ELC 2600H: Probability and Random Processes in Engineering (Honors)

Part I: Basic Probability Theory

- **Lecture 1: Course Introduction**
- Lecture 2: Build a Probability Model
- Lecture 3: Conditional Probability & Independence
- Lecture 4: Sequential Experiments

- Out-of-Class Reading: Counting Methods

ELEC 2600H
Probability and Random Processes in Engineering
Fall 2021 Semester

Teaching Team

Instructor

Prof. Bertram SHI

Instructional Assistant

ZHANG Xuning

Teaching Assistants

CHENG, Qingbei

SHEN, Yifan

ZHANG, Xin

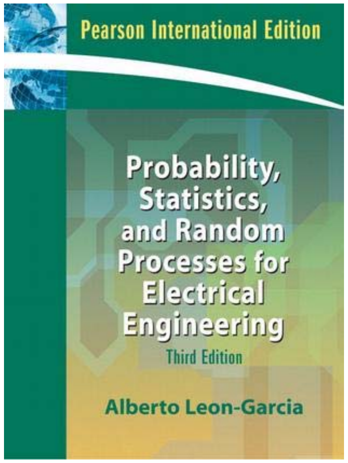
Elec 2600H: Lecture 1

- **Course Details**
- Models in Engineering
  - Their role
  - Types of models
- Relative Frequency
- Applications of Probability
- An Interesting Problem – Winning an iPad

Course Details

- Course resources available through canvas
  - Description
  - Schedule
  - Syllabus
  - Lecture notes
  - Video links
  - Grading policy
  - Contact information all available through canvas.

- Textbook:
  - *Probability, Statistics and Random Processes for Electrical Engineering*, Alberto Leon-Garcia, Addison Wesley, 3rd ed., 2009.



Course Details

- Course Structure
  - Part I: Basic probability theory. (~2-3 weeks)
  - Part II: Single random variables. (~3-4 weeks)
  - Part III: Multiple random variables. (~5 weeks)
  - Part IV: Stochastic processes. (~3 weeks)

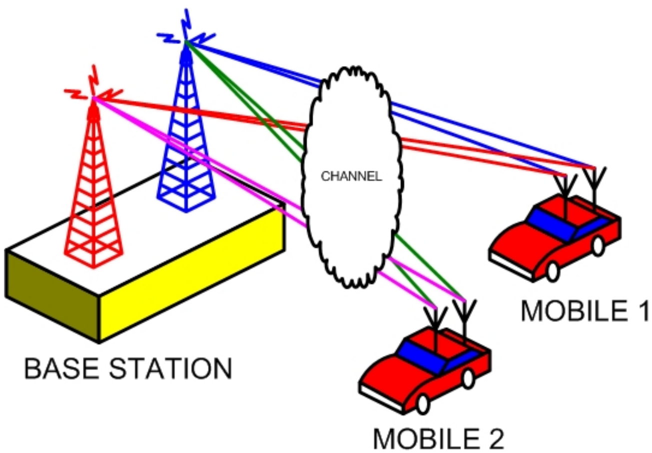
- Objective of the course
  - Basic concepts of probability theory required to **understand probability models** used in engineering.
  - Basic techniques required to **develop probability models**.

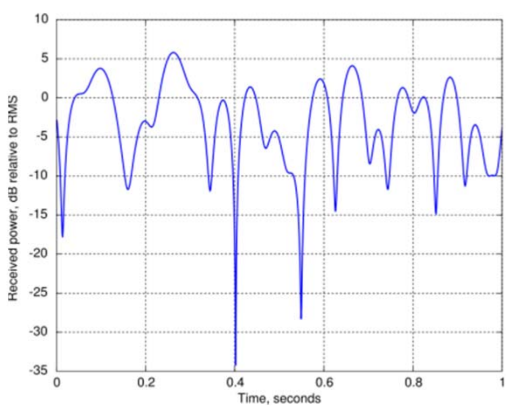
Elec 2600H: Lecture 1

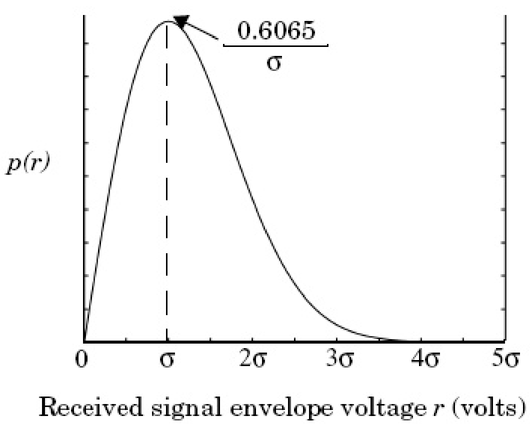
- Course details
- **Models in Engineering**
  - Their role
  - Types of models
- Relative Frequency
- Applications of Probability
- An Interesting Problem – Winning an iPad

Modelling

- In engineering as well as in daily life, we are faced with choices
- Our decisions are based on our belief – our model – of how things will behave
- A *model* is an approximate representation of a physical situation which attempts to explain observed behavior
- Good models
  - are simple and understandable
  - predict all **relevant** aspects of the situation
- Models can help us avoid costly (in time and money) experimentation







Mathematical and Computer Simulation Models

- **Mathematical model**
  - A set of assumptions about how a system or physical process works.
  - Assumptions are stated in the form of mathematical relations between the important parameters and variables.
  - By solving these relations, we can predict the outcome of an experiment, e.g. $V = I*R$
  - The simpler the model, the easier it is to solve, understand and analyze.
  - However, simple models may not accurately describe the actual system of interest.
- **Computer simulation model**
  - Assumptions are stated in the form of a computer program.
  - By running the program, we can predict the outcome of an experiment, e.g. Spice.
  - Computer models can represent systems in greater detail, and more accurately predict performance, but are generally more difficult to analyze precisely.

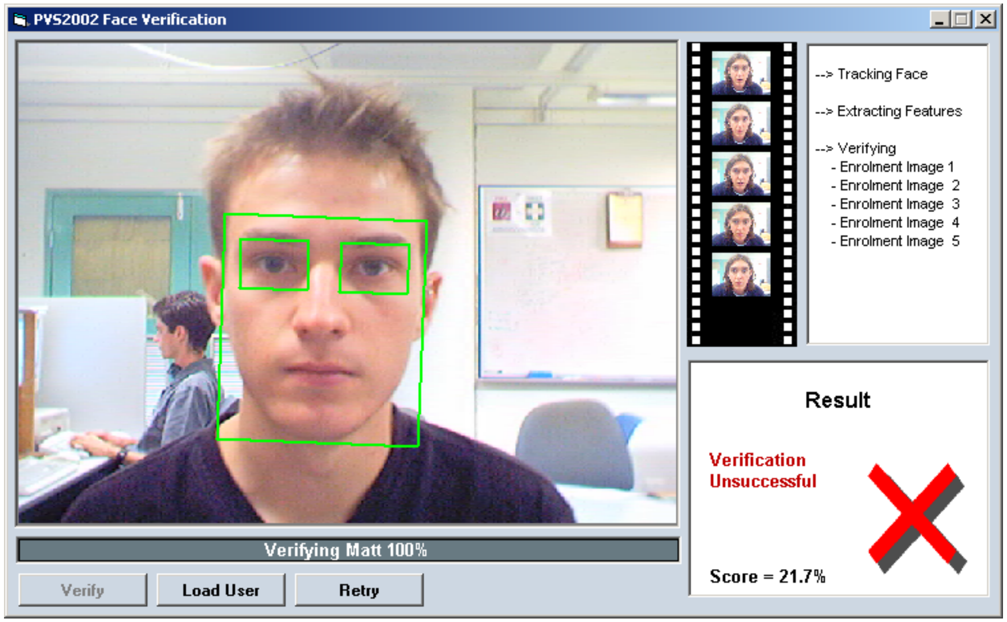
Deterministic versus Probabilistic Models

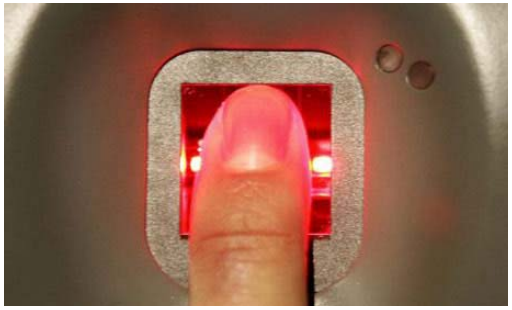
- Deterministic Model
  - The conditions of the experiment determine the **exact outcome**
  - If I do A, then B will happen.
  - Example: Ohm's Law ($V = I*R$), Kirchoff's current and voltage laws.
  - In practice, there will always be random deviations if you repeat the same experiment twice, but a deterministic model is OK if these are small.
- Probabilistic Model
  - The conditions of the experiment determine the **probability of different outcomes**.
  - If I do A, then B will probably happen, but it is *possible* that C might happen instead.
  - Example: tossing of a coin, rolling a die
  - In many cases, probabilistic models are used because we cannot model all the relevant variables required by a deterministic model.

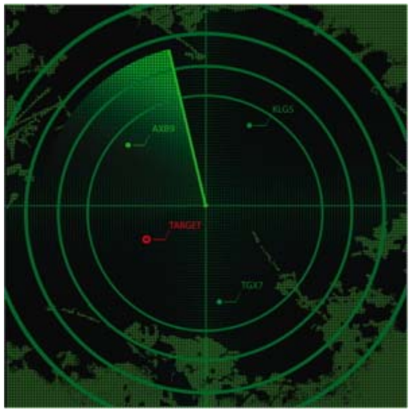
More on Probabilistic Models

- Allow designers to model very complex systems
- Nearly all practical systems cannot be described exactly, and have some sort of randomness
- Systems can be designed and optimized “off-line”, without physical implementation – **save time and $$$**
- **A complex system may involve *MANY* models!**

Other Examples





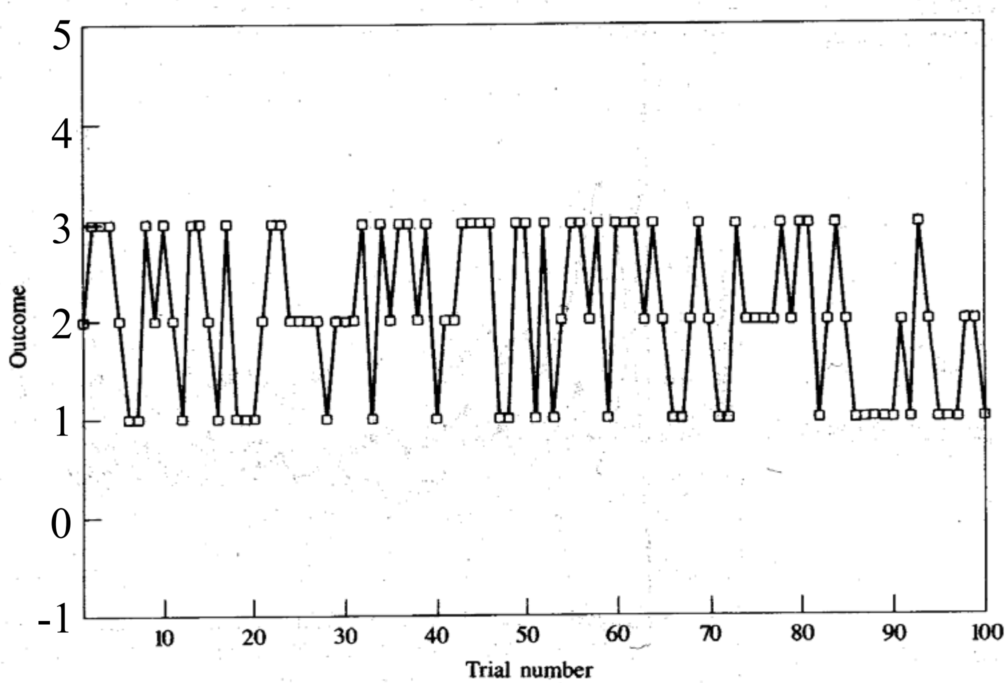


Elec 2600H: Lecture 1

- Course details
- Models in Engineering
  - Their role
  - Types of models
- **Relative Frequency**
- Applications of Probability
- An Interesting Problem – Winning an iPad

Example: Ball picking from a Urn

- An urn contains 3 balls, labeled 1, 2 or 3.
- A ball is selected, its number recorded (***outcome***) and the ball returned.
- This process is repeated for n (e.g. 100) times (***trials***).
  - **RANDOM Experiment:** Outcome cannot be exactly determined!

