
  - Hierarchical clustering: a series of partitioning steps from a single cluster containing everyone to n clusters each containing a single individual.
  - K-means clustering: find initial partitioning and start moving individuals.

Multiple linear regression analysis

- Several IVs are used to predict one DV.
  - $y = b0 + b1x1 + b2x2 + b3x3 \ldots + e$
  - Generalization of simple linear regression
  - When the IVs are orthogonal, the problem reduces to a set of univariate regressions.
  - When the IVs are correlated, their importance is estimated from the partial coefficient of correlation.

Discriminant analysis (DA)

- Predicting a nominal variable
  - Used to classify a case into one of two or more populations.
  - You need to know which population the individual belongs to for the initial sample.
  - You classify future individuals whose membership is unknown (prediction).
  - You identify which variables contribute to making the classification (description).
- Mathematically equivalent to MANOVA
- Used when a set of IVs are used to predict the group to which a given unit belongs (a nominal DV).
  - Deciding whether to approve a loan - age, income, marital status, outstanding debt, home ownership, etc..
  - Deciding whether an individual is more or less likely to be depressed - age, income, education, etc..

Logistic regression

- Binomial: for binary or dichotomous response variable, 0 and 1, such as presence and absence, success and failure, etc..
- Multinomial: for more than two categories.
- As in linear regression, we are looking for a relationship between our response variable and a set of independent variables, which are often called covariates.
- We can transform the output of a linear regression to be suitable for probabilities
  - logit(p) = log(o) = log(p/(1-p)) = b0 + b1x1 + b2x2 + ...
  - The odds: p =.8, then the odds = .8 / .2 = 4 to 1 → 4 times as likely
  - odds = p / (1 – p), p = odds / (1 + odds)
  - The inverse of the logit function is the logistic function.
    - if logit(p) = z, then p = exp(z) / (1 + exp(z))
- Some examples
  - Depression level (0 or 1)
  - People's occupational choices
  - Brand choices based on gender and age.

Multivariate analysis of variance

- IVs have the same structure as in a standard ANOVA
- Used to predict a set of DVs.
- MANOVA computes a series of ordered orthogonal linear combinations of the DVs (i.e., factors) with the constraint that the first factor generates the largest F if used in an ANOVA.

+
Factor analysis

- Examines the interrelationships among a large number of variables to determine underlying dimensions (factors)
  - Assumption: there are some underlying common factors (e.g., intelligence, social class, social/soft drug)
  - These factors cannot be directly observed – latent variables – concepts that cannot be measured directly
  - Measurable variables – manifest variables – are expected to be related to the latent variables
  - Looks at the relationships between assumed latent variables and the manifest variables.

- Confirmatory
  - Fitting a model
  - Generates one or a few models of an underlying explanatory structure, which is often expressed as a graph (how IVs related to DVs).
  - Then the correlations among the DV's are fit to this structure.
  - Models are evaluated by comparing how well they fit the data.
  - Structural equation modeling (SEM)

- Exploratory
  - No constraints on which of the manifest variables load on the common factors

Canonical correlation analysis

- Extension of multiple regression – more than one y
  - x = [x1, x2, ..., xq1], y = [y1, y2, ..., yq2]
  - R11 = cor matrix of variables in x
  - R22 = cor matrix of variables in y
  - R12 = cor between x and y (q1 by q2 matrix)
  - E1 = R11inv R12 R22inv R21
- CC combines the DVs to find pairs of new variables (called canonical variables, one for each data table) which have the highest correlation.
- CV's, even when highly correlated, do not necessarily explain a large portion of the variance of the original tables.
  - This make the interpretation of the CV sometimes difficult, but CC is nonetheless an important theoretical tool

Review of Univariate Statistics and a bit of Matrix Algebra

Review Probability

Statistics for Business and Economics

8th Edition

(Text Used in MBA 620)

Chapter 3

Probability

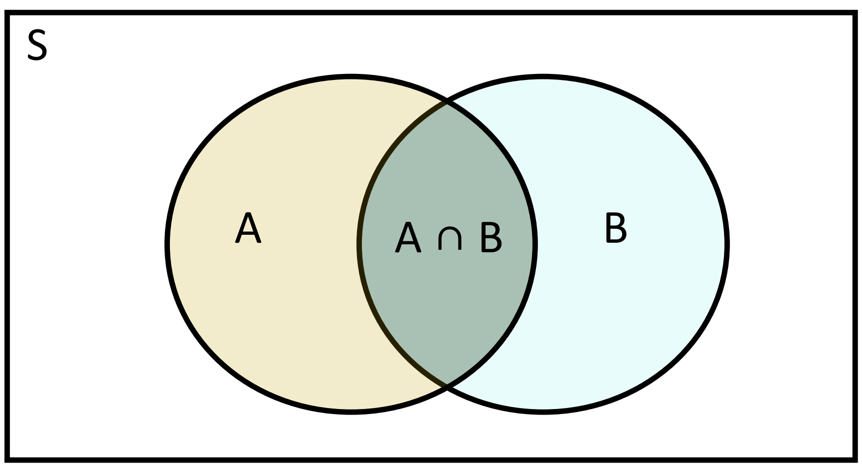
Important Terms

- Random Experiment – a process leading to an uncertain outcome
- Basic Outcome – a possible outcome of a random experiment
- Sample Space (S) – the collection of all possible outcomes of a random experiment
- Event (E) – any subset of basic outcomes from the sample space

Important Terms

(continued)

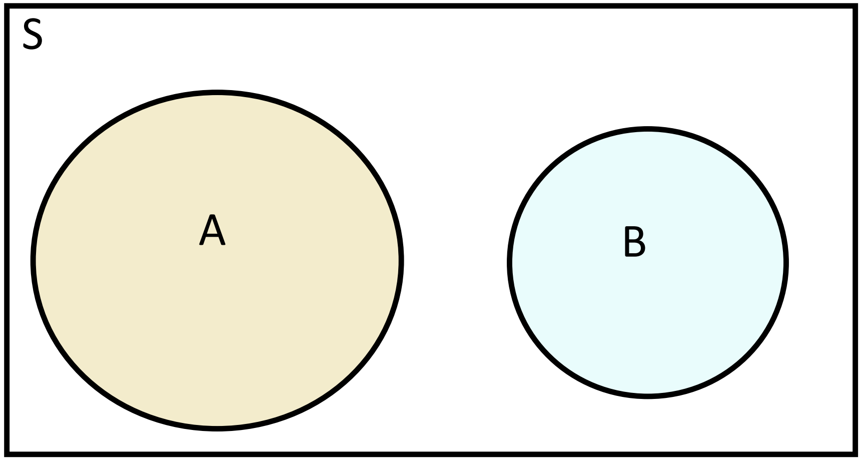
- Intersection of Events – If A and B are two events in a sample space S, then the intersection, A ∩ B, is the set of all outcomes in S that belong to both A and B



Important Terms

(continued)

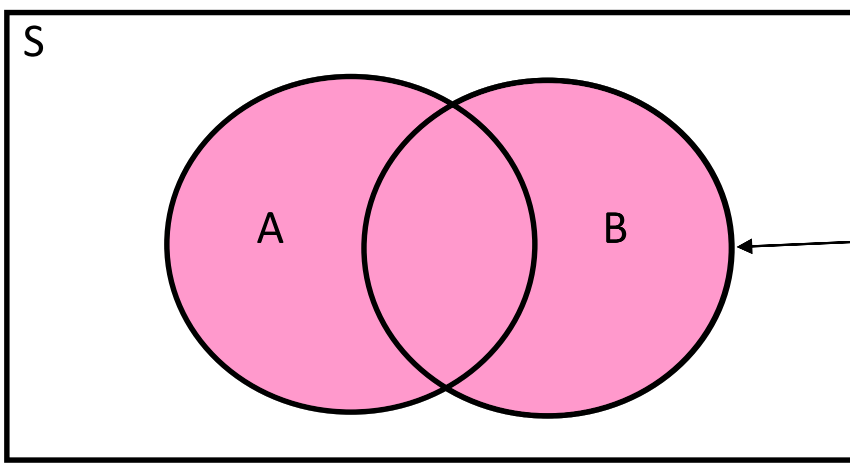
- A and B are Mutually Exclusive Events if they have no basic outcomes in common
  - i.e., the set $A \cap B$ is empty



Important Terms

(continued)

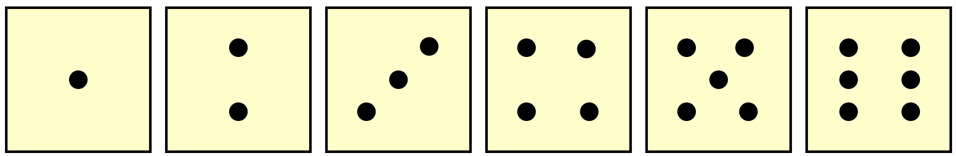
- Union of Events – If A and B are two events in a sample space S, then the union, A U B, is the set of all outcomes in S that belong to either A or B



The entire shaded area represents A U B

Examples

Let the Sample Space be the collection of all possible outcomes of rolling one dice:



S = [1, 2, 3, 4, 5, 6]

Let A be the event "Number rolled is even"

Let B be the event "Number rolled is at least 4"

Then

A = [2, 4, 6] and B = [4, 5, 6]

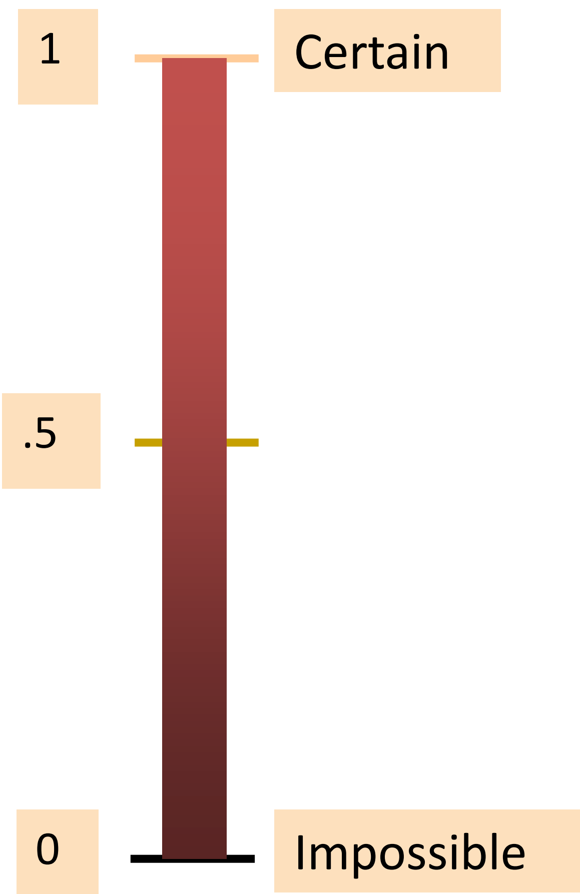
A ∩ B = [4, 6]

A U B = [2, 4, 5, 6]

3.2 Probability and Its Postulates

- **Probability** – the chance that an uncertain event will occur (always between 0 and 1)

$0 \le P(A) \le 1$ For any event A



Classical Probability

- Assumes all outcomes in the sample space are equally likely to occur

Classical probability of event A:

$$P(A) = \frac{N_A}{N} = \frac{\text{number of outcomes that satisfy the event A}}{\text{total number of outcomes in the sample space}}$$

- Requires a count of the outcomes in the sample space

Counting the Possible Outcomes

- Use the Combinations formula to determine the number of combinations of n items taken k at a time

$$C_k^n = \frac{n!}{k!(n-k)!}$$

- where
  - n! = n(n-1)(n-2)…(1)
  - 0! = 1 by definition

Permutations and Combinations

The number of possible orderings

- The total number of possible ways of arranging x objects in order is

$$x! = x(x-1)(x-2)\ldots(2)(1)$$

- x! is read as "x factorial"

Permutations and Combinations

(continued)

Permutations: the number of possible arrangements when x objects are to be selected from a total of n objects and arranged in order [with (n – x) objects left over]

$$P_x^n = n(n-1)(n-2)\ldots(n-x+1)$$

$$= \frac{n!}{(n-x)!}$$

Permutations and Combinations

(continued)

- Combinations: The number of combinations of x objects chosen from n is the number of possible selections that can be made

$$C_k^n = \frac{P_x^n}{x!}$$

$$= \frac{n!}{x!(n-x)!}$$

Permutations and Combinations Example

Suppose that two letters are to be selected from A, B, C, D and arranged in order. How many permutations are possible?

- Solution The number of permutations, with

  $n = 4$ and $x = 2$, is $$P_2^4 = \frac{4!}{(4-2)!} = 12$$

- The permutations are

  AB AC AD BA BC BD

  CA CB CD DA DB DC

Permutations and Combinations Example

(continued)

Suppose that two letters are to be selected from A, B, C, D. How many combinations are possible (i.e., order is not important)?

- Solution The number of combinations is

$$C_2^4 = \frac{4!}{2!\,(4-2)!} = 6$$

- The combinations are

AB (same as BA)
AC (same as CA)
AD (same as DA)
BC (same as CB)
BD (same as DB)
CD (same as DC)

Probability Postulates

1. If A is any event in the sample space S, then

$$\boxed{0 \le P(A) \le 1}$$

2. Let A be an event in S, and let O_i denote the basic outcomes. Then

$$\boxed{P(A) = \sum_{A} P(O_i)}$$

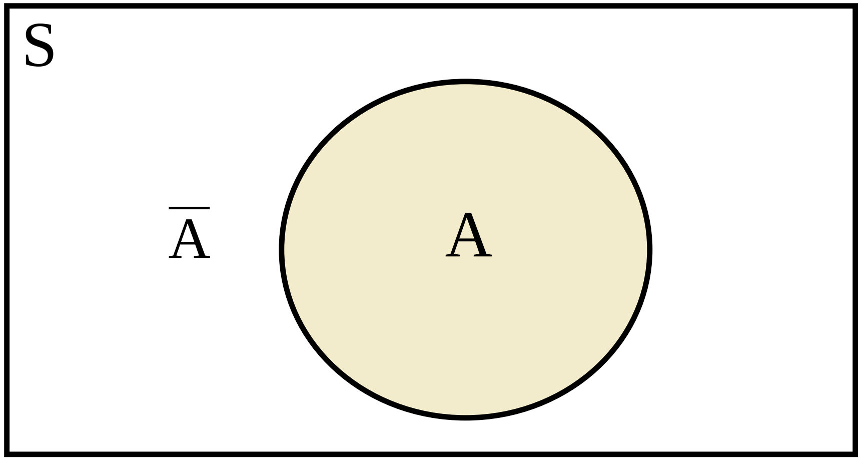
3. (the notation means that the summation is over all the basic outcomes in A)

$$\boxed{P(S) = 1}$$

Probability Rules

- The Complement rule:

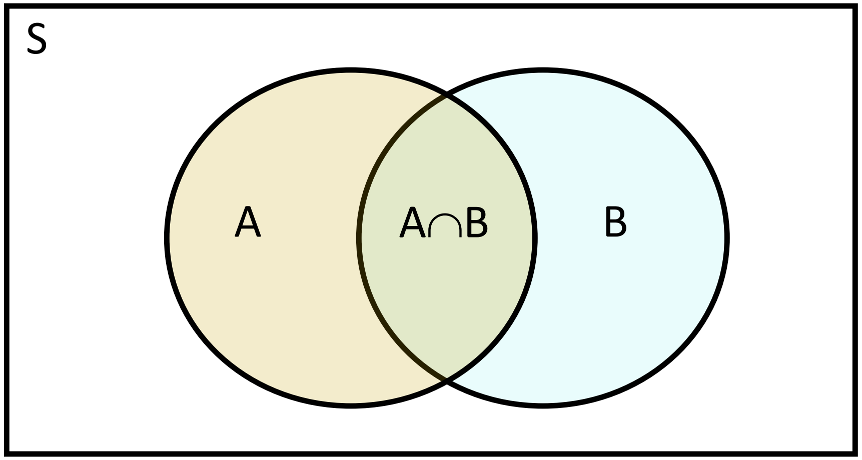
$$\boxed{P(\overline{A}) = 1 - P(A)}$$ i.e., $P(A) + P(\overline{A}) = 1$



Probability Rules

- The Addition rule:
  - The probability of the union of two events is

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



A Probability Table

Probabilities and joint probabilities for two events A and B are summarized in this table:

	B	$\overline{B}$	
A	$P(A \cap B)$	$P(A \cap \overline{B})$	$P(A)$
$\overline{A}$	$P(\overline{A} \cap B)$	$P(\overline{A} \cap \overline{B})$	$P(\overline{A})$
	$P(B)$	$P(\overline{B})$	$P(S) = 1.0$

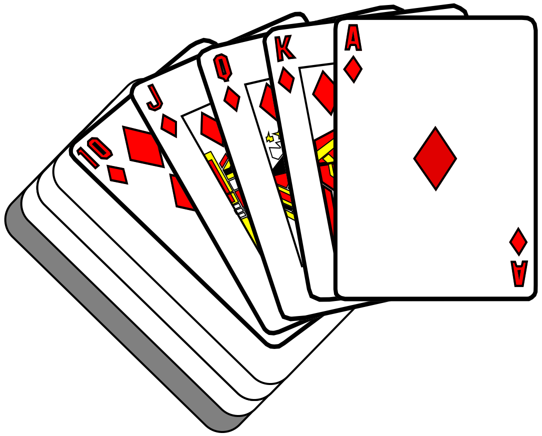
Addition Rule Example

Consider a standard deck of 52 cards, with four suits:

♥ ♣ ♦ ♠

Let event **A** = card is an Ace

Let event **B** = card is from a red suit



Addition Rule Example

(continued)

P(Red U Ace) = P(Red) + P(Ace) - P(Red ∩ Ace)

= 26/52 + 4/52 - 2/52 = 28/52

Type	Color		Total
	Red	Black	
Ace	2	2	4
Non-Ace	24	24	48
Total	26	26	52

Don't count the two red aces twice!