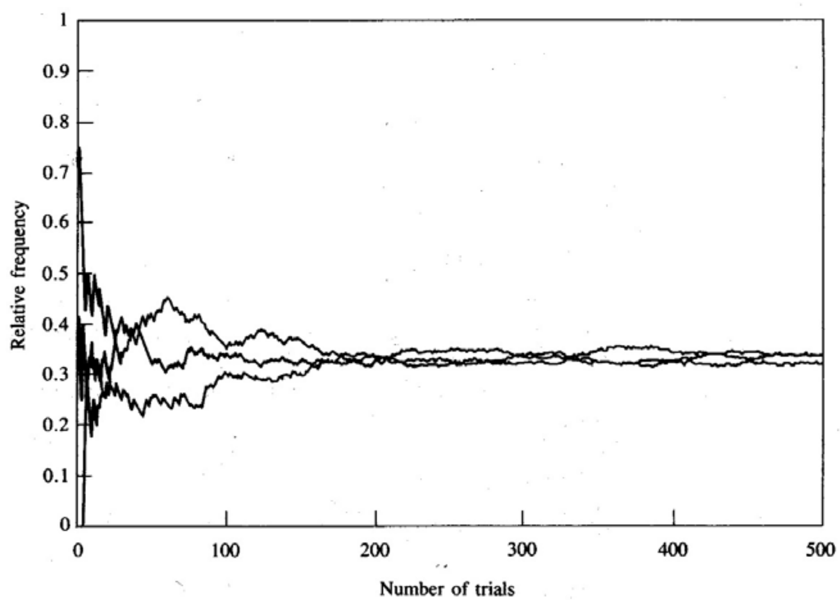


Relative Frequency

- Let $N_1(n)$, $N_2(n)$ and $N_3(n)$ be the *number of times* that we pick balls 1, 2, and 3 in n trials (***events***).
- Define the *relative frequency* of the outcome k as $f_k(n) = \frac{N_k(n)}{n}$
- This experiment exhibits *statistical regularity:* as n increases, the relative frequency approaches a ***constant value***

$$\lim_{n \to \infty} f_k(n) = p_k$$

where p_k indicates the probability of outcome k.



Provides a key connection between measurement of physical quantities and probability models!

Properties of the relative frequency (1)

- Since $0 \le N_k(n) \le n$,

$$0 \le f_k(n) = \frac{N_k(n)}{n} \le 1$$

The relative frequencies are between zero and one.

- Let K = the number of possible outcomes. Since $\sum_{k=1}^{K} N_k(n) = n$

$$\sum_{k=1}^{K} f_k(n) = \sum_{k=1}^{K} \frac{N_k(n)}{n} = 1$$

The relative frequencies sum to one.

Properties of the relative frequency (2)

- Events are groupings of the outcomes of an experiment (***sets***):
  - ONE = "the ball picked is labeled 1" = {1}
  - NOT_THREE = "the ball picked is not labeled 3" = {1 or 2}
  - ODD = "the ball picked is labelled with an odd number" = {1 or 3}
- We can often derive the relative frequency of one event from the relative frequency of other events.

  Example: Since $N_{\mathrm{ODD}}(n) = N_1(n) + N_3(n)$,

$$f_{\mathrm{ODD}}(n) = \frac{N_{\mathrm{ODD}}(n)}{n} = \frac{N_1(n) + N_3(n)}{n} = f_1(n) + f_3(n)$$

- More generally, if $C = \{A \; or \; B\}$ and A and B cannot occur simultaneously (mutually exclusive),

$$f_{\mathrm{C}}(n) = f_{\mathrm{A}}(n) + f_{\mathrm{B}}(n)$$

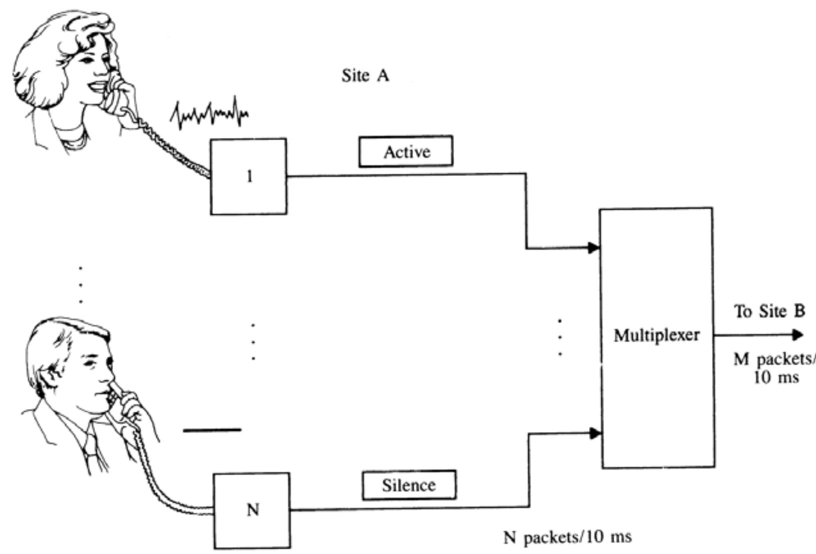
Above 3 properties coincide with AXIOMS of probability (discussed next lecture)

Elec 2600H: Lecture 1

- Course details
- Models in Engineering
  - Their role
  - Types of models
- Relative Frequency
- **Packet Voice Transmission Example**
- Applications of Probability
- An Interesting Problem – Winning an iPad

Packet Voice Transmission Example

- Suppose a communication system needs to transmit $N = 48$ simultaneous conversations using "packets" corresponding to 10ms of speech.
- Suppose each person speaks only 1/3 of the time.
- If we wish to transmit at most M packets every 10 ms, ***how should we choose M?***
- Obviously, $M = 48$ guarantees that no packets are ever lost. However, this is expensive in terms of equipment/bandwidth. In addition, most (~2/3) of the packets will be empty.



- On the other hand, if $M < 48$, some packets may get lost.
- ***How should we choose M so that on average only a small fraction (1%) of the packets are lost?***

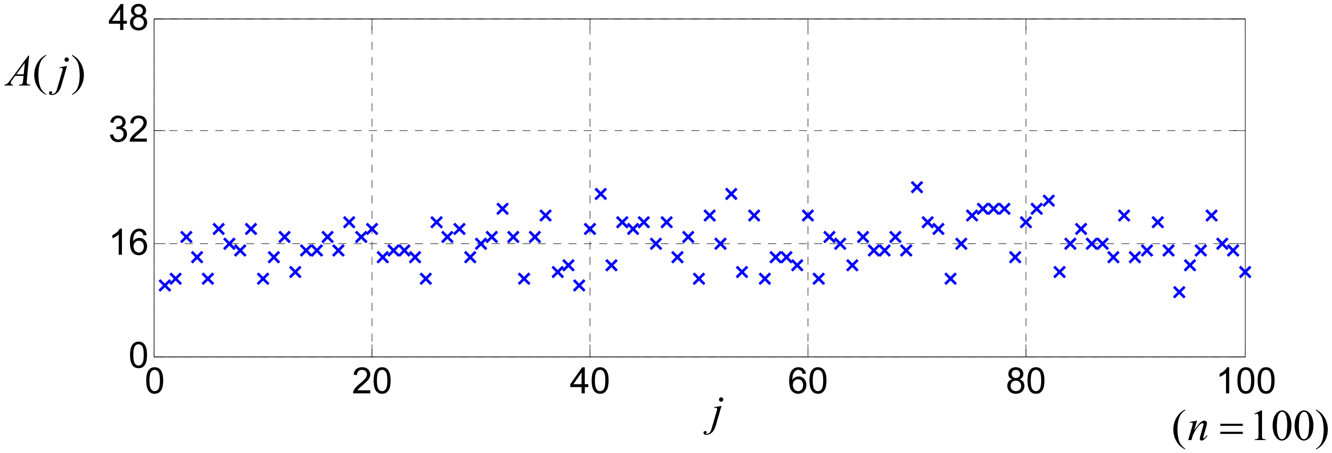
Example: Performance Metric

- We obtain good performance when the ratio below is small.

$$\text{discard fraction} = \frac{\text{average \# of active packets lost}}{\text{average \# of active packets}}$$

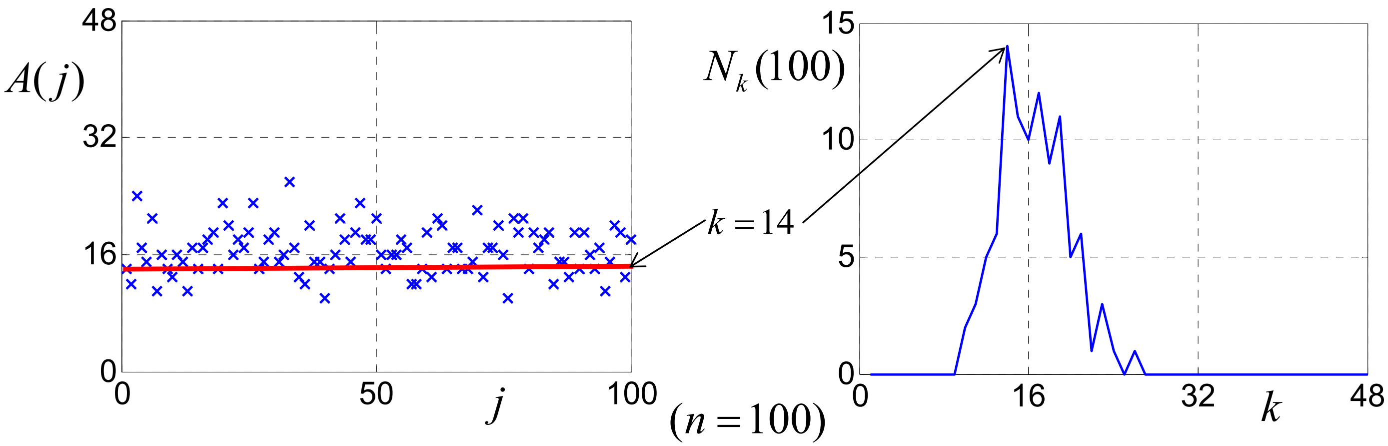
where the average is over a large number (n) of intervals.

- Let $A(j)$ be the number of active packets (i.e. speakers that speak) during the jth 10ms time interval.



Example: Average number of active packets

- Define $N_k(n)$ to be the number of intervals with k active



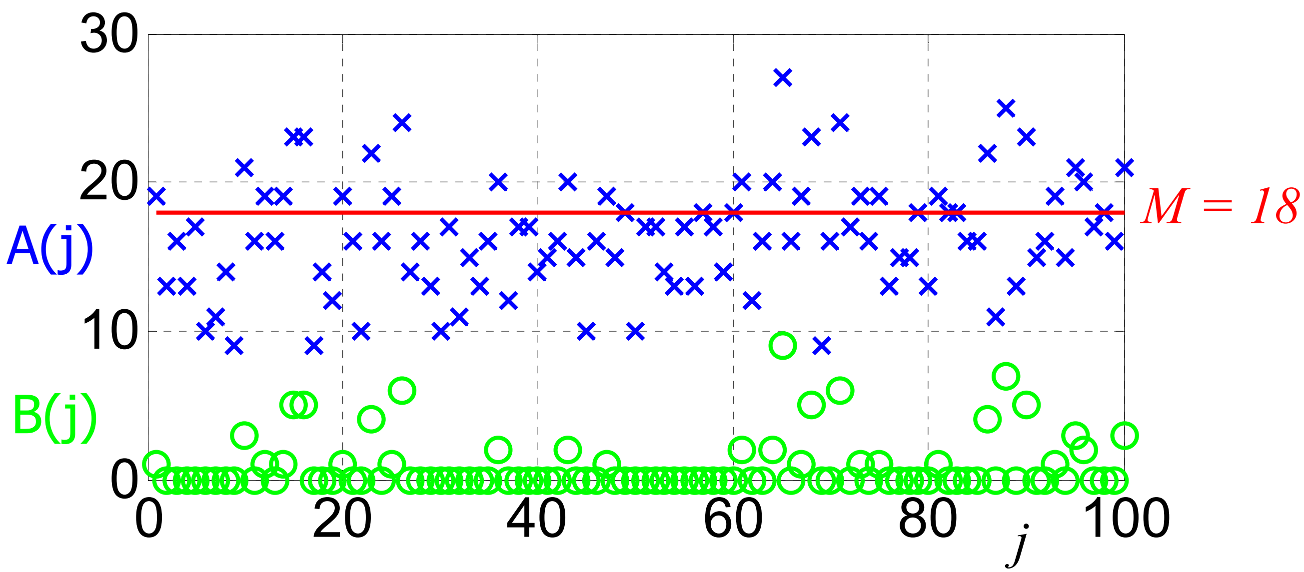
- The average number of active packets can be computed in two different ways:

$$\text{average \# of active packets} = \frac{1}{n}\underbrace{\sum_{j=1}^{n} A(j)} = \frac{1}{n}\underbrace{\sum_{k=0}^{48} k \cdot N_k(n)}$$

total number of active packets

Example: Lost packets

- Define $B(j)$ to be the number of packets lost in interval j.
- The number of lost packets is a function of the number of active packets $A(j)$.
  - If $A(j) \le M$, then all packets can be transmitted and $B(j) = 0$.
  - Otherwise, $B(j) = A(j) - M$.
  - The maximum number of lost packets is $48 - M$.



Example: Average number of lost packets

- Define $L_i(n)$ to be the number of times i packets are lost.
- As before, there are two ways to compute the average number of lost packets

$$\text{average \# of active packets lost} = \frac{1}{n}\sum_{j=1}^{n} B(j) = \frac{1}{n}\sum_{i=1}^{48-M} i \cdot L_i(n)$$

- We can also compute the average number of lost packets based on the number of active packets.
- If the number of active packets is k and $k > M$, then the number of lost packets is $i = k - M$. Thus,

$$\begin{aligned}\text{average \# of active packets lost} &= \frac{1}{n}\sum_{i=1}^{48-M} i \cdot L_i(n) \\ &= \frac{1}{n}\sum_{k=M+1}^{48} (k-M) \cdot N_k(n)\end{aligned}$$

\# packets lost

\# times k packets appear

Example: Express as relative frequency

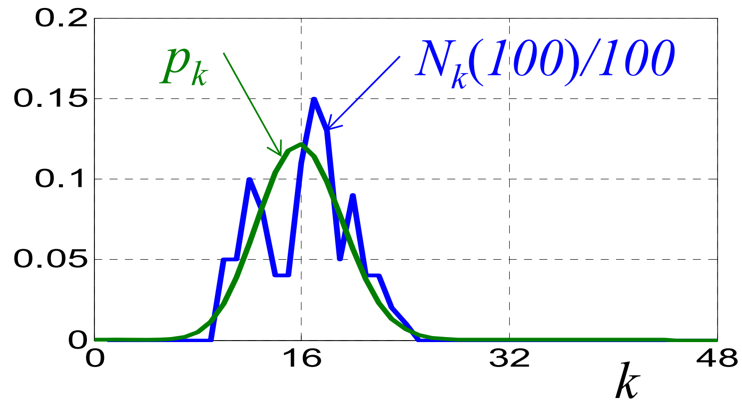
- Summarizing:

$$\text{discard fraction} = \frac{\text{average \# of active packets lost}}{\text{average \# of active packets}} = \frac{\frac{1}{n}\sum_{k=M+1}^{48}(k-M)N_k(n)}{\frac{1}{n}\sum_{k=0}^{48}k \cdot N_k(n)}$$

- Letting $n \to \infty$,

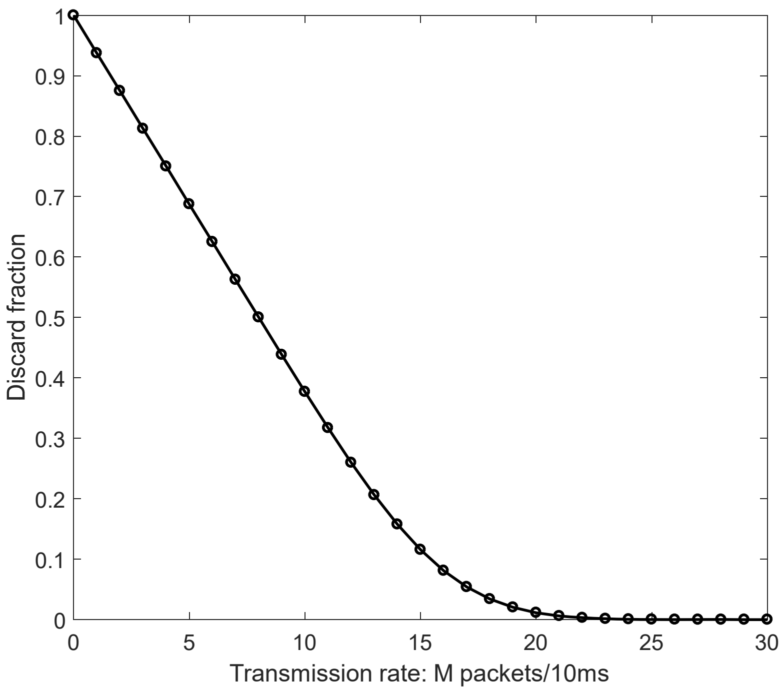
$$\text{discard fraction} = \frac{\sum_{k=M+1}^{48}(k-M)\cdot\frac{N_k(n)}{n}}{\sum_{k=0}^{48}k\cdot\frac{N_k(n)}{n}} \xrightarrow[n\to\infty]{} \frac{\sum_{k=M+1}^{48}(k-M)\cdot p_k}{\sum_{k=0}^{48}k\cdot p_k}$$

- Later in the course, we will learn how to calculate the p_k.



Example (4)

- Given the p_k, we can calculate the discard fraction

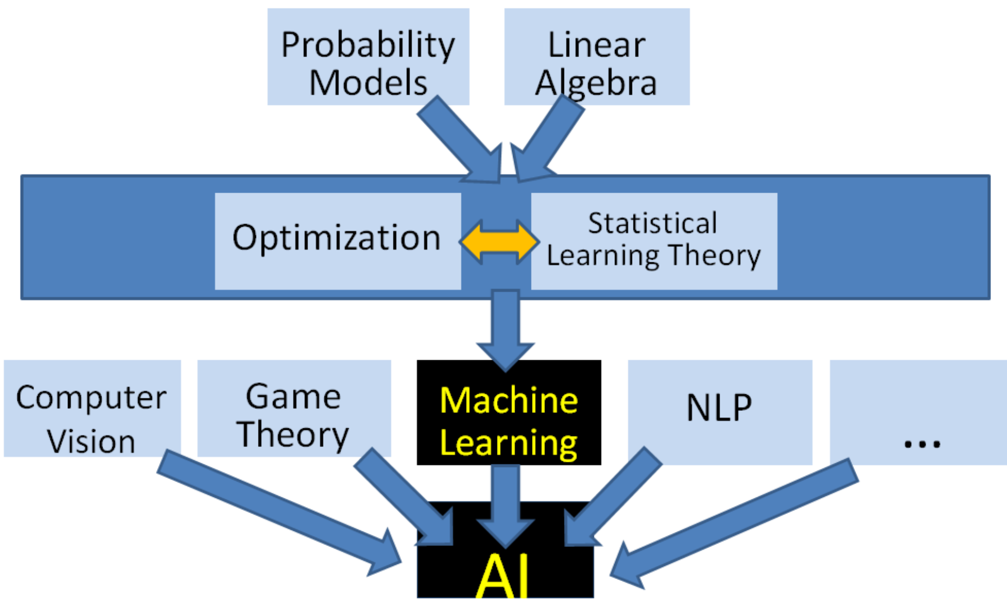


- Even if we transmit ***at most*** $M=24$ ***packets***, on average less than 1% of the active packets will be lost.

Elec 2600H: Lecture 1

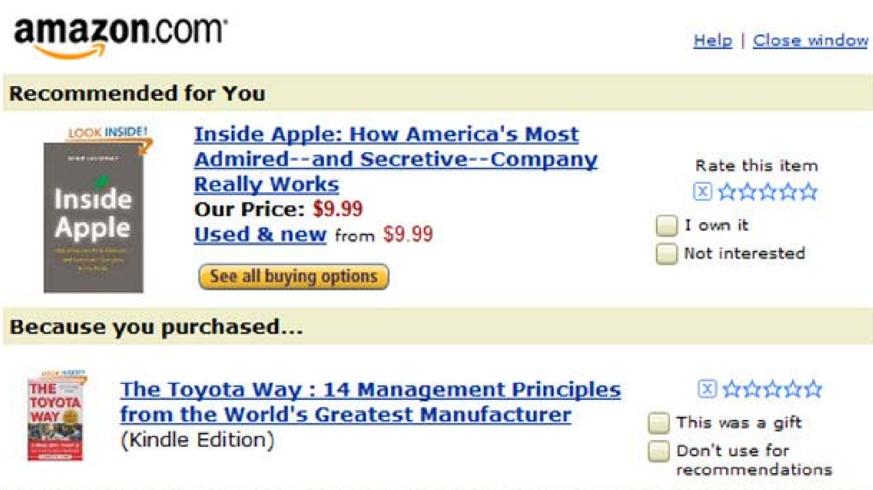
- Course details
- Models in Engineering
  - Their role
  - Types of models
- Relative Frequency
- Packet Voice Transmission Example
- **Applications of Probability**
- An Interesting Problem – Winning an iPad

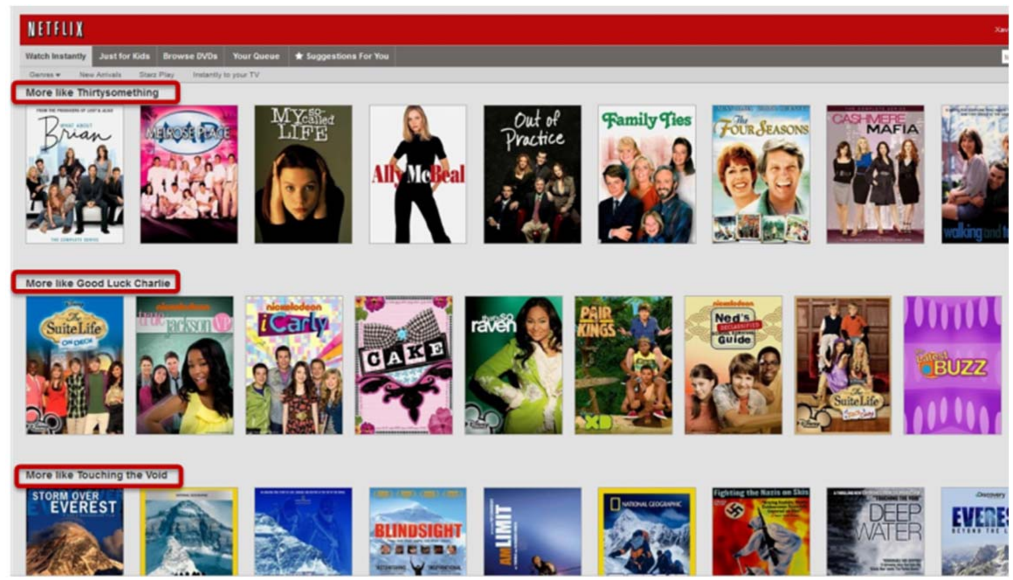
Machine Learning and AI



Reference:
https://whiteswami.wordpress.com/machine-learning/

Recommender Systems





Recommender Systems

- Beneficial to both service providers and users.
- Reduce transaction costs of finding and selecting items in an online shopping environment.
- Improve decision making process and quality.
- Enhance revenues, e.g., 35% of Amazon.com's revenue is generated by its recommendation engine.
- Three types of recommendation engines:
  - Collaborative filtering
  - Content-Based filtering
  - Hybrid recommendation systems

Reference:
1. *https://www.sciencedirect.com/science/article/pii/S1110866515000341*
2. *https://www.youtube.com/watch?v=BKCAkHn8jqA*
3. *https://www.marutitech.com/recommendation-engine-benefits/*

Recommender Systems

- Collaborative Filtering (Amazon)

Users' behaviors, Activities, preferences → Computer the Pearson correlation coefficients between users → Groups users by their similarities → Recommend items that were positively rated by users from the same group

- Content-Based Filtering (LIBRA)

Collect features extracted from items that has evaluated positively by the user → Naïve Bayes Classifier *Given sets of features, compute the probability that an item belongs to each class* → Categorize an item to the most possible class → Recommend a ranked list of items