Conditional Probability

- A conditional probability is the probability of one event, given that another event has occurred:

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

The conditional probability of A given that B has occurred

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

The conditional probability of B given that A has occurred

Conditional Probability Example

- Of the cars on a used car lot, 70% have air conditioning (AC) and 40% have a CD player (CD). 20% of the cars have both.

- What is the probability that a car has a CD player, given that it has AC ?

i.e., we want to find $P(\ CD\ |\ AC\ )$

Conditional Probability Example

(continued)

- Of the cars on a used car lot, 70% have air conditioning (AC) and 40% have a CD player (CD).
  20% of the cars have both.

	CD	No CD	Total
AC	.2	.5	.7
No AC	.2	.1	.3
Total	.4	.6	1.0

$$P(CD \mid AC) = \frac{P(CD \cap AC)}{P(AC)} = \frac{.2}{.7} = .2857$$

Conditional Probability Example

(continued)

- Given AC, we only consider the top row (70% of the cars). Of these, 20% have a CD player. 20% of 70% is 28.57%.

	CD	No CD	Total
AC	.2	.5	.7
No AC	.2	.1	.3
Total	.4	.6	1.0

$$P(CD \mid AC) = \frac{P(CD \cap AC)}{P(AC)} = \frac{.2}{.7} = .2857$$

Multiplication Rule

- Multiplication rule for two events A and B:

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

- also

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

Multiplication Rule Example

P(Red ∩ Ace) = P(Red | Ace)P(Ace)

$$= \left(\frac{2}{4}\right)\left(\frac{4}{52}\right) = \frac{2}{52}$$

$$= \frac{\text{number of cards that are red and ace}}{\text{total number of cards}} = \frac{2}{52}$$

Type	Color		Total
	Red	Black	
Ace	2	2	4
Non-Ace	24	24	48
Total	26	26	52

Statistical Independence

- Two events are statistically independent if and only if:

$$\boxed{P(A \cap B) = P(A)P(B)}$$

  - Events A and B are independent when the probability of one event is not affected by the other event

- If A and B are independent, then

$$\boxed{P(A|B) = P(A)} \quad \text{if } P(B)>0$$

$$\boxed{P(B|A) = P(B)} \quad \text{if } P(A)>0$$

Statistical Independence

(continued)

- For multiple events:

$E_1, E_2, \ldots, E_k$ are statistically independent if and only if:

$$\boxed{P(E_1 \cap E_2 \cap \ldots \cap E_k) = P(E_1)\,P(E_2)\ldots P(E_k)}$$

Statistical Independence Example

- Of the cars on a used car lot, 70% have air conditioning (AC) and 40% have a CD player (CD).
  20% of the cars have both.

	CD	No CD	Total
AC	.2	.5	.7
No AC	.2	.1	.3
Total	.4	.6	1.0

- Are the events AC and CD statistically independent?

Statistical Independence Example

(continued)

	CD	No CD	Total
AC	.2	.5	.7
No AC	.2	.1	.3
Total	.4	.6	1.0

$P(AC \cap CD) = 0.2$

$P(AC) = 0.7$

$P(CD) = 0.4$

$\left.\begin{array}{l} P(AC) = 0.7 \\ P(CD) = 0.4 \end{array}\right\} P(AC)P(CD) = (0.7)(0.4) = 0.28$

$P(AC \cap CD) = 0.2 \neq P(AC)P(CD) = 0.28$

So the two events are not statistically independent

3.4

Bivariate Probabilities

Outcomes for bivariate events:

	B_1	B_2	. . .	B_k
A_1	$P(A_1 \cap B_1)$	$P(A_1 \cap B_2)$	. . .	$P(A_1 \cap B_k)$
A_2	$P(A_2 \cap B_1)$	$P(A_2 \cap B_2)$	. . .	$P(A_2 \cap B_k)$
. . .	. . .	. . .	. . .	. . .
A_h	$P(A_h \cap B_1)$	$P(A_h \cap B_2)$	. . .	$P(A_h \cap B_k)$

Example of a Bivariate Events with Bivariate Probabilities

collectively exhaustive: at least one of the events must occur

	♥	♣	♦	♠
Ace				
King				
. . .				
2				

Joint and Marginal Probabilities

- The probability of a joint event, $A \cap B$:

$$P(A \cap B) = \frac{\text{number of outcomes satisfying A and B}}{\text{total number of elementary outcomes}}$$

- Computing a marginal probability:

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + \cdots + P(A \cap B_k)$$

  - Where B_1, B_2, …, B_k are k mutually exclusive and collectively exhaustive events

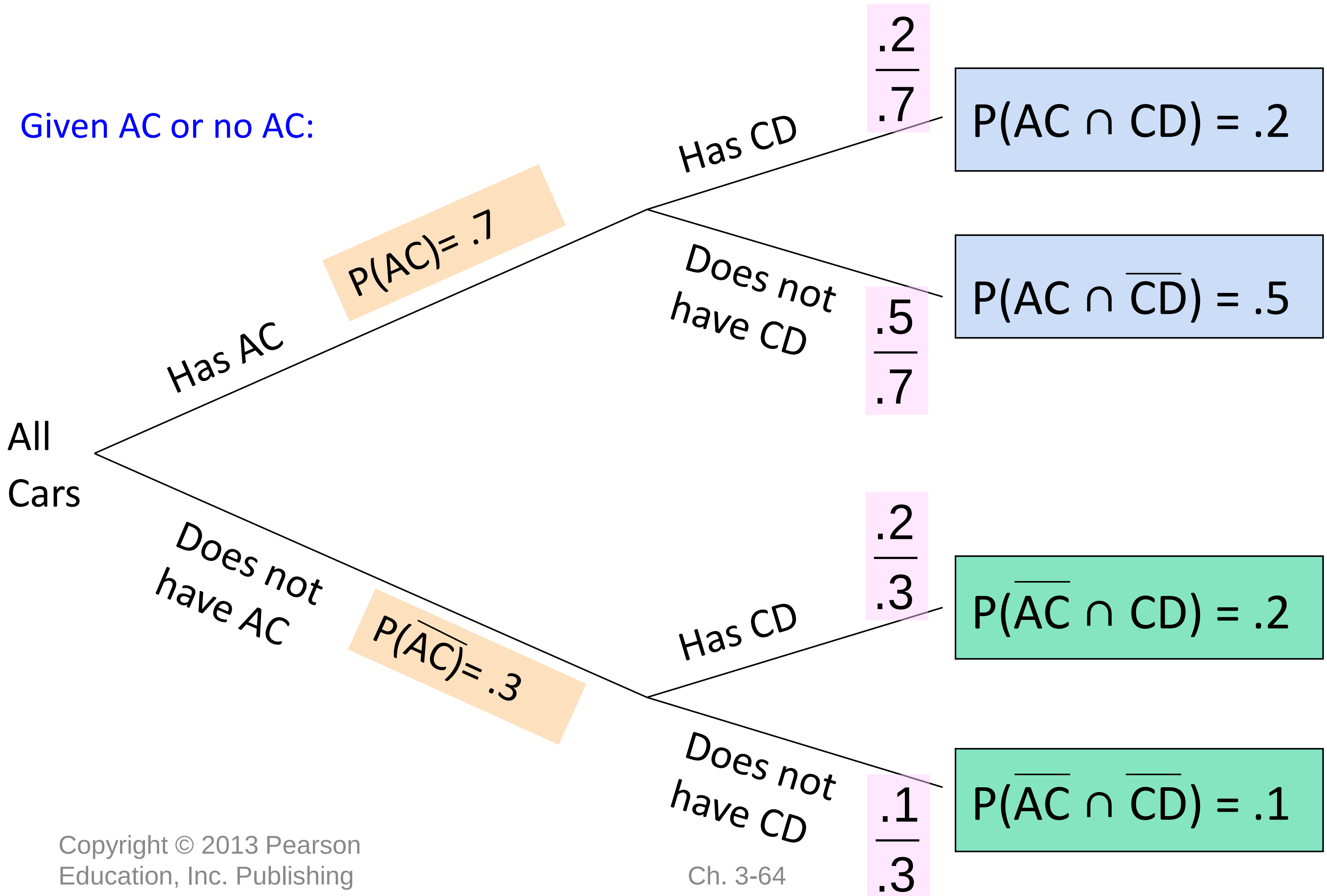
Marginal Probability Example

$$P(\text{Ace})$$

$$= P(\text{Ace} \cap \text{Red}) + P(\text{Ace} \cap \text{Black}) = \frac{2}{52} + \frac{2}{52} = \frac{4}{52}$$

Type	Color		Total
	Red	Black	
Ace	2	2	4
Non-Ace	24	24	48
Total	26	26	52

Using a Tree Diagram



Odds

- The odds in favor of a particular event are given by the ratio of the probability of the event divided by the probability of its complement
- The odds in favor of A are

$$\boxed{\text{odds} = \frac{P(A)}{1 - P(A)} = \frac{P(A)}{P(\overline{A})}}$$

Odds: Example

- Calculate the probability of winning if the odds of winning are 3 to 1:

$$\text{odds} = \frac{3}{1} = \frac{P(A)}{1 - P(A)}$$

  - Now multiply both sides by 1 – P(A) and solve for P(A):

$3 \times (1 - P(A)) = P(A)$

$3 - 3P(A) = P(A)$

$3 = 4P(A)$

$\boxed{P(A) = 0.75}$

3.5 Bayes' Theorem

Let A_1 and B_1 be two events. Bayes' theorem states that

$$P(B_1 | A_1) = \frac{P(A_1 | B_1)P(B_1)}{P(A_1)}$$

and

$$P(A_1 | B_1) = \frac{P(B_1 | A_1)P(A_1)}{P(B_1)}$$

- a way of revising conditional probabilities by using available or additional information

Bayes' Theorem

Bayes' theorem (alternative statement)

$$P(E_i \mid A) = \frac{P(A \mid E_i)P(E_i)}{P(A)}$$

$$= \frac{P(A \mid E_i)P(E_i)}{P(A \mid E_1)P(E_1) + P(A \mid E_2)P(E_2) + \ldots + P(A \mid E_k)P(E_k)}$$

- where:

  E_i = i^{th} event of k mutually exclusive and collectively exhaustive events

  A = new event that might impact $P(E_i)$

Bayes' Theorem Example

- A drilling company has estimated a 40% chance of striking oil for their new well.
- A detailed test has been scheduled for more information. Historically, 60% of successful wells have had detailed tests, and 20% of unsuccessful wells have had detailed tests.
- Given that this well has been scheduled for a detailed test, what is the probability that the well will be successful?

Bayes' Theorem Example

(continued)

- Let S = successful well

  U = unsuccessful well

- P(S) = .4 , P(U) = .6 (prior probabilities)
- Define the detailed test event as D
- Conditional probabilities:

  P(D|S) = .6 P(D|U) = .2

- Goal is to find P(S|D)

Bayes' Theorem Example

(continued)

Apply Bayes' Theorem:

$$P(S \mid D) = \frac{P(D \mid S)P(S)}{P(D \mid S)P(S) + P(D \mid U)P(U)}$$

$$= \frac{(.6)(.4)}{(.6)(.4) + (.2)(.6)}$$

$$= \frac{.24}{.24 + .12} = .667$$

So the revised probability of success (from the original estimate of .4), given that this well has been scheduled for a detailed test, is .667

Probability Distributions

Probability Distributions
Discrete Random Variables

Statistics for

Business and Economics

8^{th} Edition

Chapter 4

Discrete Random Variables and Probability Distributions

4.1

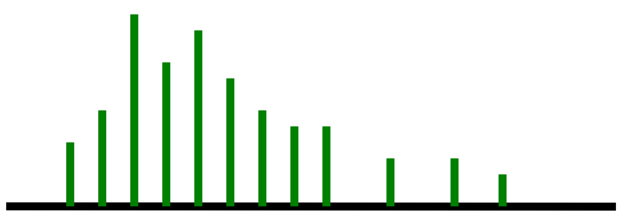
Random Variables

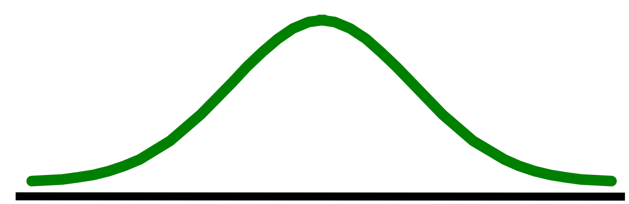
- **Random Variable**
  - Represents a possible numerical value from a random experiment

Random Variables

Ch. 4 **Discrete Random Variable**

Continuous Random Variable Ch. 5

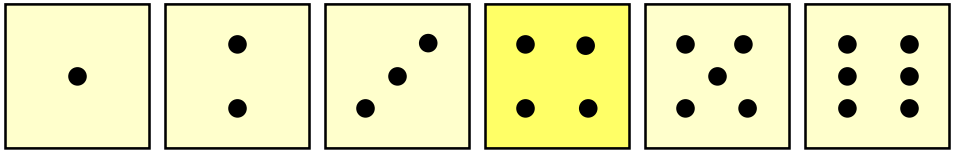




Discrete Random Variable

- Takes on no more than a countable number of values

Examples:



– **Roll a die twice**

Let X be the number of times 4 comes up

(then X could be 0, 1, or 2 times)

– **Toss a coin 5 times.**

Let X be the number of heads

(then X = 0, 1, 2, 3, 4, or 5)

Continuous Random Variable

- Can take on any value in an interval
  - Possible values are measured on a continuum

Examples:

- **Weight of packages filled by a mechanical filling process**
- **Temperature of a cleaning solution**
- **Time between failures of an electrical component**

Probability Distributions for Discrete Random Variables

Let X be a discrete random variable and x be one of its possible values

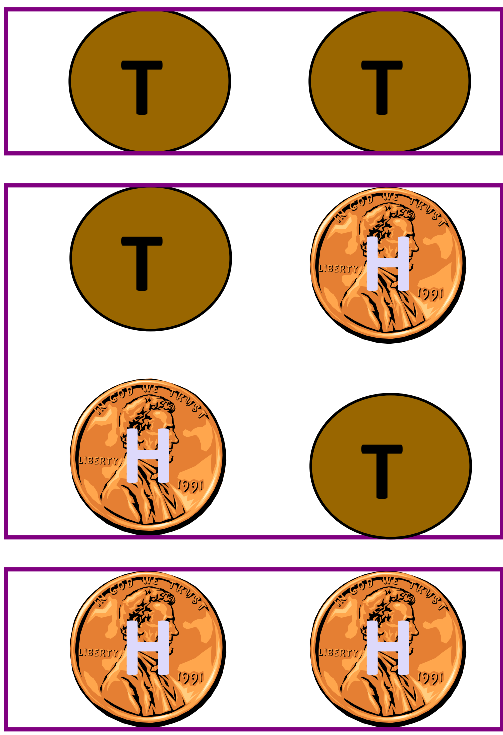
- The probability that random variable X takes specific value x is denoted $P(X = x)$
- The probability distribution function of a random variable is a representation of the probabilities for all the possible outcomes.
    - Can be shown algebraically, graphically, or with a table

Probability Distributions for Discrete Random Variables

Experiment: Toss 2 Coins. Let X = # heads.