- Hybrid recommendation systems: combine collaborative filtering and content-based filtering (Netflix)

Reference:
1. *https://www.sciencedirect.com/science/article/pii/S1110866515000341*
2. *https://www.youtube.com/watch?v=BKCAkHn8jqA*
3. *https://www.marutitech.com/recommendation-engine-benefits/*

FINANCIAL TIMES

US COMPANIES MARKETS OPINION WORK & CAREERS LIFE & ARTS

Subscribe to the FT

nd Technologies Ltd

gle's AI beats doctors at ting eye disease

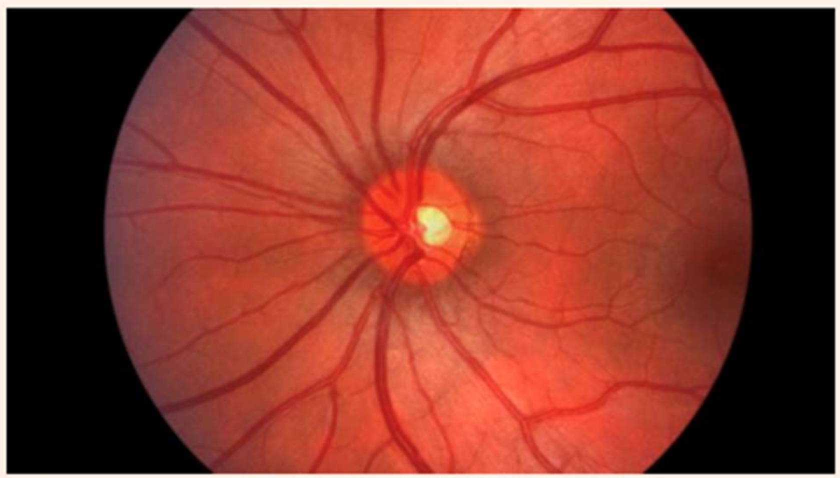
Alphabet AI unit urged to clarify its business model

DeepMind computer creates 3D model from 2D images

DeepMind Technologies Ltd + Add to myFT

DeepMind develops AI to diagnose eye diseases

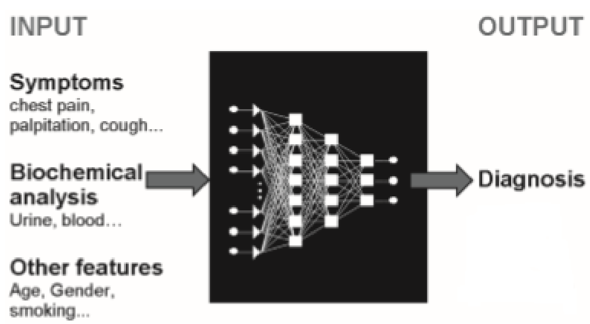
Retinal scans used to train an algorithm in 'promising' partnership with the NHS



© Alamy

Medical Diagnosis

- An efficient and accurate medical diagnosis can reduce healthcare cost.
- A medical diagnosis system will train itself based on patients' data, such as symptom, blood pressure, etc.
- The system can learn the probability distribution of a disease based on patients data.
- For a potential patient, the system computes the probability of getting a disease based on that patient's data.
- Example diagnosis systems include nephritis disease, heart disease, diabetes, etc.

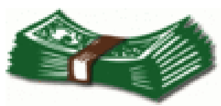


Reference:

1. *https://arxiv.org/pdf/1803.10019.pdf*
2. *http://www.zsf.jcu.cz/jab_old/11_2/havel.pdf*

More Examples

- ***Casino strategy:*** bet \$1, \$2 if losing the 1st one, \$4 if losing the 2nd one, and so on.
  - What is the winning/losing probability?



- ***Communications*** over unreliable channels: bit “0” and bit “1” may be interpreted incorrectly with an error probability
- ***Prediction*** of signals: use available signal portion to predict for the future, or to analyze and synthesize signals (eg LPC).

More Examples

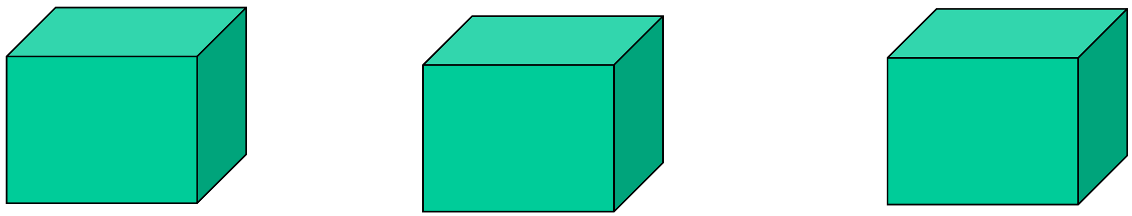
- ***System reliability:*** failure probability, average time to fail
- ***Resource-sharing*** systems: banks, telephone, etc.
- ***Internet systems:*** simple client-server, peer-to-peer
- ***Many, many more…***

Elec 2600H: Lecture 1

- Course Details
- Models in Engineering
  - Their role
  - Types of models
- Relative Frequency
- Applications of Probability
- **An Interesting Problem – Winning an iPad**

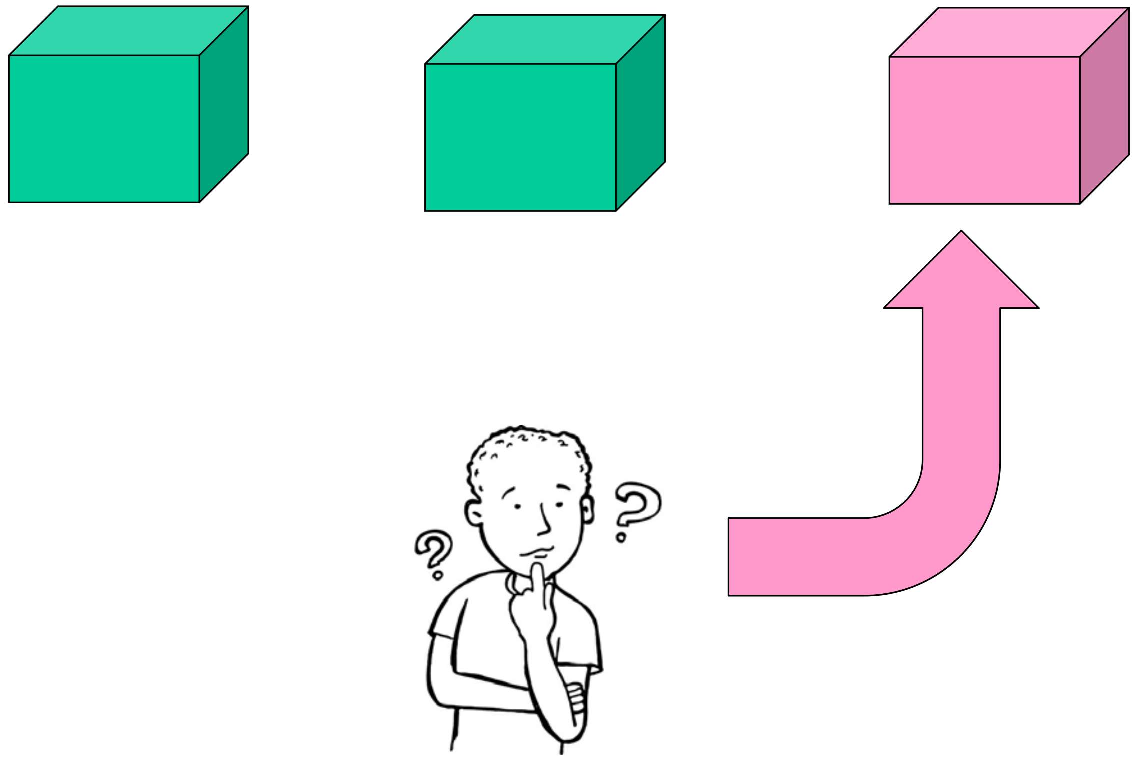
Something to think about (i):

- Suppose exactly 1 out of 3 boxes contains an **iPad**



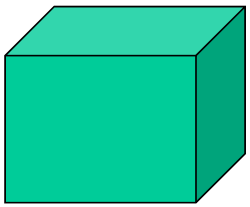
Something to think about (ii):

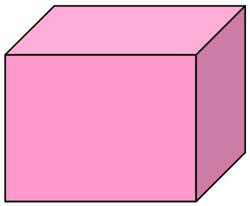
- You pick one randomly



Something to think about (iii):

- I take away one of the boxes which you did *NOT* select, and I tell you that one of the remaining boxes contains the iPad





Something to think about (iv):

- I take away one of the boxes which you did *NOT* select, and I tell you that one of the remaining boxes contains the iPad

Will you change your choice?

Why / Why not?

Summary

- This lecture we have discussed:
  - Overall course details
  - The role and types of probability models in engineering
  - The concept of relative frequency
  - Example: Winning an iPad

- Next Lecture: Random Experiments and Probability Axioms (more)!

ELEC 2600H: Probability and Random Processes in Engineering

Part I: Basic Probability Theory

- Lecture 1: Course Introduction
- Lecture 2: Build a Probability Model
- Lecture 3: Conditional Probability & Independence
- Lecture 4: Sequential Experiments

- Out-of-Class Reading: Counting Method

Lecture 2: Build a Probability Model

- Specifying Random Experiments
  - Sample spaces and events
- Set Operations
- The Three Axioms of Probability
  - Corollaries
- Probability Laws for Assigning Probabilities

Specifying Random Experiments

- A *random experiment* is specified by stating an experimental procedure and one or more measurements or observations.
- The *sample space* S is the set of all possible outcomes.
  - An outcome is a result that **cannot be decomposed into other results**. *Note that the sample space depends on the observations (compare Examples 1 and 2 or Examples 3 and 4).*
  - S is a *discrete sample space* if the number of outcomes is countable (maps to the positive integers).
  - S is a *continuous sample space* if S is not countable.
- An *event* specifies certain conditions for an outcome. It can be represented by a **subset**, A, **of the sample space** S.
  - An event A *occurs* if the outcome is a member of A.
  - The *certain event*, S , consists of all outcomes.
  - The *null event,* ϕ, contains no outcomes.
  - An *elementary event* contains only one outcome.

Discrete Sample Spaces

- *Experiment 1*: Select a ball from an urn containing balls labeled 1 to 50. Note the number on the ball.
  $S_1 = \{1, 2, 3, 4, \ldots\ldots, 49, 50\}$
  A_1 = "An even numbered ball is selected" = $\{2, 4, 6, \ldots, 48, 50\}$
- *Experiment 2*: Select a ball from an urn containing balls labeled 1 to 50. Note the number in the tens place.
  $S_2 = \{0, 1, 2, 3, 4, 5\}$
  A_2 = "The number is more than 2" $\{3, 4, 5\}$
- *Experiment 3a* : Toss a coin in the air and note which side faces up when it lands. The two sides of a coin are often called "heads" (H) and "tails" (T).
  $S_{3a} = \{H, T\}$
  A_{3a} = "The coin lands with heads up" = $\{H\}$
- *Experiment 3b:* Toss a coin three times. Note the sequence of heads/tails.
  $S_{3b} = \{HHH, HHT, HTH, THH, TTH, THT, HTT, TTT\}$
  A_{3b} = "The three tosses give the same outcome" = $\{HHH, TTT\}$

Discrete Sample Spaces

- *Experiment 4*: Toss a coin three times and note the number of heads.

  $S_4 = \{0, 1, 2, 3\}$

  A_4 = "The number of heads equals the number of tails" = ϕ

- *Experiment 5* : A block of information is transmitted repeatedly over a noisy channel until an error-free block arrives at a receiver. Count the number of transmissions required.