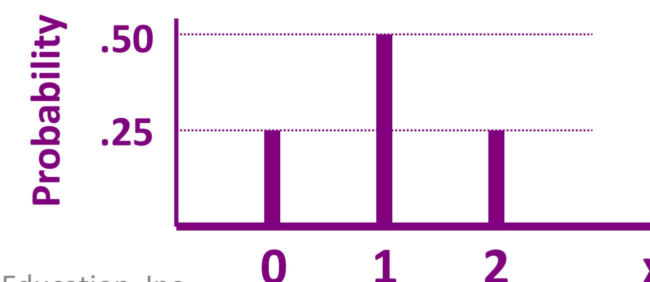
Show P(x) , i.e., P(X = x) , for all values of x:

4 possible outcomes



Probability Distribution

x Value	Probability
0	1/4 = .25
1	2/4 = .50
2	1/4 = .25



Probability Distribution Required Properties

- $\boxed{0 \le P(x) \le 1}$ for any value of x

- The individual probabilities sum to 1;

$$\boxed{\sum_{x} P(x) = 1}$$

(The notation indicates summation over all possible x values)

Cumulative Probability Function

- The cumulative probability function, denoted $F(x_0)$, shows the probability that X does not exceed the value x_0

$$\boxed{F(x_0) = P(X \leq x_0)}$$

Where the function is evaluated at all values of x_0

Cumulative Probability Function

(continued)

Example: Toss 2 Coins. Let X = # heads.

x Value	P(x)	F(x)
0	0.25	0.25
1	0.50	0.75
2	0.25	1.00

Properties of Discrete Random Variables

- Expected Value (or mean) of a discrete random variable X:

$$E[X] = \mu = \sum_{x} xP(x)$$

– Example: Toss 2 coins,

x = # of heads,

compute expected value of x:

x	P(x)
0	.25
1	.50
2	.25

E(x) = (0 x .25) + (1 x .50) + (2 x .25) = 1.0

Variance and Standard Deviation

- Variance of a discrete random variable X

$$\sigma^2 = E[(X-\mu)^2] = \sum_x (x-\mu)^2 P(x)$$

Can also be expressed as

$$\sigma^2 = E[X^2] - \mu^2 = \sum_x x^2 P(x) - \mu^2$$

- Standard Deviation of a discrete random variable X

$$\sigma = \sqrt{\sigma^2} = \sqrt{\sum_x (x-\mu)^2 P(x)}$$

Standard Deviation Example

– Example: Toss 2 coins, X = # heads, compute standard deviation (recall E[X] = 1)

$$\sigma = \sqrt{\sum_x (x-\mu)^2 P(x)}$$

$$\sigma = \sqrt{(0-1)^2(.25)+(1-1)^2(.50)+(2-1)^2(.25)} = \sqrt{.50} = .707$$

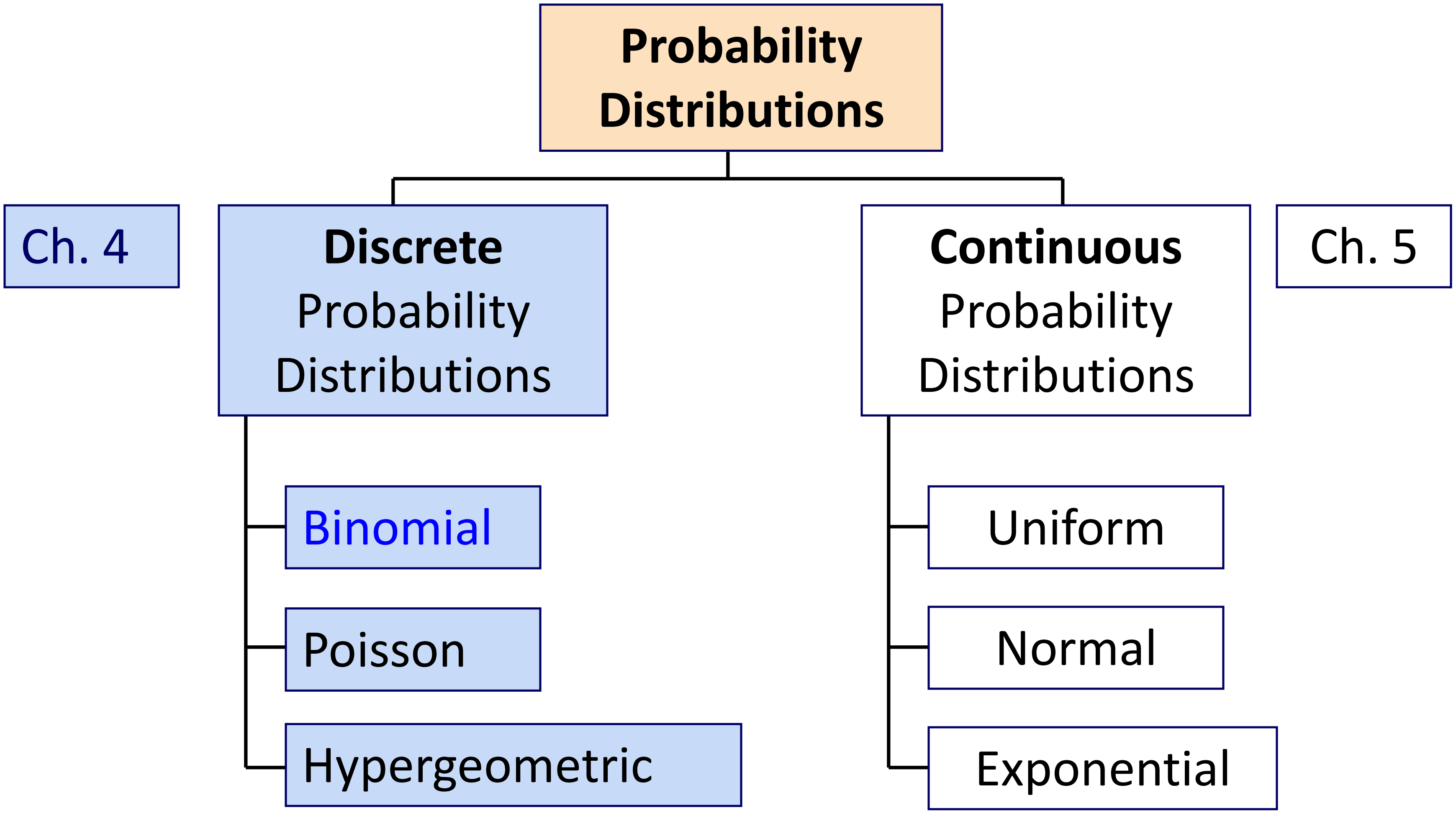
Possible number of heads = 0, 1, or 2

Functions of Random Variables

- If $P(x)$ is the probability function of a discrete random variable X, and $g(X)$ is some function of X, then the expected value of function g is

$$\boxed{E[g(X)] = \sum_{x} g(x)P(x)}$$

Probability Distributions



4.4

The Binomial Distribution

- **Probability Distributions**
  - **Discrete** Probability Distributions
    - Binomial
    - Hypergeometric
    - Poisson

Bernoulli Distribution

- Consider only two outcomes: "success" or "failure"
- Let P denote the probability of success
- Let $1 - P$ be the probability of failure
- Define random variable X:

$$x = 1 \text{ if success, } x = 0 \text{ if failure}$$

- Then the Bernoulli probability distribution is

$$\boxed{P(0) = (1 - P) \quad \text{and} \quad P(1) = P}$$

Mean and Variance of a Bernoulli Random Variable

- The mean is $\mu_x = P$

$$\mu_x = E[X] = \sum_X xP(x) = (0)(1-P) + (1)P = P$$

- The variance is $\sigma^2_x = P(1 - P)$

$$\sigma_x^2 = E[(X-\mu_x)^2] = \sum_X (x-\mu_x)^2 P(x)$$

$$= (0-P)^2(1-P) + (1-P)^2 P = P(1-P)$$

Developing the Binomial Distribution

- The number of sequences with x successes in n independent trials is:

$$\boxed{C_x^n = \frac{n!}{x!(n-x)!}}$$

Where n! = n·(n – 1)·(n – 2)· . . . ·1 and 0! = 1

- These sequences are mutually exclusive, since no two can occur at the same time

Binomial Probability Distribution

- A fixed number of observations, n
  - e.g., 15 tosses of a coin; ten light bulbs taken from a warehouse
- Two mutually exclusive and collectively exhaustive categories
  - e.g., head or tail in each toss of a coin; defective or not defective light bulb
  - Generally called “success” and “failure”
  - Probability of success is P , probability of failure is 1 – P
- Constant probability for each observation
  - e.g., Probability of getting a tail is the same each time we toss the coin
- Observations are independent
  - The outcome of one observation does not affect the outcome of the other

Possible Binomial Distribution Settings

- A manufacturing plant labels items as either defective or acceptable
- A firm bidding for contracts will either get a contract or not
- A marketing research firm receives survey responses of “yes I will buy” or “no I will not”
- New job applicants either accept the offer or reject it

The Binomial Distribution

$$P(x) = \frac{n\,!}{x\,!\,(n - x)!} P^{x} (1 - P)^{n - x}$$

P(x) = probability of **x** successes in **n** trials, with probability of success **P** on each trial

x = number of 'successes' in sample, (x = 0, 1, 2, ..., n)

n = sample size (number of independent trials or observations)

P = probability of "success"

Example: Flip a coin four times, let x = # heads:

n = 4

P = 0.5

1 - P = (1 - 0.5) = 0.5

x = 0, 1, 2, 3, 4

Example:
Calculating a Binomial Probability

What is the probability of one success in five observations if the probability of success is 0.1?

x = 1, n = 5, and P = 0.1

$$P(x=1) = \frac{n!}{x!(n-x)!}P^X(1-P)^{n-X}$$

$$= \frac{5!}{1!(5-1)!}(0.1)^1(1-0.1)^{5-1}$$

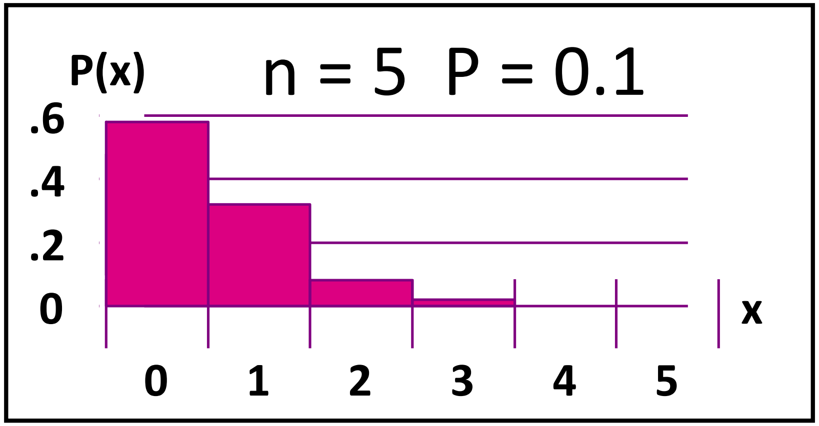
$$= (5)(0.1)(09)^4$$

$$= .32805$$

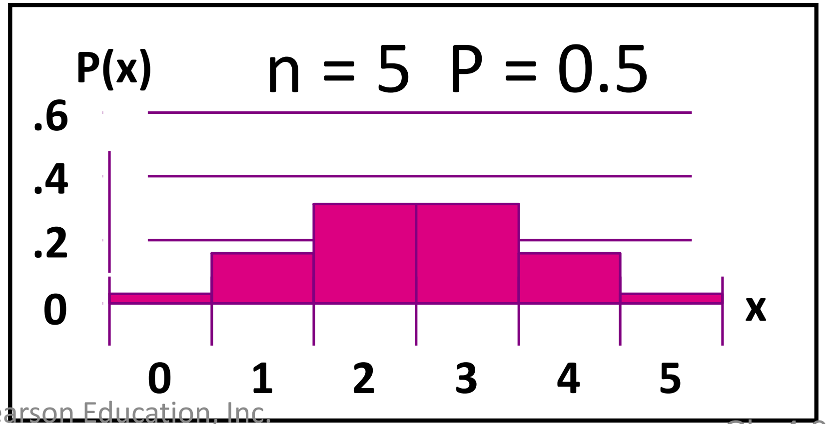
Shape of Binomial Distribution

- The shape of the binomial distribution depends on the values of P and n

  - Here, n = 5 and P = 0.1



  - Here, n = 5 and P = 0.5



Mean and Variance of a Binomial Distribution

- Mean

$$\mu = E(x) = nP$$

- Variance and Standard Deviation

$$\sigma^2 = nP(1- P)$$

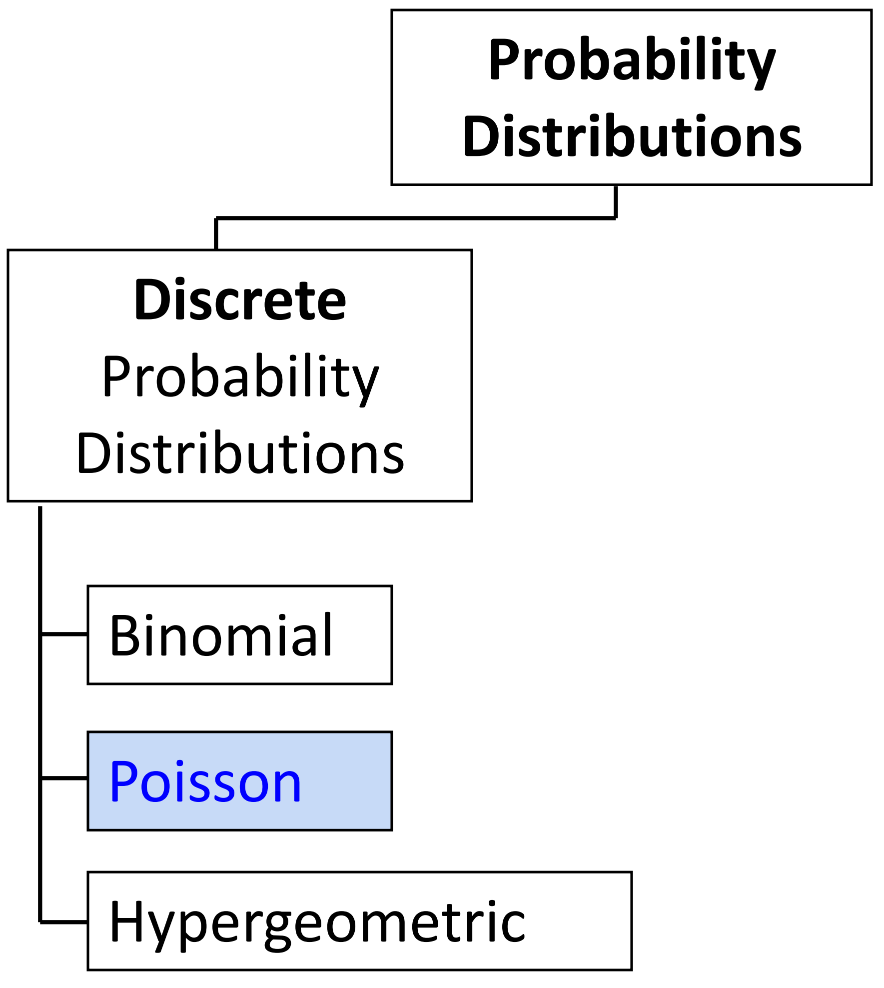
$$\sigma = \sqrt{nP(1-P)}$$

Where n = sample size

P = probability of success

(1 – P) = probability of failure

The Poisson Distribution



The Poisson Distribution

- The Poisson distribution is used to determine the probability of a random variable which characterizes the number of occurrences or successes of a certain event in a given continuous interval (such as time, surface area, or length).