  $S_5 = \{1, 2, 3, 4, \ldots\ldots\ldots\}$

  A_5 = "Less than 10 transmissions are required." = $\{1, 2,.., 8, 9\}$

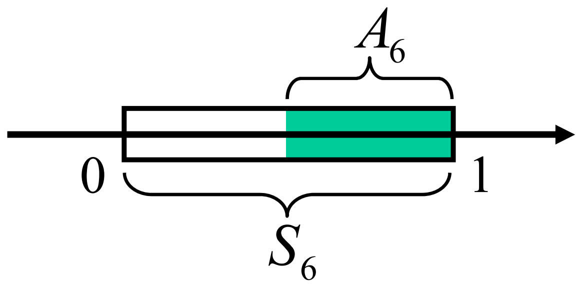
Continuous Sample Space

- *Experiment 6*: Pick a number at random between zero and one.

$$S_6 = \{x : 0 \le x \le 1\} = [0,1]$$

$$A_6 = \text{"The number is greater than 0.5"}$$

$$= \{x : 0.5 < x \le 1\} = (0.5,1]$$

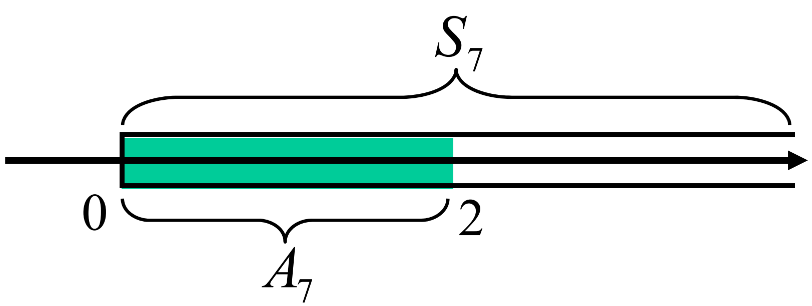


- *Experiment 7*: Measure the time between two typhoons.

$$S_7 = \{t : t \ge 0\} = [0, \infty)$$

$$A_7 = \text{"Less than 2 months elapse"}$$

$$= \{t : 0 \le t < 2\} = [0,2)$$



Elec 2600H: Lecture 2

- Specifying Random Experiments
  - Sample spaces and events
- Set Operations (Review)
- The Three Axioms of Probability
  - Corollaries
- Probability Laws for Assigning Probabilities

Set Theory

- We use set operations to **construct more complex events.**

- Interesting article(s) on the History of Set Theory

  http://www-groups.dcs.st-and.ac.uk/~history/HistTopics/Beginnings_of_set_theory.html
  http://www.absoluteastronomy.com/topics/Georg_Cantor

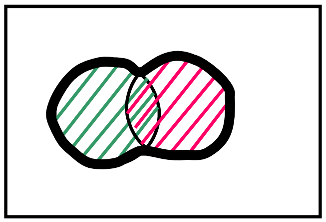
- The Father of Set Theory:

  Georg Cantor (1845-1918)

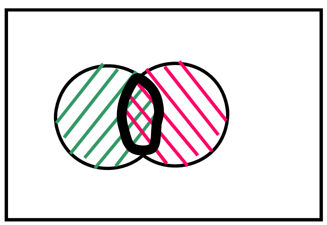
Set Operations

- We use set operations to construct more complex events.

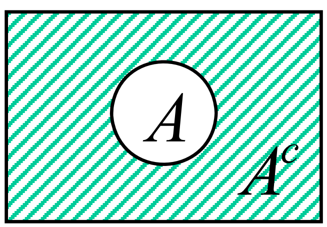
- Union $A \cup B$
"A, B or both occur"



- Intersection $A \cap B$
"Both A and B occur"



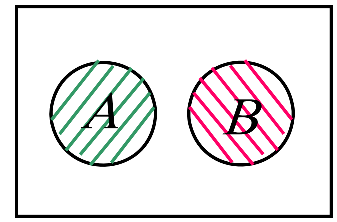
- Complement A^c or $\bar{A}$
"A does not occur"



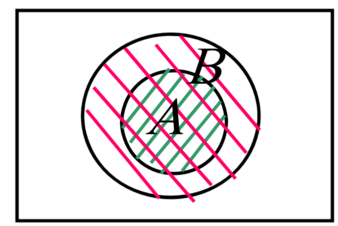
Set Relationships

- $A = B$ if they contain the same outcomes

- Mutual Exclusivity $A \cap B = \phi$
"If A occurs, then B does not"

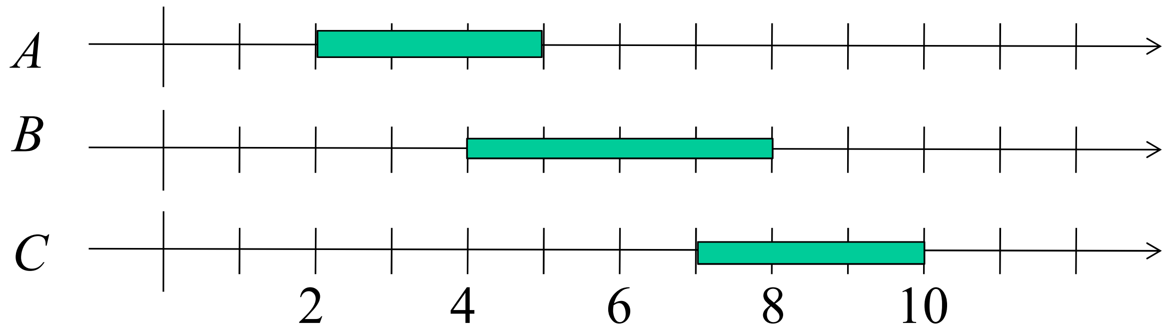


- Subset or containment $A \subset B$
"If A occurs, then B occurs"



Example

Let $A = \{x \mid 2 \le x \le 5\}$, $B = \{x \mid 4 \le x \le 8\}$, $C = \{x \mid 7 \le x \le 10\}$



$A \cup B = \{x \mid 2 \le x \le 8\}$

$A \cup C = \{x \mid 2 \le x \le 5 \text{ or } 7 \le x \le 10\}$

$A \cap B = \{x \mid 4 \le x \le 5\}$

$A \cap C = \phi$

A and C are mutually exclusive

A and B are **not** mutually exclusive

Useful Properties of Set Operations

- Commutative Properties:

$$A \cup B = B \cup A \text{ and } A \cap B = B \cap A$$

- Associative Properties:

$$A \cup (B \cup C) = (A \cup B) \cup C \text{ and } A \cap (B \cap C) = (A \cap B) \cap C$$

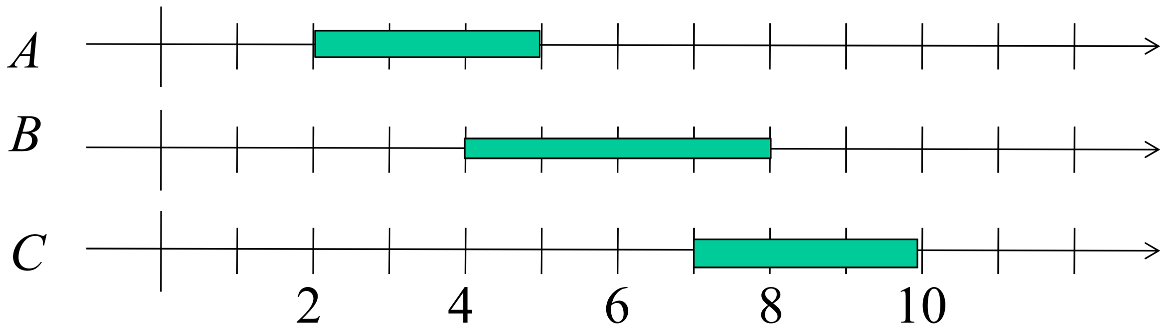
- Distributive Properties:

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \text{ and}$$
$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

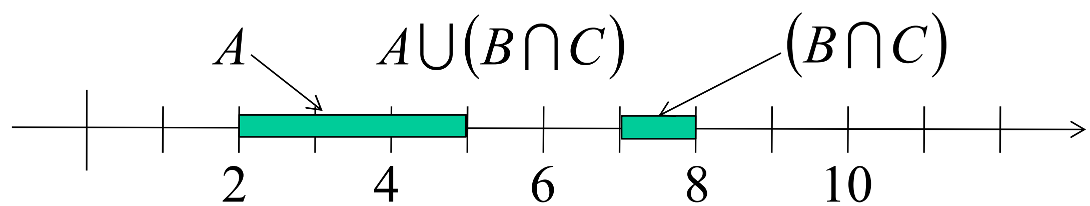
- DeMorgan's Rule:

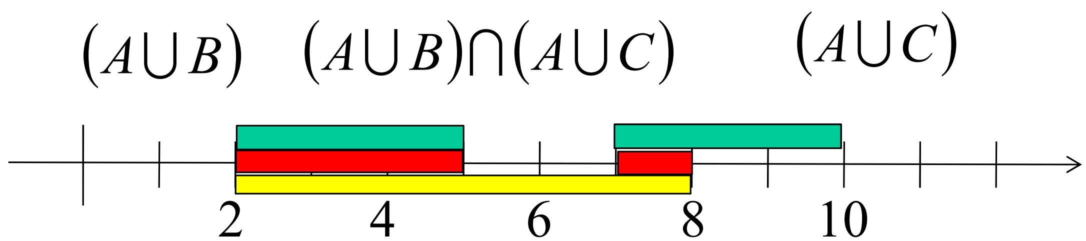
$$(A \cap B)^c = A^c \cup B^c \text{ and } (A \cup B)^c = A^c \cap B^c$$

Set Operations: Examples



Distributive Properties: $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$





Elec 2600H: Lecture 2

- Specifying Random Experiments
  - Sample spaces and events
- Set Operations
- The Three Axioms of Probability
  - Corollaries
- Probability Laws for Assigning Probabilities

Axiomatic Approach to Probability

- Assume a random experiment with sample space S. A *probability law* is a rule that assigns to each event A a number $P[A]$ that satisfies

  - **Axiom I:** $0 \leq P[A]$
  - **Axiom II :** $P[S] = 1$
  - **Axiom III:** if $A \cap B = \phi$, then $P[A \cup B] = P[A] + P[B]$

- We occasionally need a more general form of Axiom III:
  If A_1, A_2, A_3,… is a sequence of events such that $A_i \cap A_j = \phi \;\; \forall\, i \neq j$,

$$\text{then } P\left[\bigcup_{k=1}^{\infty} A_k\right] = \sum_{k=1}^{\infty} P[A_k]$$

Mass Analogy

- The probability has attributes similar to physical mass. The axioms can be interpreted using this analogy.

- Axiom 1: $0 \le P[A]$
  The probability (mass) of any event (object) is non-negative.

- Axiom 2: $P[S] = 1$
  The total probability (mass) is always equal to one.

- Axiom 3: $\text{if} \quad A \cap B = \phi, \ \text{then} \ P[A \cup B] = P[A] + P[B]$
  The total probability (mass) of two disjoint events (objects) is equal to the sum of the individual probabilities (masses).

Corollary 1

$$\boxed{P\left[A^c\right] = 1 - P[A]}$$

Proof:

By Axiom III,

$$A \cap A^c = \phi \;\Rightarrow P\left[A \cup A^c\right] = P[A] + P\left[A^c\right]$$

By Axiom II,

$$P\left[A \cup A^c\right] = P[S] = 1$$

Combining these equations,

$$P[A] + P\left[A^c\right] = 1 \Rightarrow P\left[A^c\right] = 1 - P[A]$$

Corollary 2

$$\boxed{P[A] \leq 1}$$

Proof:

By Corollary 1,

$$P[A] = 1 - P[A^c]$$

By Axiom I

$$P[A^c] \geq 0$$

Therefore,

$$P[A] \leq 1$$

Corollary 3

$$\boxed{P[\phi] = 0}$$

Proof:

Let $A = S$ and $A^c = \phi$ in Corollary 1,

$$P[\phi] = 1 - P[S] = 0$$

Corollary 4

If $A_1, A_2, \ldots A_n$ are pair-wise mutually exclusive, then $\boxed{P\left[\bigcup_{k=1}^{n} A_k\right] = \sum_{k=1}^{n} P[A_k] \quad \text{for } n \geq 2}$

Proof: Use mathematical induction.

By Axiom III, the statement is true for $n = 2$.