The Poisson Distribution

(continued)

- Assume an interval is divided into a very large number of equal subintervals where the probability of the occurrence of an event in any subinterval is very small.

Poisson distribution assumptions

1. The probability of the occurrence of an event is constant for all subintervals.
2. There can be no more than one occurrence in each subinterval.
3. Occurrences are independent; that is, an occurrence in one interval does not influence the probability of an occurrence in another interval.

Poisson Distribution Function

The expected number of events per unit is the parameter λ (lambda), which is a constant that specifies the average number of occurrences (successes) for a particular time and/or space

$$\boxed{P(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}}$$

where:

P(x) = the probability of x successes over a given time or space, given λ

λ = the expected number of successes per time or space unit, $\lambda > 0$

e = base of the natural logarithm system (2.71828...)

Poisson Distribution Characteristics

Mean and variance of the Poisson distribution

- Mean

$$\boxed{\mu_x = E[X] = \lambda}$$

- Variance and Standard Deviation

$$\boxed{\sigma_x^2 = E[(X - \mu_x)^2] = \lambda}$$

$$\boxed{\sigma = \sqrt{\lambda}}$$

where λ = expected number of successes per time or space unit

Using Poisson Tables

X	λ								
	0.10	0.20	0.30	0.40	**0.50**	0.60	0.70	0.80	0.90
0	0.9048	0.8187	0.7408	0.6703	0.6065	0.5488	0.4966	0.4493	0.4066
1	0.0905	0.1637	0.2222	0.2681	0.3033	0.3293	0.3476	0.3595	0.3659
2	0.0045	0.0164	0.0333	0.0536	**0.0758**	0.0988	0.1217	0.1438	0.1647
3	0.0002	0.0011	0.0033	0.0072	0.0126	0.0198	0.0284	0.0383	0.0494
4	0.0000	0.0001	0.0003	0.0007	0.0016	0.0030	0.0050	0.0077	0.0111
5	0.0000	0.0000	0.0000	0.0001	0.0002	0.0004	0.0007	0.0012	0.0020
6	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002	0.0003
7	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Example: Find P(X = 2) if $\lambda = .50$

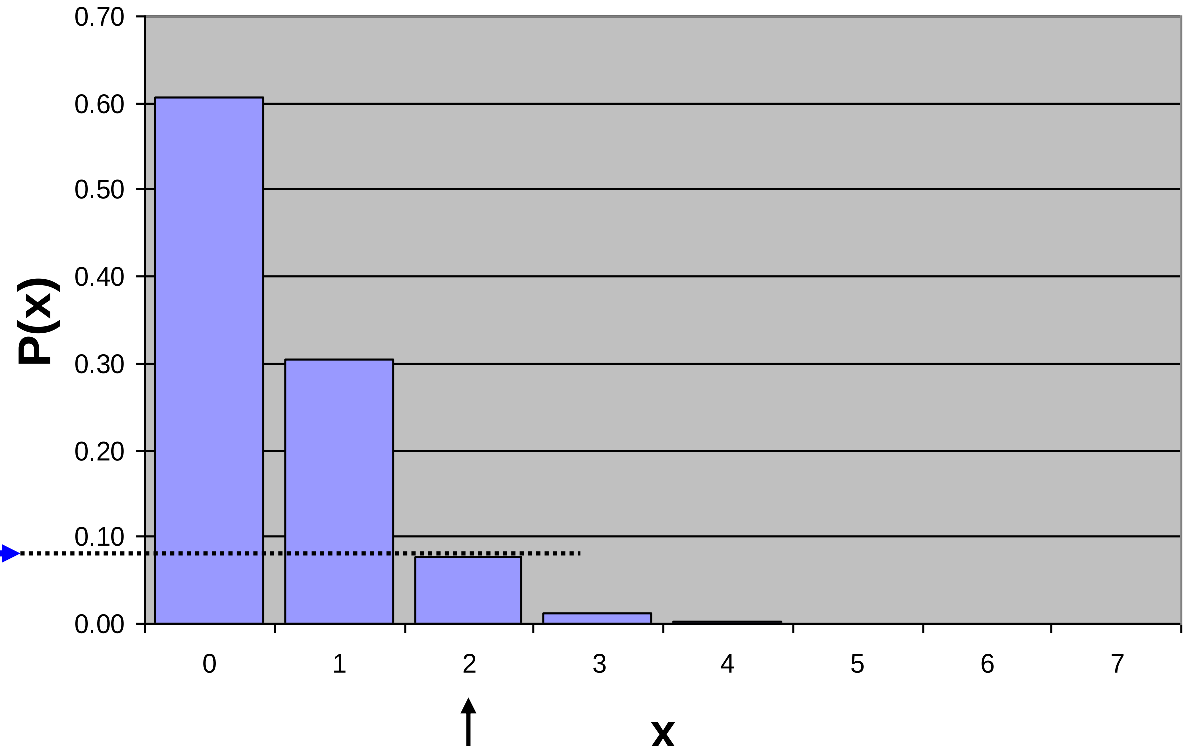
$$P(X=2) = \frac{e^{-\lambda}\lambda^X}{X!} = \frac{e^{-0.50}(0.50)^2}{2!} = .0758$$

Graph of Poisson Probabilities

Graphically:

λ = .50

X	**λ = 0.50**
0	0.6065
1	0.3033
2	0.0758
3	0.0126
4	0.0016
5	0.0002
6	0.0000
7	0.0000

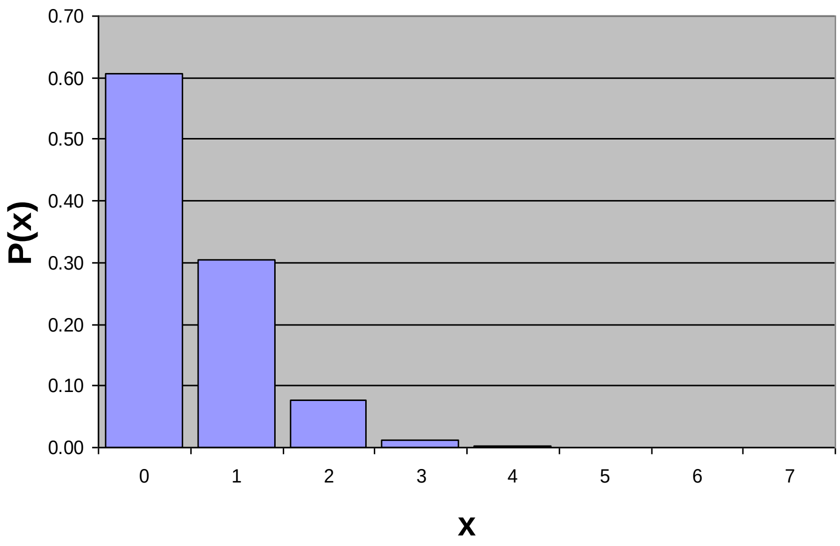


P(X = 2) = .0758

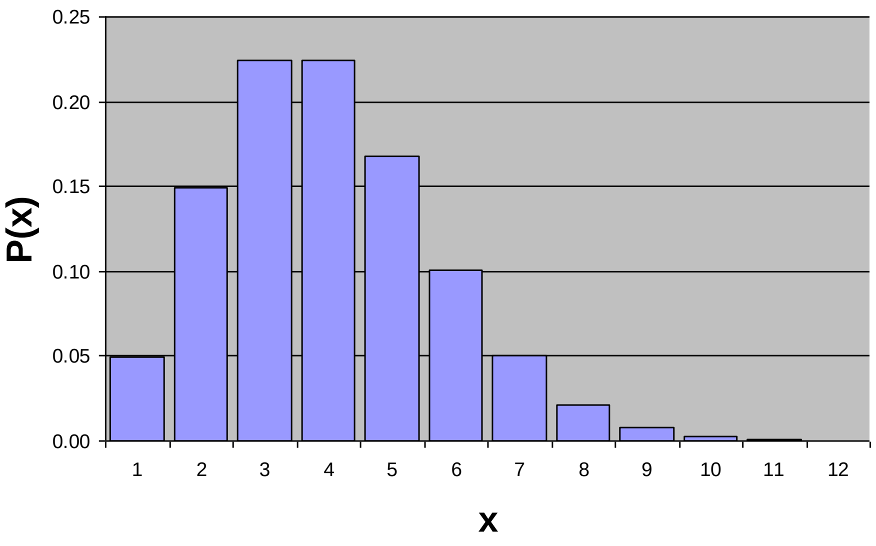
Poisson Distribution Shape

- The shape of the Poisson Distribution depends on the parameter λ :

$\lambda = 0.50$



$\lambda = 3.00$

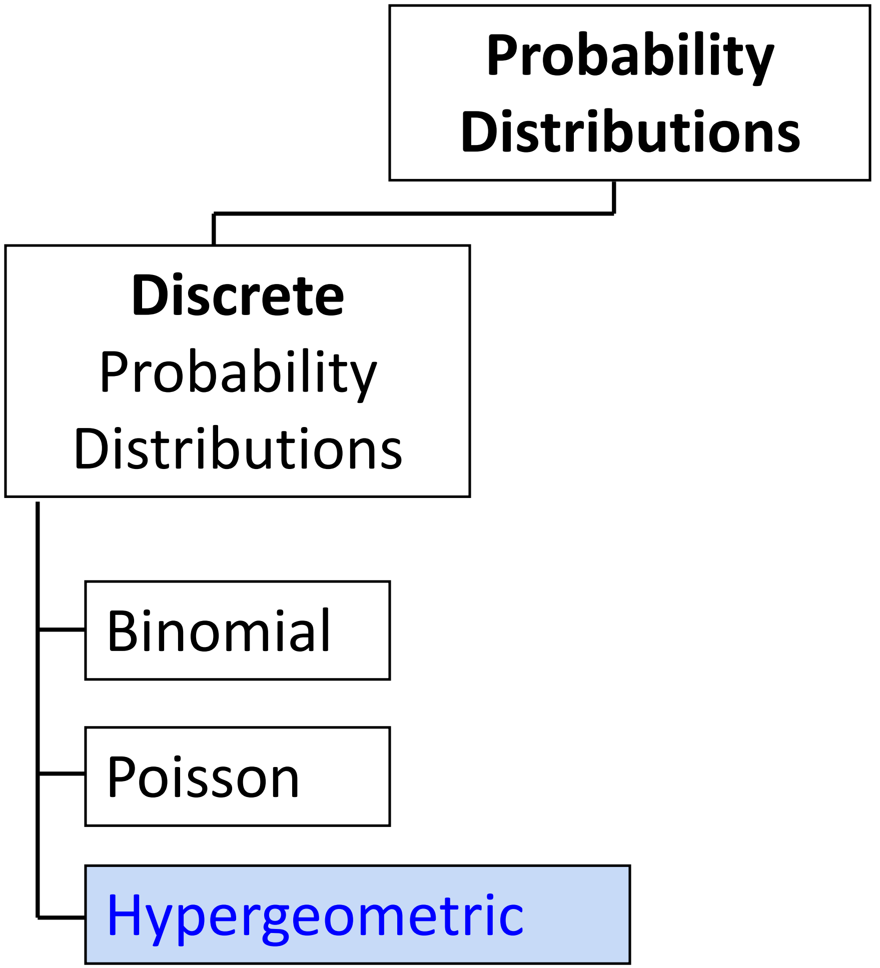


Poisson Approximation to the Binomial Distribution

Let X be the number of successes from n independent trials, each with probability of success P. The distribution of the number of successes, X , is binomial, with mean nP. If the number of trials, n , is large and nP is of only moderate size (preferably nP ≤ 7), this distribution can be approximated by the Poisson distribution with λ = nP. The probability distribution of the approximating distribution is

$$P(x) = \frac{e^{-nP}(nP)^x}{x!} \quad \text{for} \quad x = 0,1,2,\ldots$$

The Hypergeometric Distribution



The Hypergeometric Distribution

- "n" trials in a sample taken from a finite population of size N
- Sample taken without replacement
- Outcomes of trials are dependent
- Concerned with finding the probability of "X" successes in the sample where there are "S" successes in the population

Probability Distributions
Continuous Random Variables

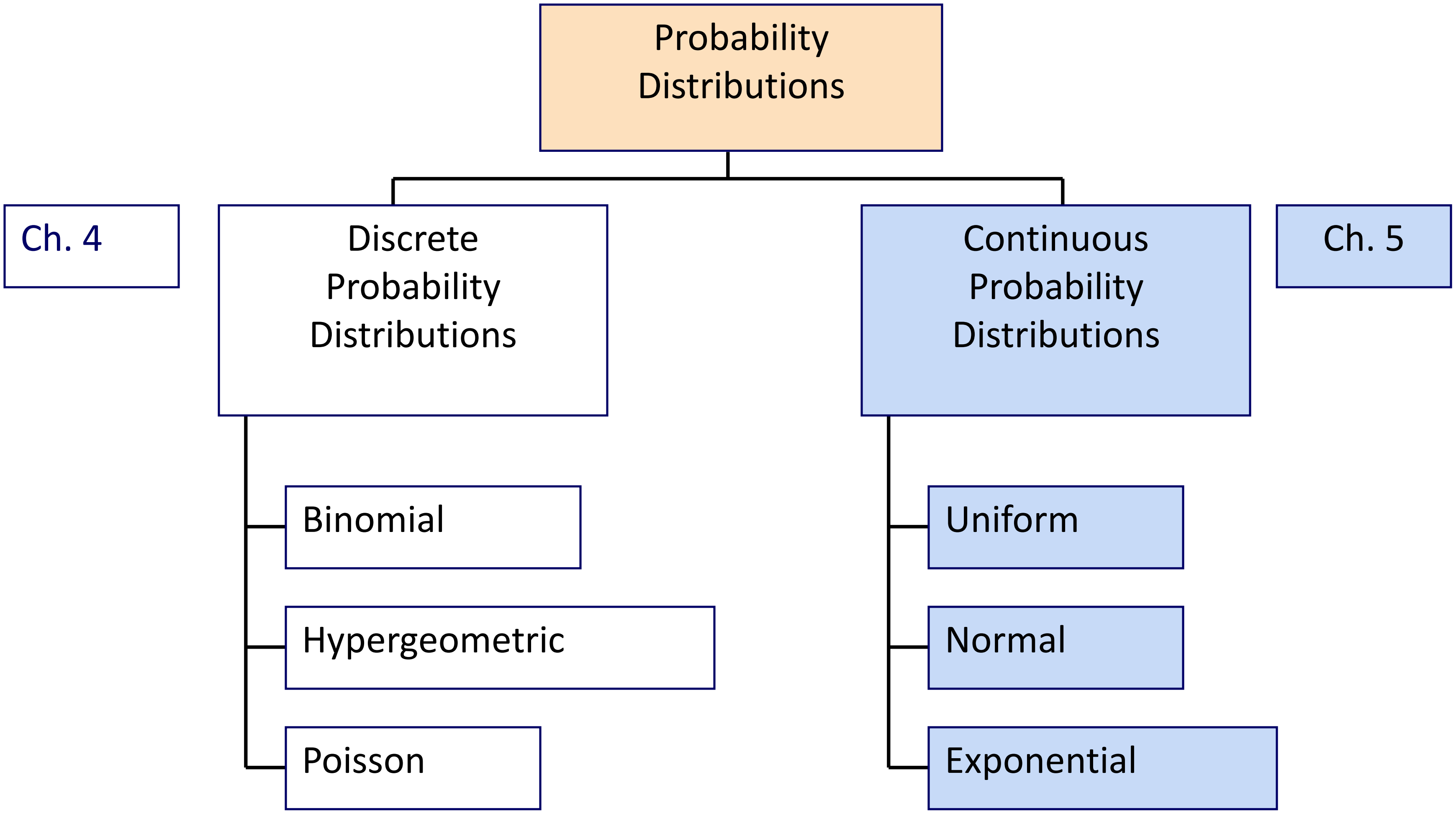
Statistics for Business and Economics

8th Edition

Chapter 5

Continuous Random Variables and Probability Distributions

Probability Distributions



5.1

Continuous Random Variables

- A continuous random variable is a variable that can assume any value in an interval
  - thickness of an item
  - time required to complete a task
  - temperature of a solution
  - height, in inches
- These can potentially take on any value, depending only on the ability to measure accurately.

Cumulative Distribution Function

- The cumulative distribution function, $F(x)$, for a continuous random variable X expresses the probability that X does not exceed the value of x

$$\boxed{F(x) = P(X \leq x)}$$

- Let a and b be two possible values of X, with $a < b$. The probability that X lies between a and b is

$$\boxed{P(a < X < b) = F(b) - F(a)}$$

Probability Density Function

The probability density function, f(x), of random variable X has the following properties:

1. $f(x) > 0$ for all values of x
2. The area under the probability density function f(x) over all values of the random variable X within its range, is equal to 1.0
3. The probability that X lies between two values is the area under the density function graph between the two values

$$\boxed{P(a < X < b) = \int_a^b f(x)dx}$$

Probability Density Function

(continued)

The probability density function, $f(x)$, of random variable X has the following properties:

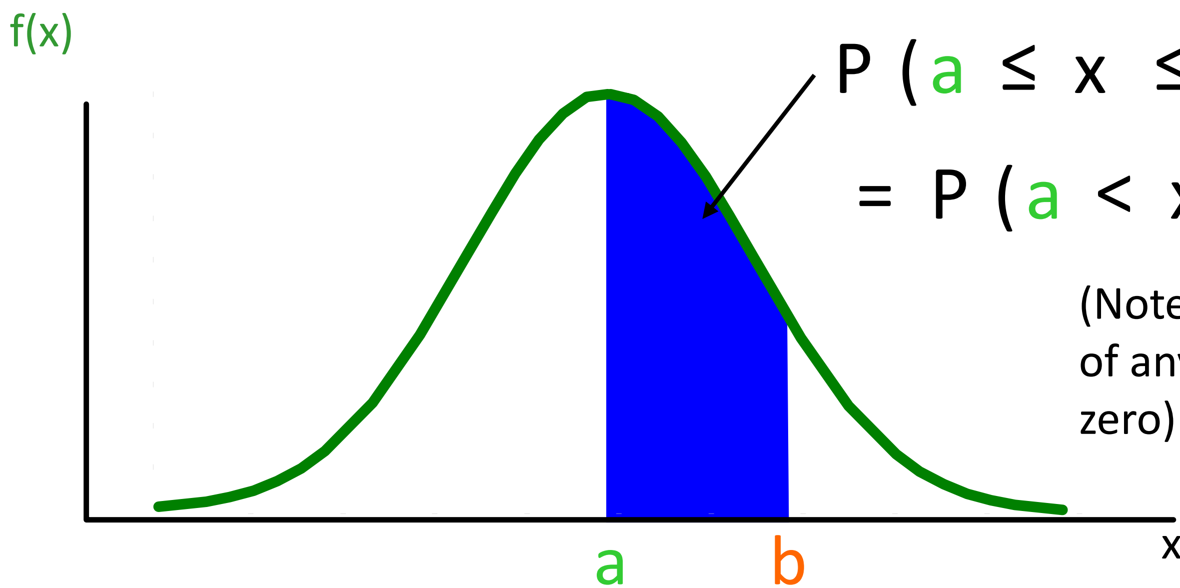
4. The cumulative density function $F(x_0)$ is the area under the probability density function $f(x)$ from the minimum x value up to x_0

$$F(x_0) = \int_{x_m}^{x_0} f(x)dx$$

where x_m is the minimum value of the random variable x

Probability as an Area

Shaded area under the curve is the probability that X is between a and b



$$P(a \le x \le b) = P(a < x < b)$$

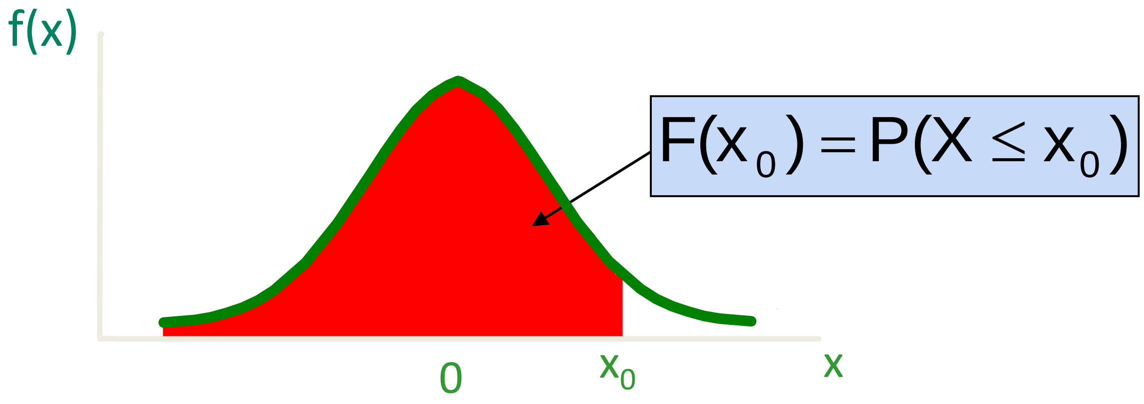
(Note that the probability of any individual value is zero)

Probability as an Area

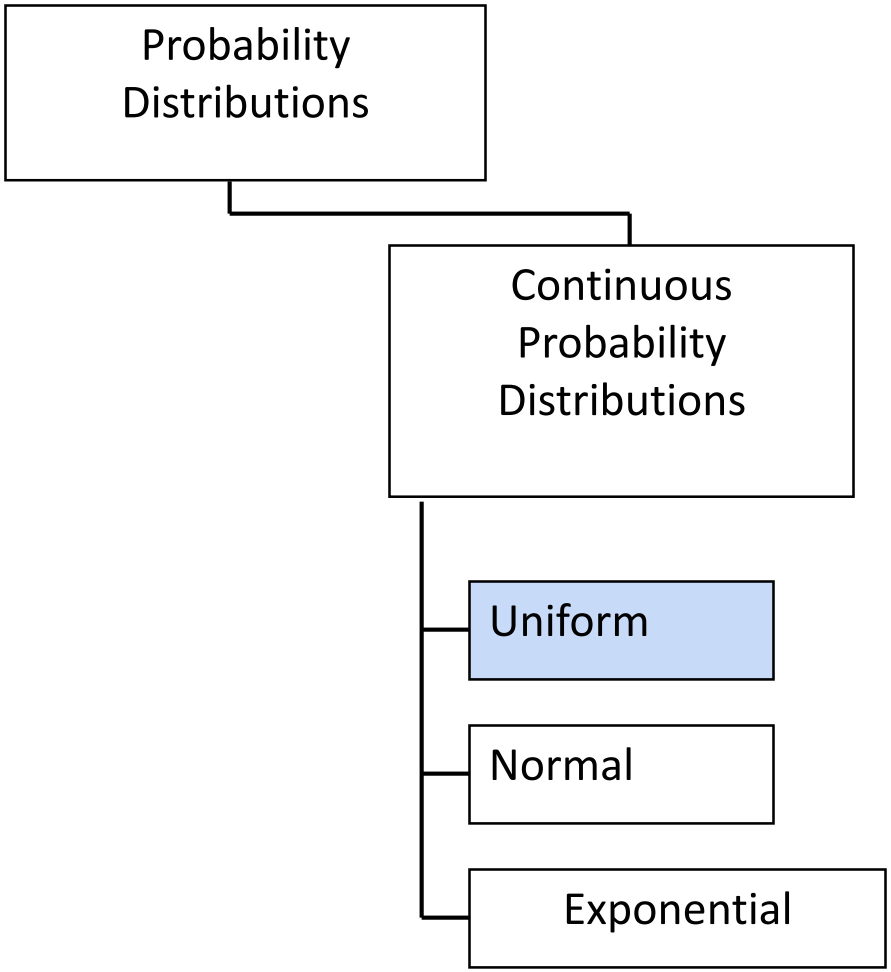
(continued)

1. The total area under the curve $f(x)$ is 1

2. The area under the curve $f(x)$ to the left of x_0 is $F(x_0)$, where x_0 is any value that the random variable can take.

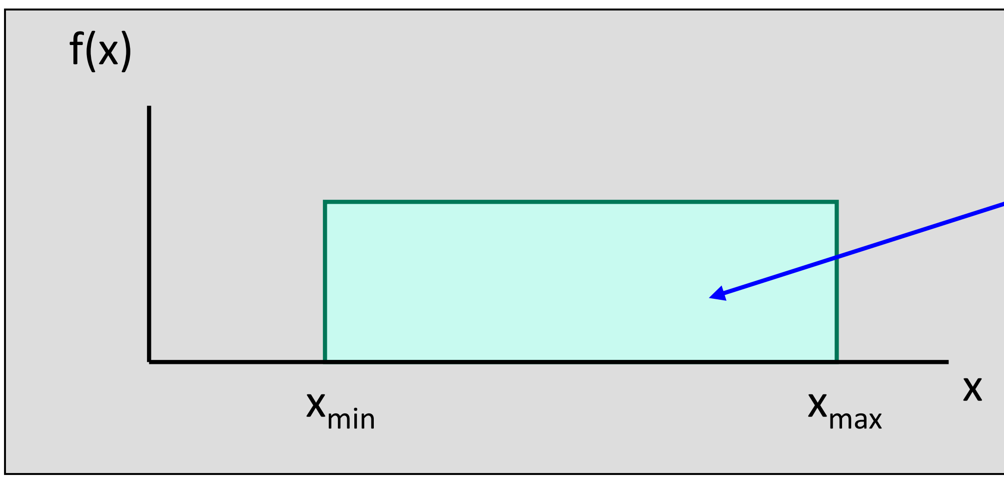


The Uniform Distribution



The Uniform Distribution

- The uniform distribution is a probability distribution that has equal probabilities for all equal-width intervals within the range of the random variable



Total area under the uniform probability density function is 1.0

The Uniform Distribution

(continued)

The Continuous Uniform Distribution:

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{if } a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

where

$f(x)$ = value of the density function at any x value

a = minimum value of x

b = maximum value of x

5.2

Expectations for Continuous Random Variables

- The mean of X, denoted μ_X, is defined as the expected value of X

$$\boxed{\mu_X = E[X]}$$

- The variance of X, denoted σ_X^2, is defined as the expectation of the squared deviation, $(X - \mu_X)^2$, of a random variable from its mean

$$\boxed{\sigma_X^2 = E[(X - \mu_X)^2]}$$

Mean and Variance of the Uniform Distribution

- The mean of a uniform distribution is

$$\boxed{\mu = \frac{a+b}{2}}$$

- The variance is

$$\boxed{\sigma^2 = \frac{(b-a)^2}{12}}$$

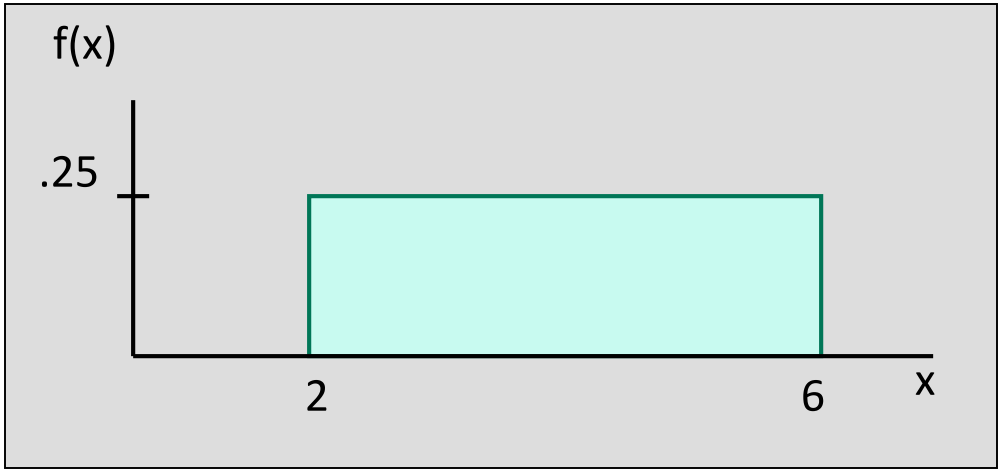
Where a = minimum value of x

b = maximum value of x

Uniform Distribution Example

Example: Uniform probability distribution over the range $2 \le x \le 6$:

$$f(x) = \frac{1}{6-2} = .25 \quad \text{for } 2 \le x \le 6$$



$$\mu = \frac{a+b}{2} = \frac{2+6}{2} = 4$$

$$\sigma^2 = \frac{(b-a)^2}{12} = \frac{(6-2)^2}{12} = 1.333$$

Linear Functions of Random Variables

- Let $\boxed{W = a + bX}$, where X has mean μ_X and variance σ_X^2, and a and b are constants

- Then the mean of W is

$$\boxed{\mu_W = E[a + bX] = a + b\mu_X}$$

- the variance is

$$\boxed{\sigma_W^2 = Var[a + bX] = b^2\sigma_X^2}$$

- the standard deviation of W is

$$\boxed{\sigma_W = |b|\sigma_X}$$

Linear Functions of Random Variables

(continued)

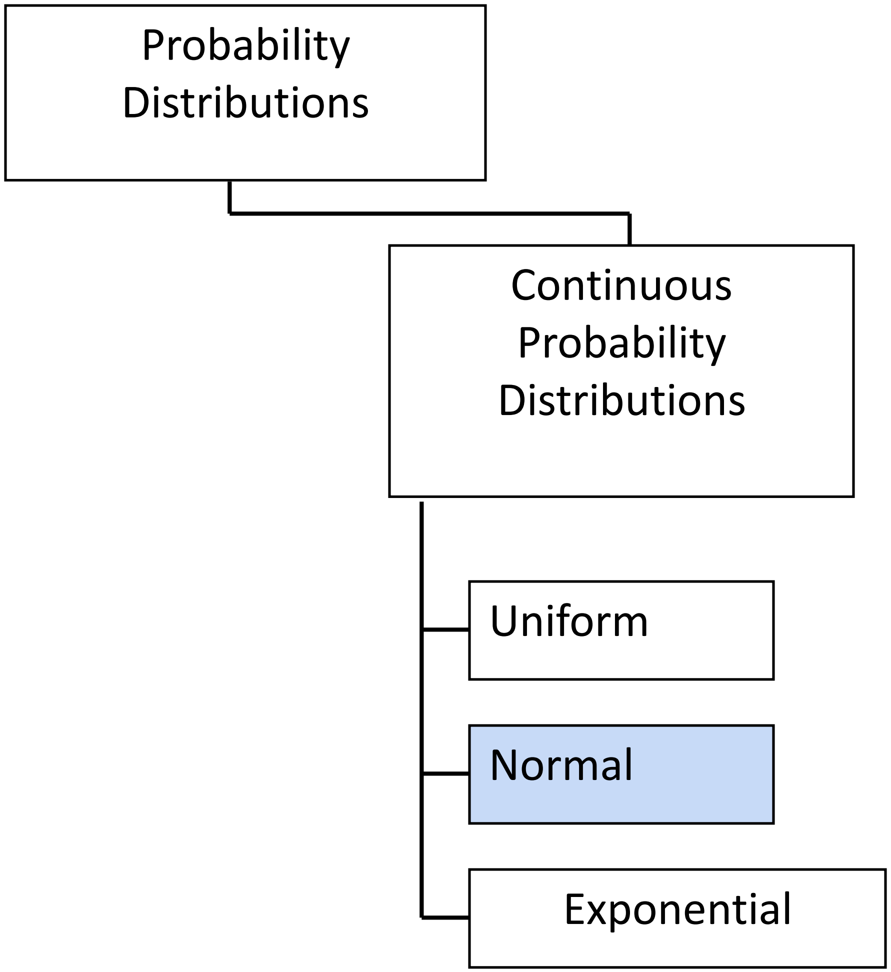
- An important special case of the result for the linear function of random variable is the standardized random variable

$$\boxed{Z = \frac{X - \mu_X}{\sigma_X}}$$

- which has a mean 0 and variance 1

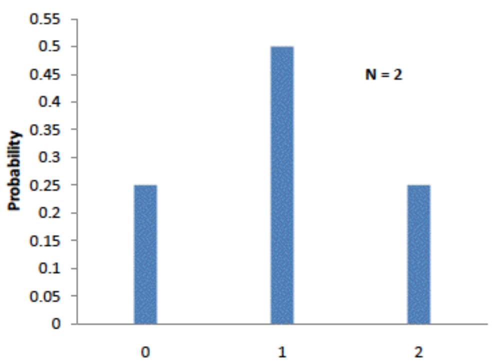
5.3

The Normal Distribution

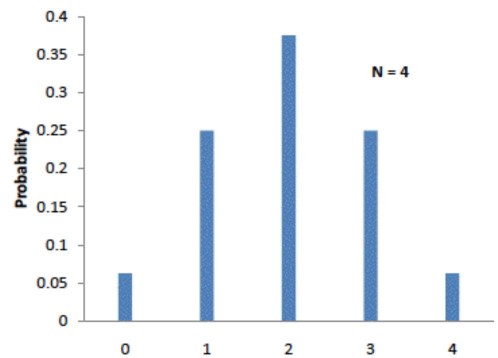


History of the Normal Distribution

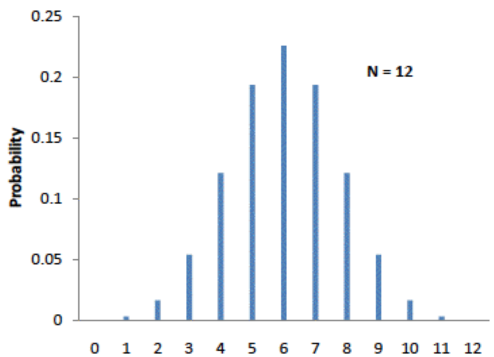
- Abraham de Moivre: an 18th century statistician and consultant to gamblers
  - when the number of events (coin flips) increased, the shape of the binomial distribution approached a very smooth curve.