We assume the statement is true for $n \geq 2$, and prove it for $n+1$:

$$P\left[\bigcup_{k=1}^{n+1} A_k\right] = P\left[\left\{\bigcup_{k=1}^{n} A_k\right\} \cup A_{n+1}\right]$$

$$= P\left[\bigcup_{k=1}^{n} A_k\right] + P[A_{n+1}]$$

$$= \sum_{k=1}^{n} P[A_k] + P[A_{n+1}] = \sum_{k=1}^{n+1} P[A_k]$$

$\left\{\bigcup_{k=1}^{n} A_k\right\} \cap A_{n+1} = \phi$

by the distributive law

and our assumption of

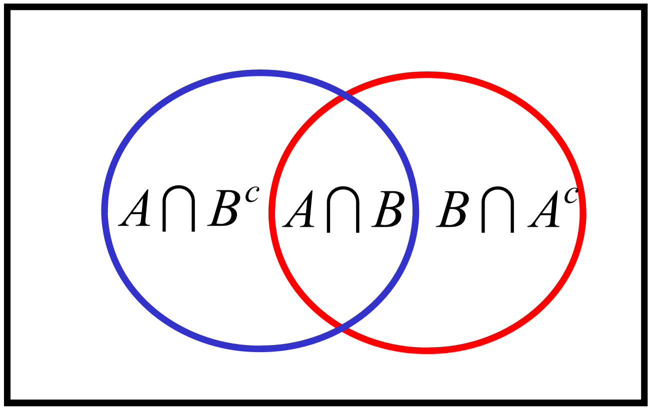
pairwise mutual exclusion

Corollary 5

For any events A and B, $\boxed{P[A \cup B] = P[A] + P[B] - P[A \cap B]}$

Proof:

Decompose $A \cup B$ as shown.



By Axiom 3, $P[A] = P[A \cap B^c] + P[A \cap B]$

$$P[B] = P[B \cap A^c] + P[A \cap B]$$

Thus, $P[A] + P[B] = \underbrace{P[A \cap B^c] + P[A \cap B] + P[B \cap A^c]}_{P[A \cup B]} + P[A \cap B]$

By Corollary 4:

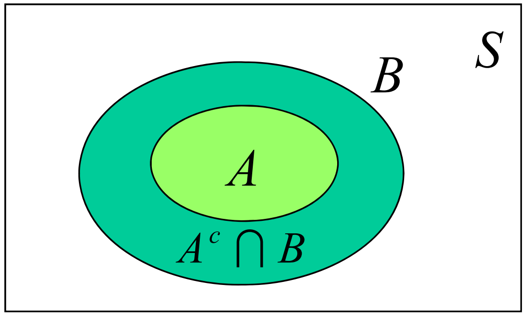
Corollary 6

If $\boxed{A \subset B, \text{ then } P[A] \le P[B]}$

Proof:

B can be split into two mutually exclusive parts: $B = A \cup (A^c \cap B)$

By Axiom III, $P[B] = P[A] + P[A^c \cap B]$



By Axiom I, $P[A^c \cap B] \ge 0$

Thus, $P[A] \le P[B]$

Elec 2600H: Lecture 2

- Specifying Random Experiments
    - Sample spaces and events
- Set Operations
- The Three Axioms of Probability
    - Corollaries
- Probability Laws for Assigning Probabilities

Probability Laws

- The axioms and corollaries provide with a set of rules for computing probabilities of events in terms of other events (*mathematical manipulation*).
- Note that any assignment of probabilities to events that satisfies the three axioms can be manipulated in this way. It is up to the **engineer** to determine a "reasonable" or "accurate" assignment of probabilities.
- Assigning probabilities to some **sub-class of events** gives you enough information to determine the probability of any event (by manipulating the axioms and corollaries).

Finite Discrete Sample Spaces

- In experiments with a finite discrete sample space, a probability law can be specified by giving the **probabilities of the elementary events** (events containing only one outcome).
- One commonly used probability law is the case of **equally likely (equiprobable) outcomes:**
  - Assume the sample space has n elements: $S = \{a_1, a_2, \ldots, a_n\}$
  - **Assign** each elementary event probability: $1/n$: $P[a_i] = \frac{1}{n}$ for all i
- For this case, the probability of any event that contains k outcomes is k/n. Thus, the probability of an event can be computed by counting the number of outcomes in it.
- Equiprobable outcomes are often said to be chosen "at random."

Example 2.9 (page 49)

- An urn contains 10 identical balls numbered 0, 1, 2, 3, ……, 9. A random experiment involves selecting a ball *at random* from the urn and noting the number on the ball.
- Consider the following events
  - o A = "ball number selected is odd"
  - o B = "ball number selected is a multiple of 3"
  - o C = "ball number selected is less than 5"
- Since
  - o A = {1, 3, 5, 7, 9} ; P[A] = 5/10
  - o B = {3, 6, 9}; P[B] = 3/10
  - o C = {0, 1, 2, 3, 4}; P[C] = 5/10
- By Corollary 5, $P\left[A\cup B\right]=P[A]+P[B]-P\left[A\cap B\right]\Rightarrow\frac{5}{10}+\frac{3}{10}-\frac{2}{10}=\frac{6}{10}$

  since $A\cap B=\{3,9\}\Rightarrow P\left[A\cap B\right]=\frac{2}{10}$

Example 2.10 (page 50)

- Suppose a coin is tossed three times and we observe the sequence of heads and tails:
  S = {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}

  If we **assume** equiprobable outcomes, then $P[\text{"2 heads in 3 tosses"}] = P[\{\text{HHT, HTH, THH}\}] = 3/8$

- Suppose a coin is tossed three times and we observe the number of heads:
  S = {0, 1, 2, 3}

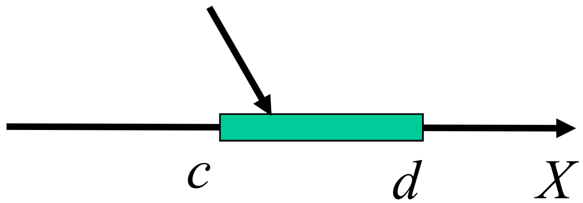
  If we **assume** equiprobable outcomes, then $P[\text{"2 heads in 3 tosses"}] = P[\{2\}] = 1/4$

- Both are mathematically valid, but **one model is more realistic…**

Continuous Sample Spaces

- For continuous sample spaces, there are an uncountably infinite number of outcomes.
- For the real line, the events of interest are intervals and their complements, unions and intersections.

$$A = [c,d] = \{X | c \le X \le d\}$$



- Examples
  - Queuing time for a data packet (network)
  - Received signal power on your mobile phone/WiFi
  - Number of minutes a student is early / late (?) to class.
  - *What about: Number of students that show up to class?*
  - *Can you think of more examples?*

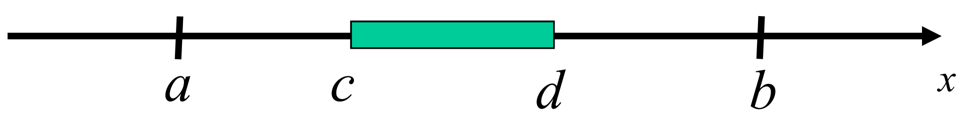
Equiprobable outcomes on the real line

- Consider an experiment where the outcome is real valued between a and b.

$$S = [a, b]$$

- A probability law that captures the idea of equiprobable events is one where the probability of an event that is a subinterval of S is equal to the length of the subinterval divided by the length of S, i.e.

$$\text{if } A = [c, d], \text{ then } \mathrm{P}[\mathrm{A}] = \frac{d-c}{b-a}$$



- It is easy to verify that the three axioms are satisfied.
- Note that the probability of any particular outcome is zero.

$$\mathbf{P}[\{x_0\}] = \mathbf{P}[[x_0, x_0]] = \frac{x_0 - x_0}{b-a} = 0$$

ELEC 2600H: Probability and Random Processes in Engineering

Part I: Basic Probability Theory

- Lecture 1: Course Introduction
- Lecture 2: Build a Probability Model
- Lecture 3: Conditional Probability & Independence
- Lecture 4: Sequential Experiments

- Out-of-Class Reading: Counting Method

Elec 2600H: Lecture 3

- **Conditional Probability**
  - Properties
  - Total Probability Theorem
  - Bayes' Rule
- Independence
- Example: Medical Testing

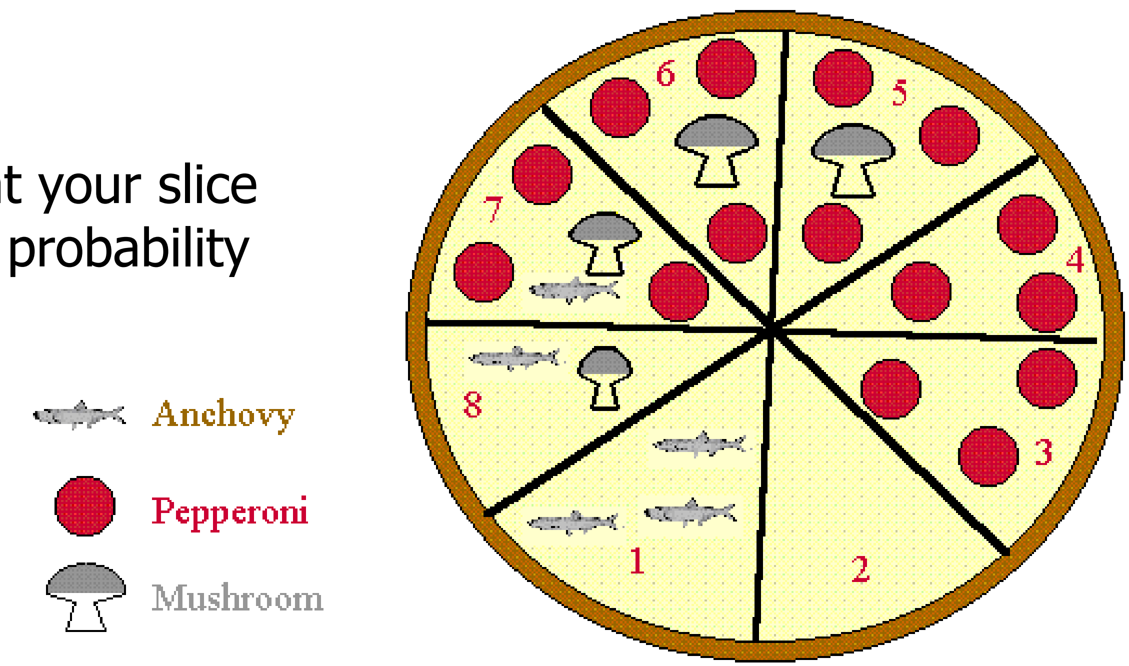
Thus far, we have looked at the probability of events occurring "individually", without regards to any other event.

However, if we KNOW that a particular event occurs, how does this change the probabilities of other events?

Pizza Time

- You are blindfolded and choose a slice of pizza. What is the probability that it contains mushrooms?

- Suppose your friend tells you that your slice contains pepperoni. What is the probability that it has mushrooms?



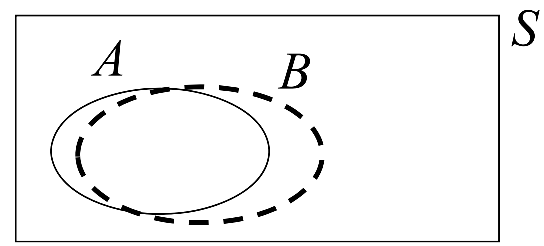
Journal of Statistics Education v.6, n.1 (1998)

Conditional Probability

- The conditional probability of event A given an event B, is defined as:

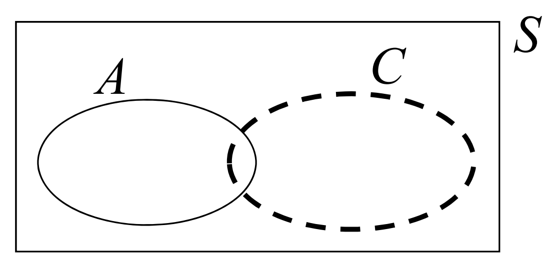
$$P[A \mid B] = \frac{P[A \cap B]}{P[B]}$$

  where we assume that $P[B] > 0$



$P[A|B] > P[A]$
Our belief that A has occurred has increased

- The conditional probability expresses how knowledge that an event B has occurred alters the probability that an event A occurs.
  - $\mathrm{P}[A]$ is the a priori probability (before we know B occurs)
  - $\mathrm{P}[A|B]$ is the a posteriori probability (after we know B occurs)