2 flips



4 flips



12 flips

- Independently, the mathematicians Adrain in 1808 and Gauss in 1809 developed the formula for the normal distribution.

Lane, D. M. (2000). *Online statistics education: An interactive multimedia course of study*. Rice University.

The Normal Distribution

(continued)

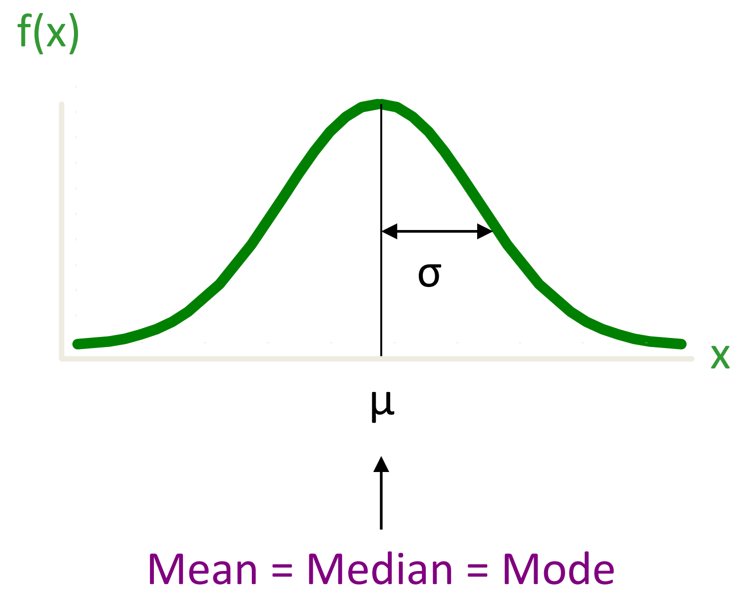
- Bell Shaped
- Symmetrical
- Mean, Median and Mode are Equal

Location is determined by the mean, μ

Spread is determined by the standard deviation, σ

The random variable has an infinite theoretical range:

$-\infty$ to $+\infty$

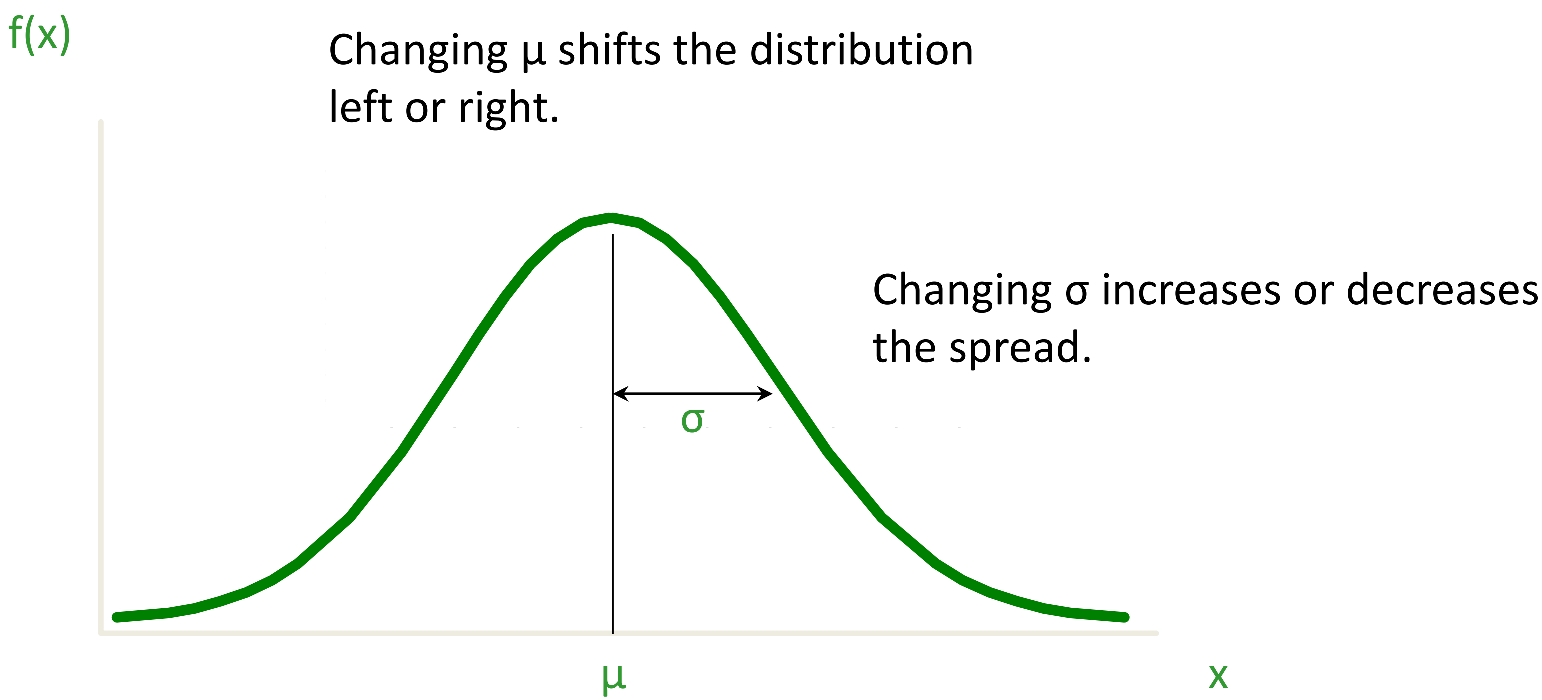


The Normal Distribution

(continued)

- The normal distribution closely approximates the probability distributions of a wide range of random variables
- Distributions of sample means approach a normal distribution given a "large" sample size
- Computations of probabilities are direct and elegant
- The normal probability distribution has led to good business decisions for a number of applications

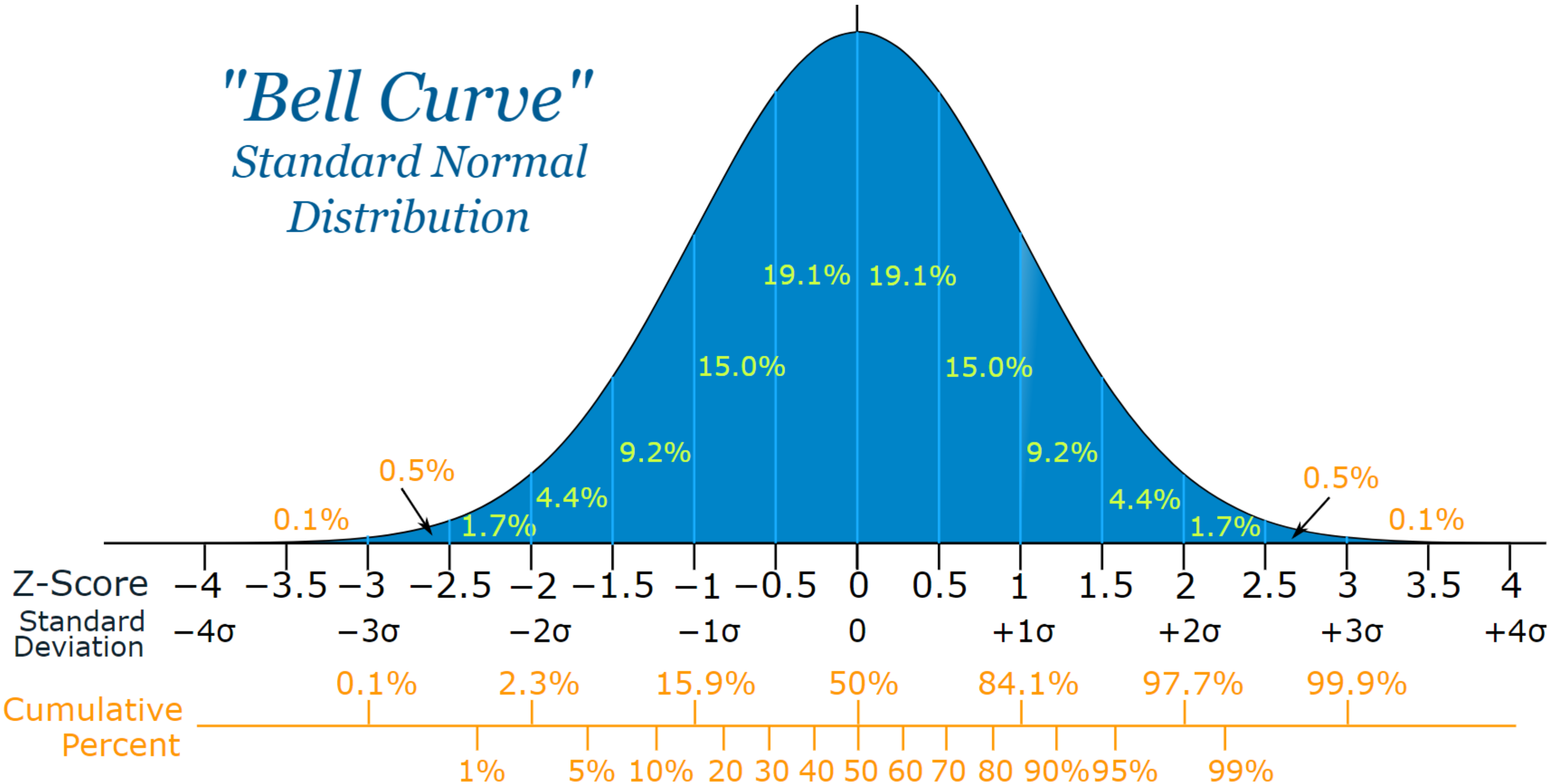
The Normal Distribution Shape



Given the mean μ and variance σ^2 we define the normal distribution using the notation

$$\boxed{X \sim N(\mu, \sigma^2)}$$

The Normal Distribution Shape



https://www.mathsisfun.com/data/standard-normal-distribution.html

The Normal Probability Density Function

- The formula for the normal probability density function is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$$

Where e = the mathematical constant approximated by 2.71828

π = the mathematical constant approximated by 3.14159

μ = the population mean

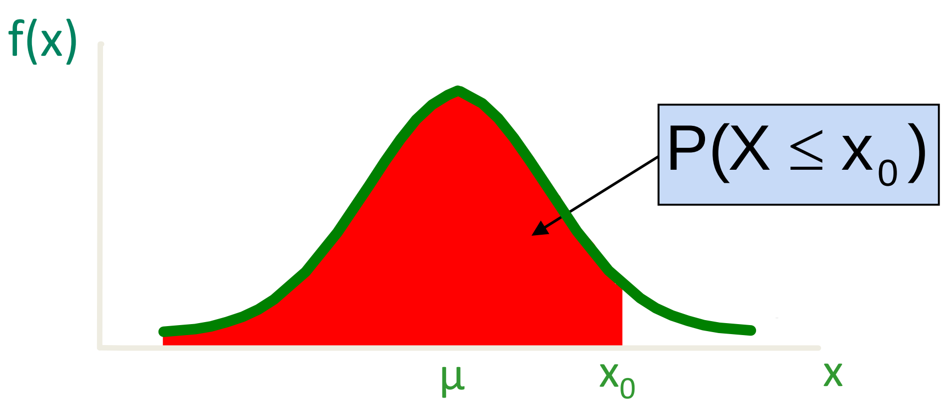
σ^2 = the population variance

x = any value of the continuous variable, $-\infty < x < \infty$

Cumulative Normal Distribution

- For a normal random variable X with mean μ and variance σ^2 , i.e., $X \sim N(\mu, \sigma^2)$, the cumulative distribution function is

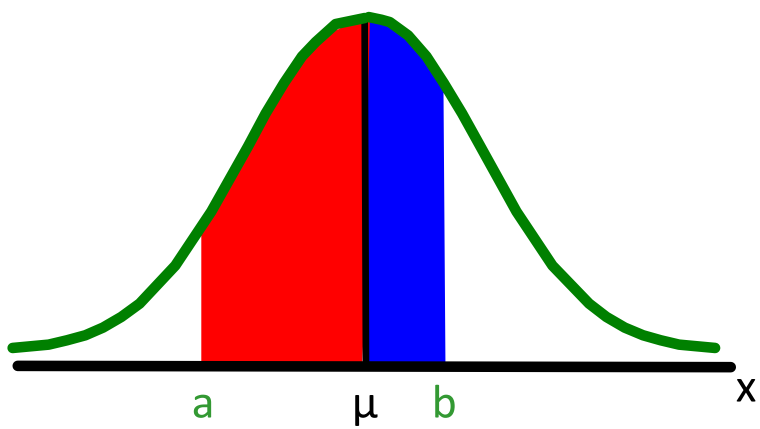
$$\boxed{F(x_0) = P(X \le x_0)}$$



Finding Normal Probabilities

The probability for a range of values is measured by the area under the curve

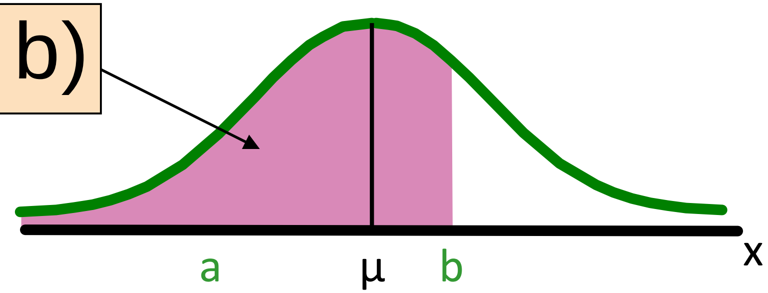
$$\boxed{P(a < X < b) = F(b) - F(a)}$$



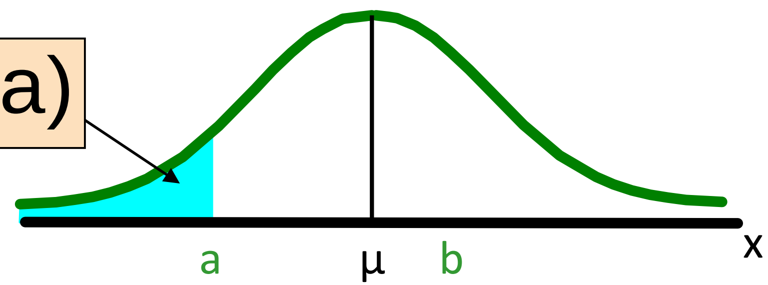
Finding Normal Probabilities

(continued)

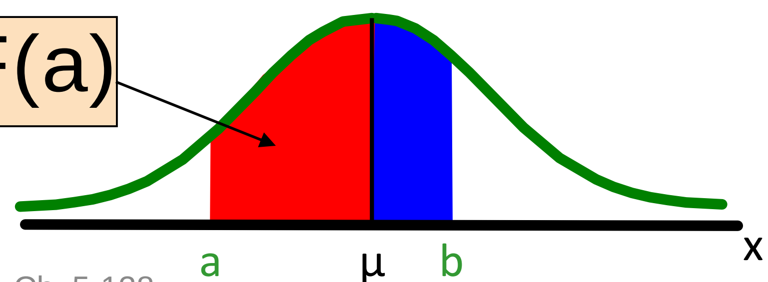
$$F(b) = P(X < b)$$



$$F(a) = P(X < a)$$



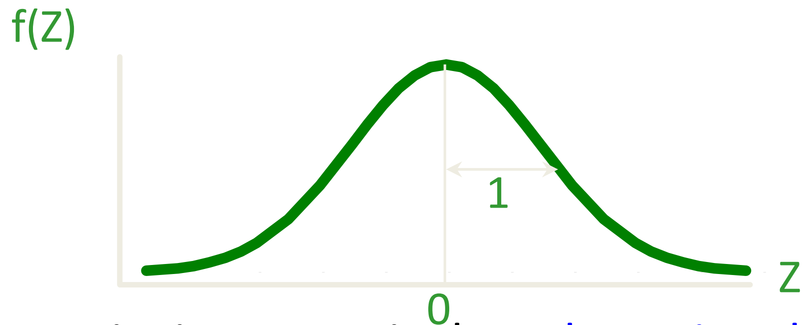
$$P(a < X < b) = F(b) - F(a)$$



The
Standard Normal Distribution

- Any normal distribution (with any mean and variance combination) can be transformed into the standardized normal distribution (Z), with mean 0 and variance 1

$$Z \sim N(0,1)$$



- Need to transform X units into Z units by subtracting the mean of X and dividing by its standard deviation

$$Z = \frac{X - \mu}{\sigma}$$

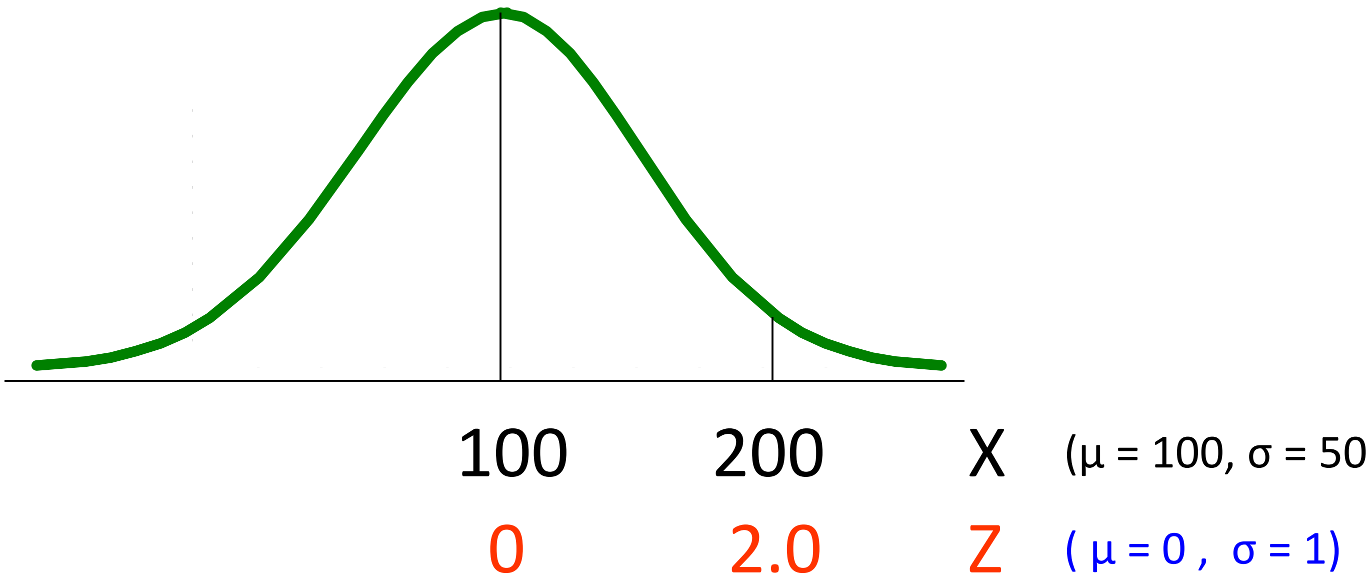
Example

- If X is distributed normally with mean of 100 and standard deviation of 50, the Z value for X = 200 is

$$Z = \frac{X - \mu}{\sigma} = \frac{200 - 100}{50} = 2.0$$

- This says that X = 200 is two standard deviations (2 increments of 50 units) above the mean of 100.

Comparing X and Z units

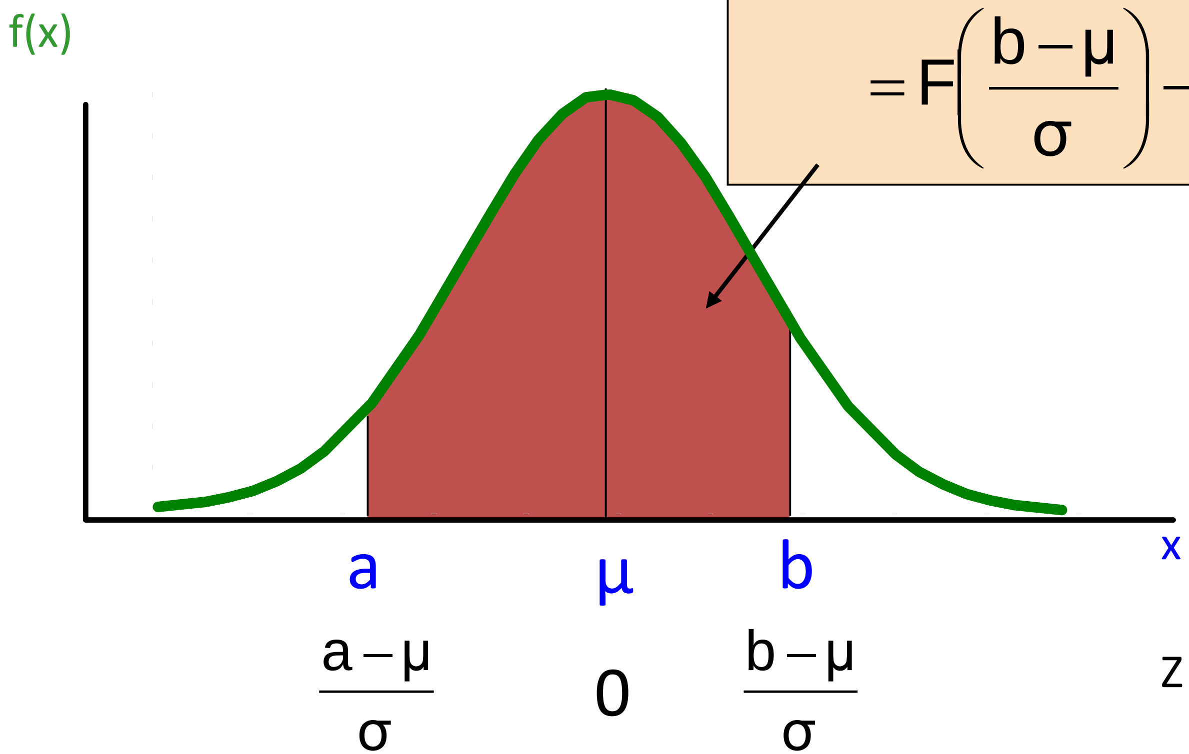


Note that the distribution is the same, only the scale has changed. We can express the problem in original units (X) or in standardized units (Z)

Finding Normal Probabilities

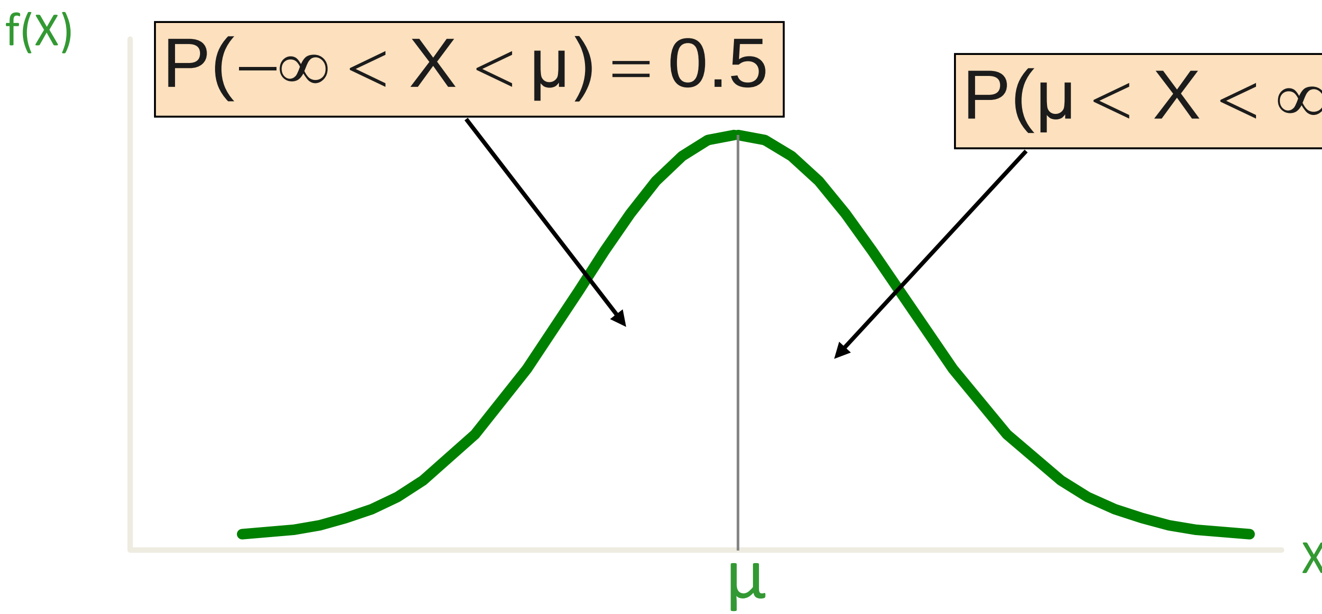
$$P(a < X < b) = P\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right)$$

$$= F\left(\frac{b-\mu}{\sigma}\right) - F\left(\frac{a-\mu}{\sigma}\right)$$



Probability as Area Under the Curve

The total area under the curve is 1.0, and the curve is symmetric, so half is above the mean, half is below

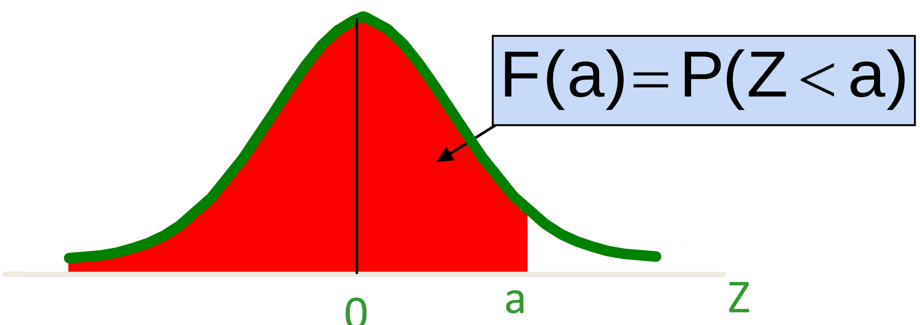


$$P(-\infty < X < \infty) = 1.0$$

Appendix Table 1

- The Standard Normal Distribution calculator (e.g. http://vassarstats.net/tabs.html) shows values of the cumulative normal distribution function
- For a given Z-value a , the table shows F(a)

  (the area under the curve from negative infinity to a)



General Procedure for Finding Probabilities

To find $P(a < X < b)$ when X is distributed normally:

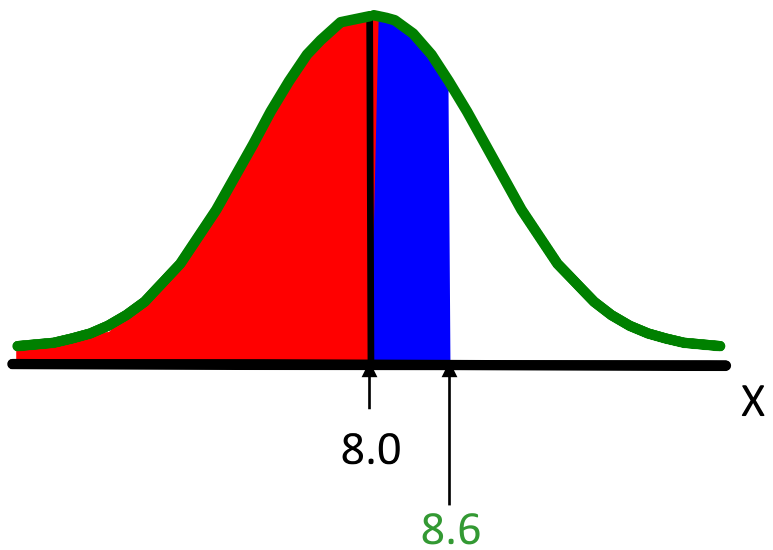
- Draw the normal curve for the problem in terms of X

- Translate X-values to Z-values

- Use the Cumulative Normal Table

Finding Normal Probabilities

- Suppose X is normal with mean 8.0 and standard deviation 5.0
- Find P(X < 8.6)

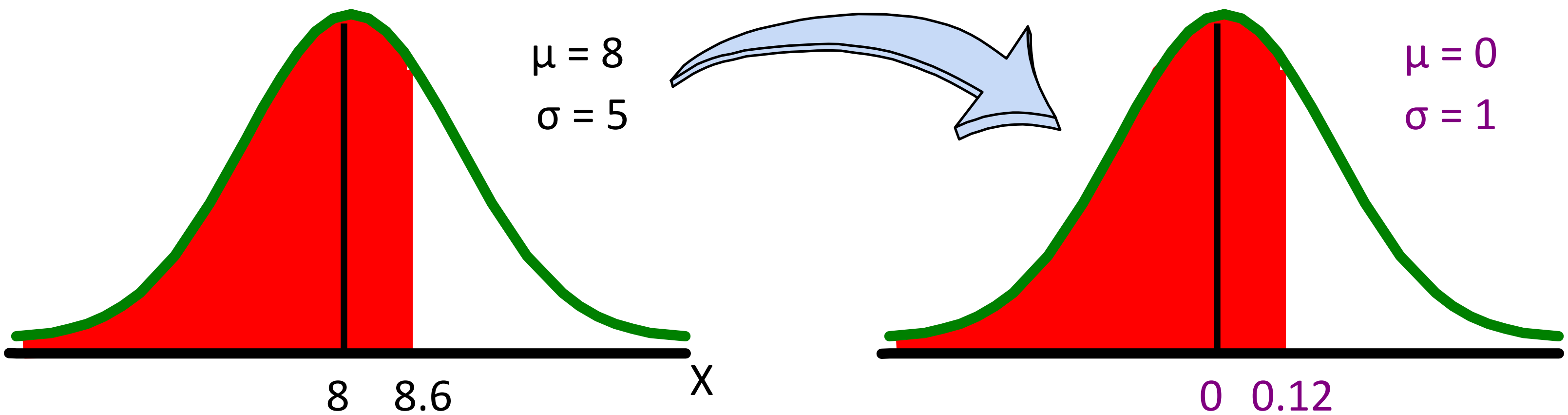


Finding Normal Probabilities

(continued)