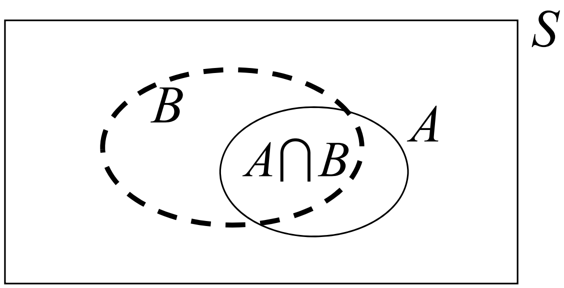


$P[A|C] < P[A]$
Our belief that A has occurred has decreased

Relative Frequency Interpretation

- If we interpret probability as relative frequency, then $P[A|B]$ is the relative frequency of the event $A \cap B$ in experiments where B occurred.
- Suppose the experiment is performed n times, and suppose that the event B occurs n_B times, and the event $A \cap B$ occurs $n_{A \cap B}$ times, the relative frequency of interest is:

$$\frac{n_{A \cap B}}{n_B} = \frac{\frac{n_{A \cap B}}{n}}{\frac{n_B}{n}} \to \frac{P[A \cap B]}{P[B]}$$



If B is known to have occurred, then A can occur only if $A \cap B$ occurs

Example

- Prior information: An integer between 1 and 10 is chosen at random. Suppose we are interested in two events:
  - A_2 = "The integer is 2"
  - A_8 = "The integer is 8"
- Since the integer is chosen at random, our prior probabilities are

$$P[A_2] = 0.1 \qquad P[A_8] = 0.1$$

- Suppose we learn that the integer chosen is greater than 5, how does that change our probabilities?
- Let B = "The number is greater than 5" = $\{6,7,8,9,10\}$

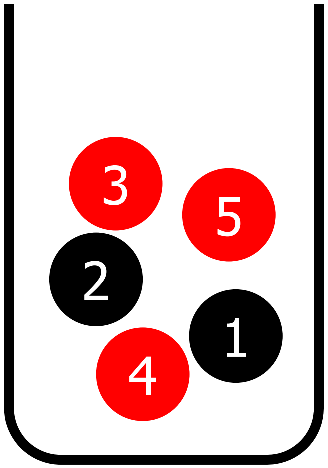
$$P[A_2 \mid B] = \frac{P[A_2 \cap B]}{P[B]} = \frac{P[\{2\} \cap \{6,7,8,9,10\}]}{P[\{6,7,8,9,10\}]} = \frac{P[\{\}]}{P[\{6,7,8,9,10\}]} = \frac{0}{0.5} = 0$$

$$P[A_8 \mid B] = \frac{P[A_8 \cap B]}{P[B]} = \frac{P[\{8\} \cap \{6,7,8,9,10\}]}{P[\{6,7,8,9,10\}]} = \frac{P[\{8\}]}{P[\{6,7,8,9,10\}]} = \frac{0.1}{0.5} = 0.2$$

Example

An urn contains two black balls numbered 1 and 2 and three red balls numbered 3 through 5. One ball is selected at random

- Find the probability that the ball is numbered 2.
- Find the probability that the ball is numbered 2, given that the ball is black.
- Find the probability that the ball is numbered 2, given that the ball is red.



Elec 2600H: Lecture 3

- **Conditional Probability**
  - **Properties**
  - Total Probability Theorem
  - Bayes' Rule

- Independence

- Example: Medical Testing

Properties of Conditional Probability

1. If $\boxed{B \subset A, \text{ then } P[A \mid B] = 1}$

Proof:

$$B \subset A \Rightarrow P[A \cap B] = P[B] \Rightarrow \frac{P[A \cap B]}{P[B]} = 1$$

2. If $\boxed{A \subset B \text{ then } P[A \mid B] = \frac{P[A]}{P[B]} \geq P[A]}$

Proof:

$$A \subset B \Rightarrow P[A \cap B] = P[A] \Rightarrow \frac{P[A \cap B]}{P[B]} = \frac{P[A]}{P[B]} \geq P[A]$$

$$\text{since } P[B] \leq 1$$

3. For fixed B, $P[A|B]$ is a probability measure

To prove property 3, we need to show that $P[A|B]$ satisfies the three axioms for probability measures.

Proof of Properties 3 of Conditional Probability

Axiom I: $P[A|B] \geq 0$

Proof $P[A \cap B] \geq 0$ and $P[B] > 0 \Rightarrow P[A \mid B] = \dfrac{P[A \cap B]}{P[B]} \geq 0$

Axiom II: $P[S|B] = 1$

Proof $S \cap B = B,\ P[S \mid B] = \dfrac{P[S \cap B]}{P[B]} = \dfrac{P[B]}{P[B]} = 1$

Axiom III: If A_1 and A_2 are mutually exclusive, then

$$P[A_1 \cup A_2 \mid B] = P[A_1 \mid B] + P[A_2 \mid B]$$

Proof $P[A_1 \cup A_2 \mid B] = \dfrac{P[(A_1 \cup A_2) \cap B]}{P[B]} = \dfrac{P[A_1 B \cup A_2 B]}{P[B]}$

Since A_1 and A_2 are mutually exclusive, so are A_1B and A_2B,

therefore $P[A_1 \cup A_2 \mid B] = \dfrac{P[A_1 B]}{P[B]} + \dfrac{P[A_2 B]}{P[B]} = P[A_1 \mid B] + P[A_2 \mid B]$

Why use Conditional Probability?

- To simplify the computation of desired probabilities
  - Joint probability: Example 2.25
  - Total probability: Example 2.27

- To change our model based on observations.
  - Example of Bayes' theorem

Computing **Joint Probability** from Conditional Probability

- Start with the definition of the conditional probability, $P[A \mid B] = \dfrac{P[A \cap B]}{P[B]}$
- Multiplying both sides by $P[B]$, $P[A \cap B] = P[A|B] \times P[B]$
- This is known as the product rule.

switch B and A

- Similarly, it can be shown that $P[A \cap B] = P[B|A] \times P[A]$

Example 2.25

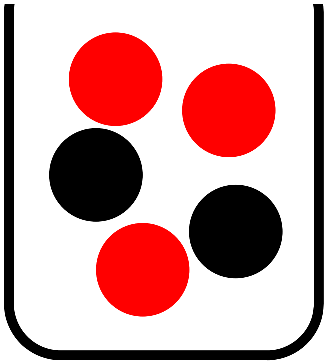
An urn contains two black balls and three red balls. Two balls are selected at random without replacement. Find the probability that both balls are black.

Solution

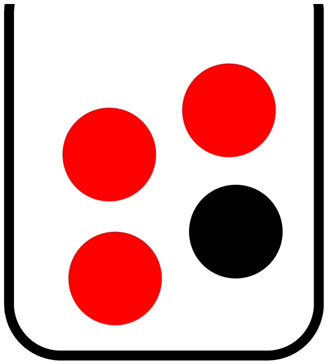
Let B_1 = "first ball black"
Let B_2 = "second ball black"

By the product rule:

$$P[B_1 \cap B_2] = P[B_2|B_1] \times P[B_1] = \frac{1}{4} \times \frac{2}{5} = \frac{1}{10}$$



At start



After B_1

Elec 2600H: Lecture 3

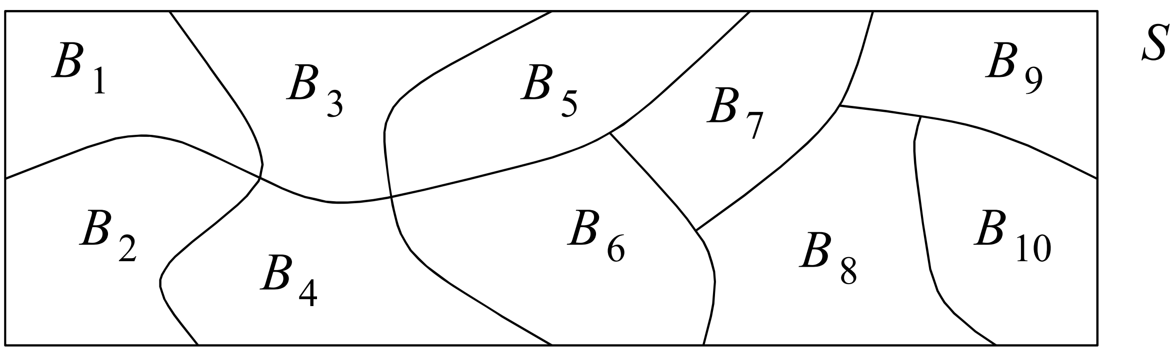
- **Conditional Probability**
  - Properties
  - **Total Probability Theorem**
  - Bayes' Rule

- Independence

- Example: Medical Testing

Partitioning the Sample Space

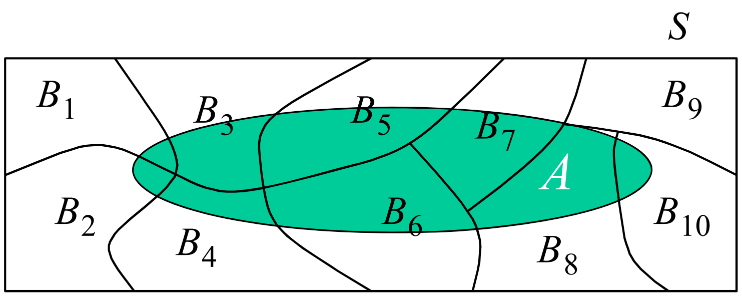
- Definition: $B_1, B_2, B_3, \ldots\ldots, B_n$ form a ***partition* of the sample space** S if both of the following are true
  - $B_1, B_2, B_3, \ldots, B_n$ are mutually exclusive
  - Their union equals S, i.e., $\bigcup_{i=1}^{n} B_i = S$



Total Probability Theorem

If $B_1, B_2, B_3, \ldots\ldots, B_n$ partition the sample space, then for any event A,

$$P[A] = \sum_{i=1}^{n} P[A \mid B_i] P[B_i]$$



Proof:

$$A = A \cap S = A \cap \left(\bigcup_{i=1}^{n} B_i \right) = \bigcup_{i=1}^{n} AB_i$$

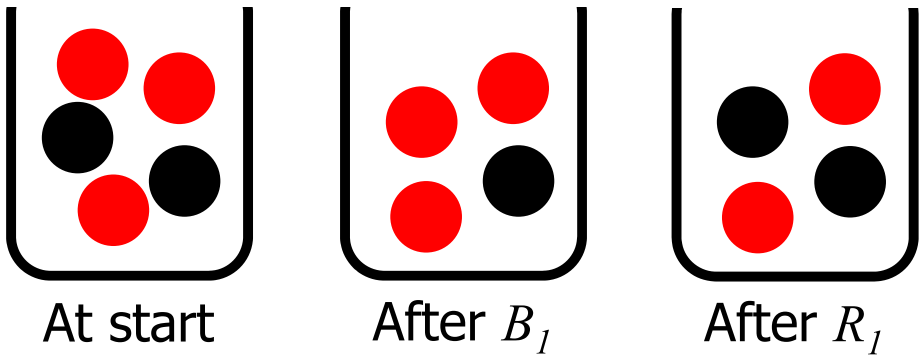
where $AB_1, AB_2, \ldots, AB_n$ are mutually exclusive.

- By Corollary 4, $P[A] = \sum_{i=1}^{n} P[AB_i] = \sum_{i=1}^{n} P[A|B_i] P[B_i]$

- This theorem lets us split the problem of computing the probability of A into smaller (easier) sub-problems!

Example 2.27

An urn contains two black balls and three red balls. Two balls are selected at random without replacement. What is the probability of the event R_2 = "the second ball is red"?



Solution

Let B_1 = "the first ball is black"

Let R_1 = "the first ball is red"

Let R_2 = "the second ball is red"

Since B_1 and R_1 partition the sample space,

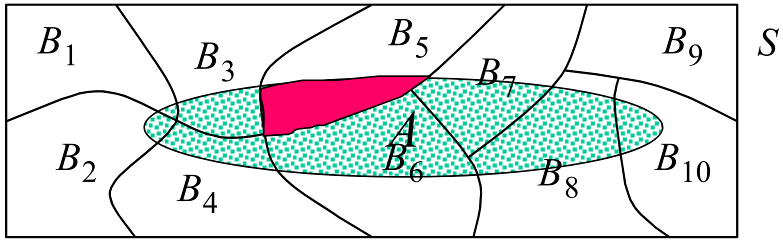
$$P[R_2] = P[R_2|B_1]P[B_1] + P[R_2|R_1]P[R_1] = \frac{3}{4}\times\frac{2}{5} + \frac{2}{4}\times\frac{3}{5} = \frac{12}{20} = \frac{3}{5}$$

Elec 2600H: Lecture 3

- **Conditional Probability**
  - Properties
  - Total Probability Theorem
  - **Bayes' Rule**
- Independence
- Example: Medical Testing

Reverend Thomas Bayes

Bayes' Rule



Let $B_1, B_2, B_3, \ldots, B_n$ be a partition of the sample space S. For any event A,

$$P[B_j \mid A] = \frac{P[AB_j]}{P[A]} = \frac{P[A \mid B_j]P[B_j]}{\sum_{k=1}^{n} P[A \mid B_k]P[B_k]}$$

- Bayes' rule lets us ***switch*** which event is the conditioning event. This simplifies the calculation when the $P[A|B_k]$ are easy to compute.