- Suppose X is normal with mean 8.0 and standard deviation 5.0. Find P(X < 8.6)

$$Z = \frac{X - \mu}{\sigma} = \frac{8.6 - 8.0}{5.0} = 0.12$$

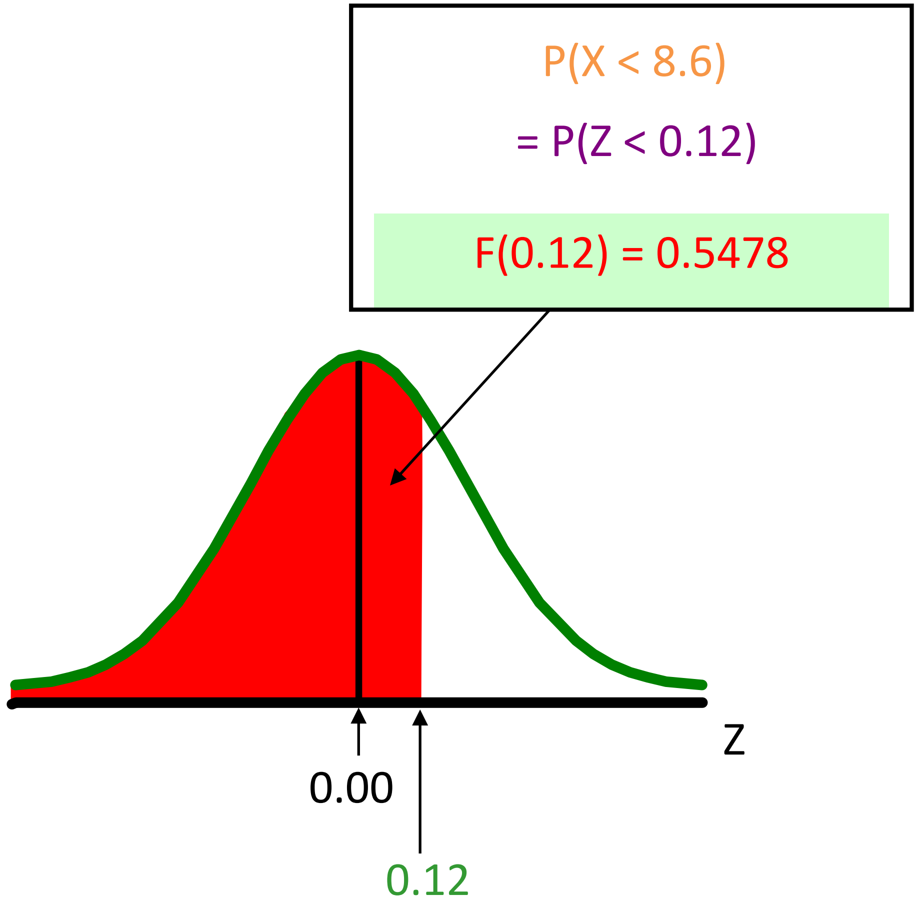


P(X < 8.6) P(Z < 0.12)

Solution: Finding P(Z < 0.12)

Standardized Normal Probability Table (Portion)

z	F(z)
.10	.5398
.11	.5438
.12	.5478
.13	.5517



Finding the X value for a Known Probability

- Steps to find the X value for a known probability:

1. Find the Z value for the known probability
2. Convert to X units using the formula:

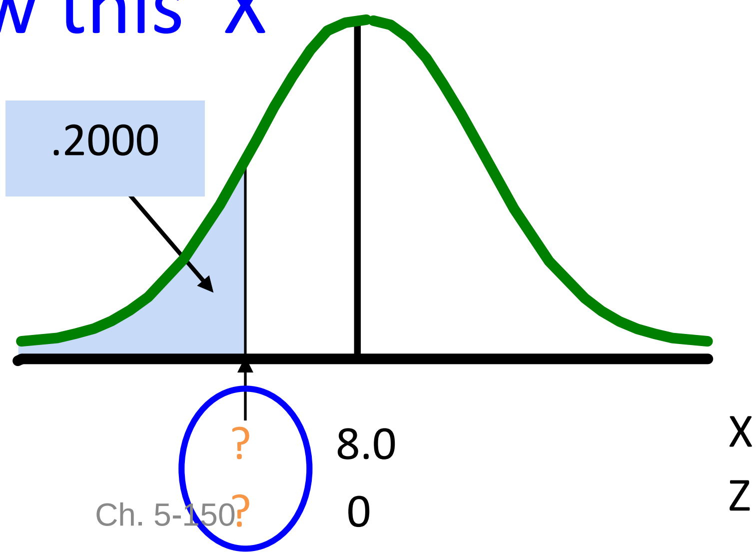
$$\boxed{X = \mu + Z\sigma}$$

Finding the X value for a Known Probability

(continued)

Example:

- Suppose X is normal with mean 8.0 and standard deviation 5.0.
- Now find the X value so that only 20% of all values are below this X



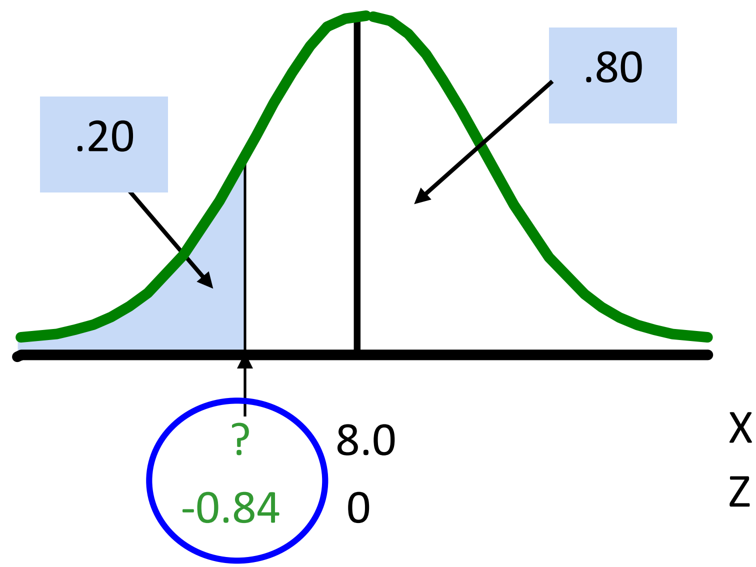
Find the Z value for 20% in the Lower Tail

1. Find the Z value for the known probability

Standardized Normal Probability Table (Portion)

z	F(z)
.82	.7939
.83	.7967
.84	.7995
.85	.8023

- 20% area in the lower tail is consistent with a Z value of -0.84



Finding the X value

2. Convert to X units using the formula:

$$
\begin{aligned}
X &= \mu + Z\sigma \\
&= 8.0 + (-0.84)5.0 \\
&= 3.80
\end{aligned}
$$

So 20% of the values from a distribution with mean 8.0 and standard deviation 5.0 are less than 3.80

Assessing Normality

- Not all continuous random variables are normally distributed
- It is important to evaluate how well the data is approximated by a normal distribution

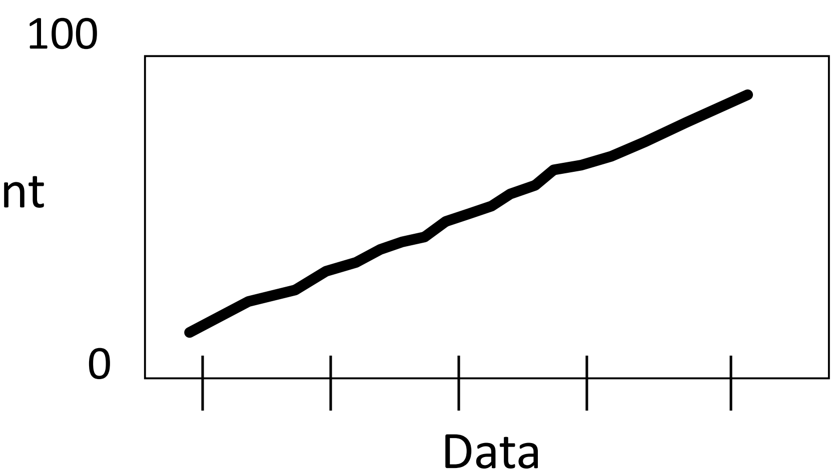
The Normal Probability Plot

- Normal probability plot
  - Arrange data from low to high values
  - Find cumulative normal probabilities for all values
  - Examine a plot of the observed values vs. cumulative probabilities (with the cumulative normal probability on the vertical axis and the observed data values on the horizontal axis)
  - Evaluate the plot for evidence of linearity

The Normal Probability Plot

(continued)

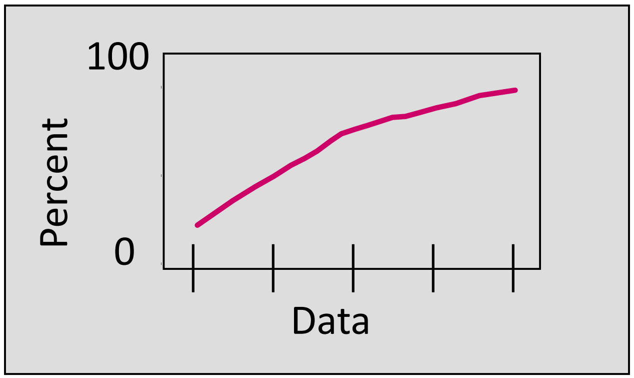
A normal probability plot for data from a normal distribution will be approximately linear:



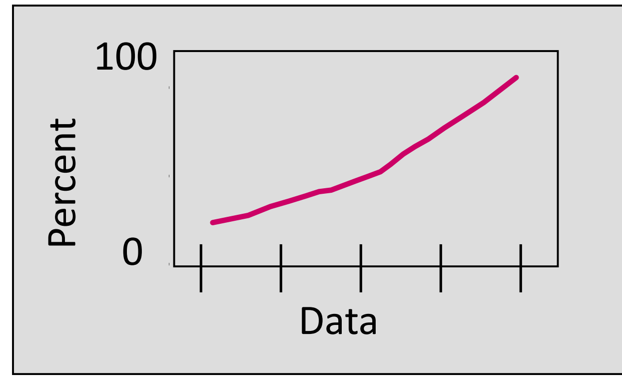
The Normal Probability Plot

(continued)

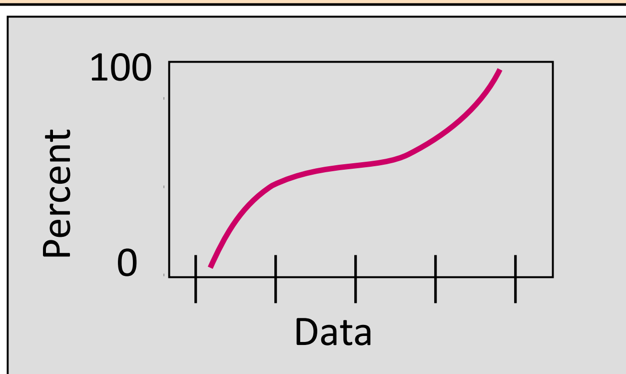
Left-Skewed



Right-Skewed



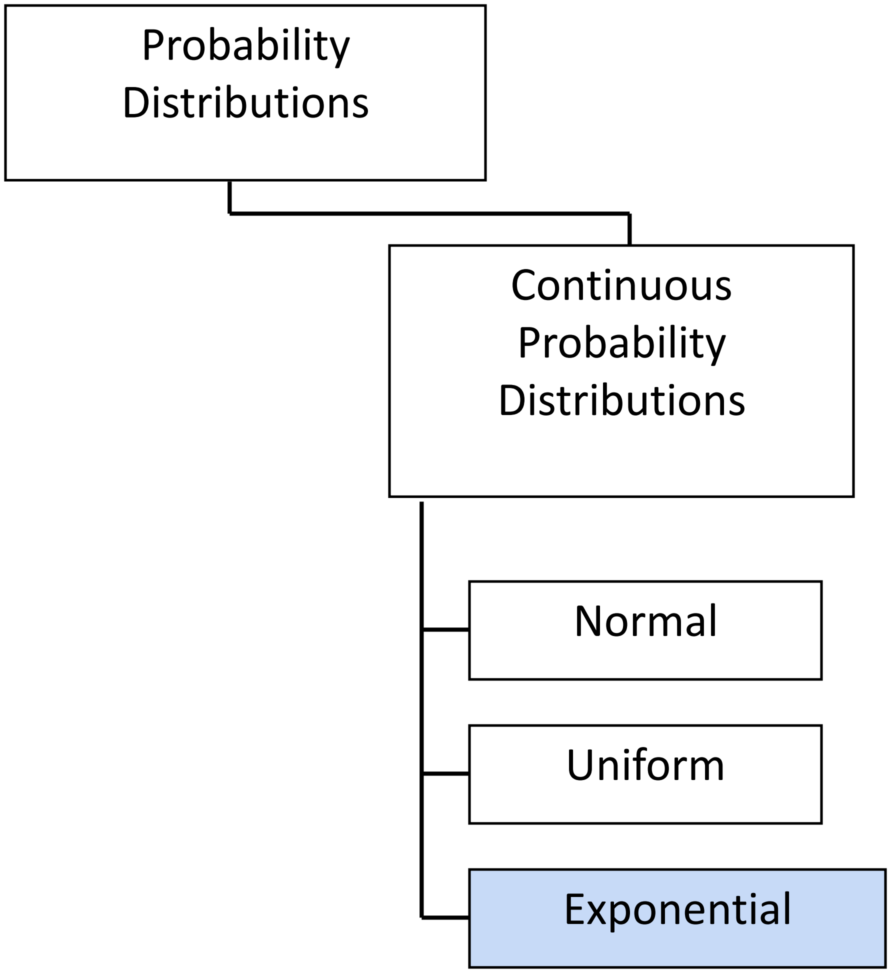
Uniform



Nonlinear plots indicate a deviation from normality

5.5

The Exponential Distribution



The Exponential Distribution

- Used to model the length of time between two occurrences of an event (the time between arrivals)
  - Examples:
    - Time between trucks arriving at an unloading dock
    - Time between transactions at an ATM Machine
    - Time between phone calls to the main operator

The Exponential Distribution

(continued)

- The exponential random variable T (t>0) has a probability density function

$$\boxed{f(t) = \lambda e^{-\lambda t} \quad \text{for } t > 0}$$

- Where
  - λ is the mean number of occurrences per unit time
  - t is the number of time units until the next occurrence
  - e = 2.71828
- T is said to follow an exponential probability distribution

The Exponential Distribution

(continued)

- Defined by a single parameter, its mean λ (lambda)

- The cumulative distribution function (the probability that an arrival time is less than some specified time t) is

$$F(t) = 1 - e^{-\lambda t}$$

where e = mathematical constant approximated by 2.71828

λ = the population mean number of arrivals per unit

t = any value of the continuous variable where $t > 0$

Exponential Distribution Example

Example: Customers arrive at the service counter at the rate of 15 per hour. What is the probability that the arrival time between consecutive customers is less than three minutes?

- The mean number of arrivals per hour is 15, so $\lambda = 15$
- Three minutes is .05 hours
- P(arrival time < .05) $= 1 - e^{-\lambda X} = 1 - e^{-(15)(.05)} = 0.5276$
- So there is a 52.76% probability that the arrival time between successive customers is less than three minutes

5.6

Jointly Distributed Continuous Random Variables

- Let $X_1, X_2, \ldots, X_k$ be continuous random variables
- Their joint cumulative distribution function,

  $$F(x_1, x_2, \ldots, x_k)$$

  defines the probability that simultaneously X_1 is less than x_1, X_2 is less than x_2, and so on; that is

$$\boxed{F(x_1, x_2, \ldots, x_k) = P(X_1 < x_1 \cap X_2 < x_2 \cap \cdots X_k < x_k)}$$

STEVENS
INSTITUTE of TECHNOLOGY
School of Business

stevens.edu

Amir H Gandomi; PhD
Assistant Professor of Analytics & Information Systems
a.h.gandomi@stevens.edu