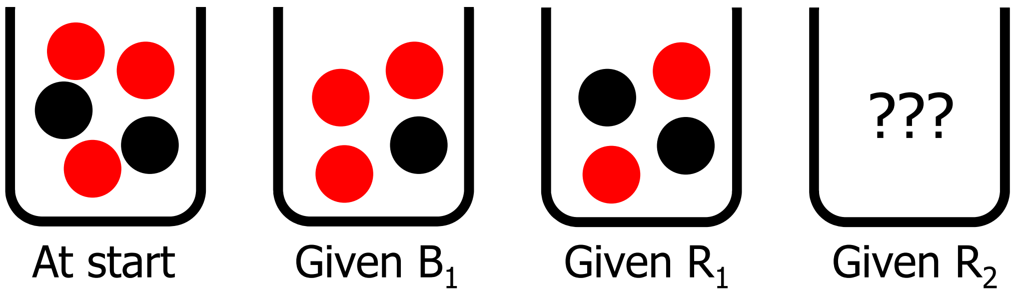
Example of Bayes Theorem

An urn contains two black balls and three red balls. Two balls are selected at random without replacement. What is the probability that we picked a black ball first, given that we picked a red ball second, $P[B_1|R_2]$?

Solution

$$P[B_1|R_2] = \frac{P[R_2|B_1]P[B_1]}{P[R_2|B_1]P[B_1] + P[R_2|R_1]P[R_1]}$$

$$= \frac{\frac{3}{4} \times \frac{2}{5}}{\frac{3}{4} \times \frac{2}{5} + \frac{2}{4} \times \frac{3}{5}} = \frac{1}{2}$$



Since $P[B_1|R_2] = \frac{2}{4} > \frac{2}{5} = P[B_1]$, knowing we picked a red ball second increases our belief that we picked a black ball first.

Elec 2600H: Lecture 3

- Conditional Probability
  - Properties
  - Total Probability Theorem
  - Bayes' Rule

- **Independence**

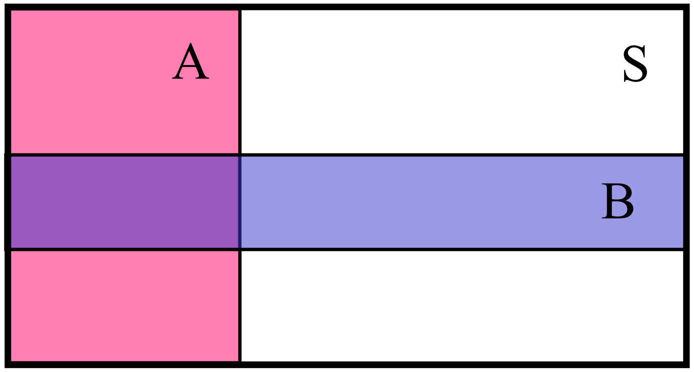
- Example: Medical Testing

Independence of Events

- Two events A and B are *independent* if $P[AB]$ = $P[A]$ x $P[B]$.
- An event A is independent of event B if knowledge that B occurs does not alter the probability of event A and vice versa:

$$P[A \mid B] = \frac{P[A \cap B]}{P[B]} = \frac{P[A]P[B]}{P[B]} = P[A]$$
$$P[B \mid A] = \frac{P[A \cap B]}{P[A]} = \frac{P[A]P[B]}{P[A]} = P[B]$$

- In other words, the *a posteriori* probability equals the *a priori* probability.



Example 2.31

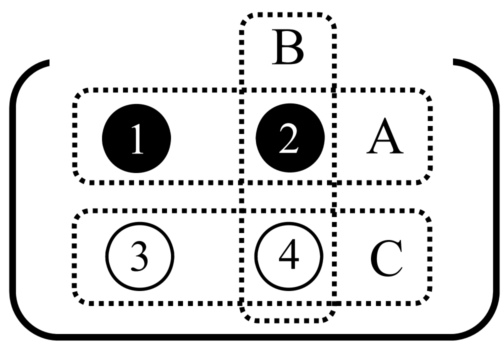
- A ball is selected from an urn containing two black balls, numbered 1 and 2, and two white balls, numbered 3 and 4.

  The event space is: $S = \{(1,b), (2,b), (3,w), (4,w)\}$

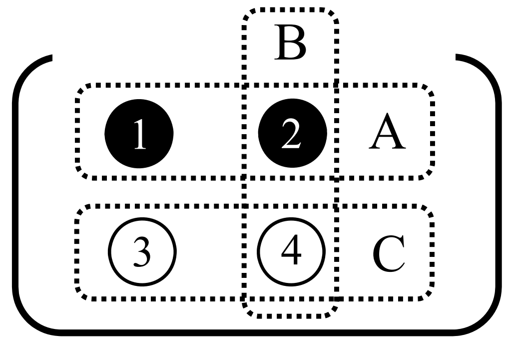
  Let events A, B and C be defined as follows:

  $A = \{(1,b), (2,b)\}$ "a black ball is selected"
  $B = \{(2,b), (4,w)\}$ "an even-numbered ball is selected"
  $C = \{(3,w), (4,w)\}$ "number on ball selected is greater than 2"

- Are A and B independent? Are A and C independent?



Example 2.31 - Solution



- A, B and C have equal probabilities

$$P[A] = P[B] = P[C] = \frac{1}{2}$$

- A and B are independent

$$P[A \cap B] = P[\{(2,\mathrm{b})\}] = \frac{1}{4} = P[A]P[B]$$

  - Another way to look at it: The proportion of outcomes in S that leads to A is equal to the proportion of outcomes in B that leads to A.

$$P[A \mid B] = \frac{P[A \cap B]}{P[B]} = \frac{P[\{(2,\mathrm{b})\}]}{P[\{(2,\mathrm{b}),\ (4,\mathrm{w})\}]} = \frac{1/4}{2/4} = \frac{1}{2} = P[A]$$

- A and C are dependent (not independent)

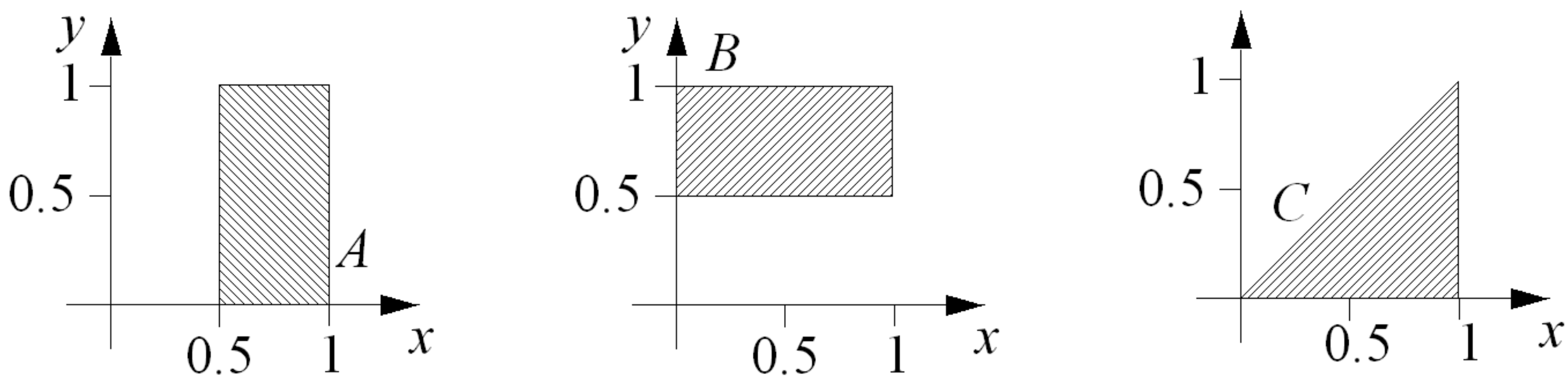
$$P[A \cap C] = 0 \neq \frac{1}{4} = P[A]P[C]$$

  - Intuition: Since A and C are mutually exclusive, knowledge that A occurs, rules out the possibility that C occurs.

$$P[A \mid C] = 0 \neq P[A]$$

Example (Continuous Sample Space)

- Suppose two numbers (X and Y) are chosen at random on $[0,1]$. In this case, the probability of any event, is equal to its area in the (X,Y) plane.
- Consider the following events:



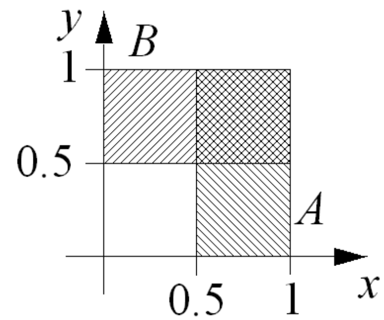
$A = \{X > 0.5\}$ $B = \{Y > 0.5\}$ $C = \{X > Y\}$

- The probability of each event is equal to its area. All events have the same probability of 0.5.

Example (cont.)

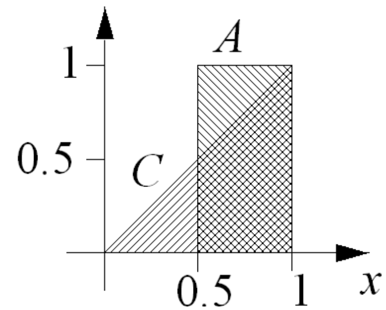
- A and B are independent

$$P[A \mid B] = \frac{P[AB]}{P[B]} = \frac{1/4}{1/2} = \frac{1}{2} = P[A]$$



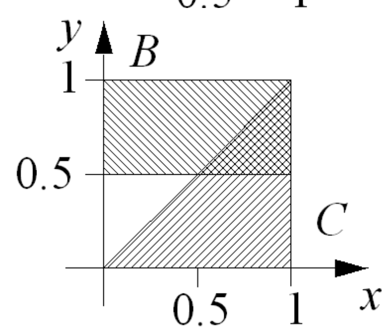
- A and C are dependent

$$P[A \mid C] = \frac{P[AC]}{P[C]} = \frac{3/8}{1/2} = \frac{3}{4} > P[A]$$



- B and C are dependent

$$P[B \mid C] = \frac{P[BC]}{P[C]} = \frac{1/8}{1/2} = \frac{1}{4} < P[B]$$



Independence of 3 or more events

- Events A_1, A_2 and A_3 are independent if and only if
  - They are pairwise independent:
    - $P[A_1A_2] = P[A_1] \times P[A_2]$
    - $P[A_2A_3] = P[A_2] \times P[A_3]$
    - $P[A_1A_3] = P[A_1] \times P[A_3]$
  - $P[A_1A_2A_3] = P[A_1] \times P[A_2] \times P[A_3]$

- Events $A_1, A_2, \ldots, A_n$ are independent if and only if
  - Any group of k of them are independent for any $k < n$.
  - $P[A_1A_2 \ldots An] = P[A_1] \times P[A_2] \times \cdots \times P[A_n]$

Example

Toss two six sided die. Are the following events independent?

A = "first die shows 3"

B = "sum of die equals 7"

C = "second die shows 2"

Solution

Note that $P[A] = P[B] = P[C] = 1/6$.

Test pairwise independence:

$P[AB] = P[\{3,4\}] = 1/36 = P[A] \times P[B]$

$P[AC] = P[\{3,2\}] = 1/36 = P[A] \times P[C]$

$P[BC] = P[\{5,2\}] = 1/36 = P[B] \times P[C]$

But,

$P[ABC] = 0 \neq P[A] \times P[B] \times P[C]$

Thus, the events are **not** independent.

Example 2.35 (System Reliability)

- A system consists of a controller and three peripheral units. The system is said to be "up" if the controller and at least two of the peripherals are functioning. Find the probability that the system is up, assuming that the components fail independently.

Define the following events

A = "the controller is functioning"

B_i = "peripheral i is functioning"

Assume the failure probabilities are $P[A^c] = p$ and $P[B_i^c] = a$.

This implies that $P[A] = (1\text{-}p)$ and $P[B_i] = (1\text{-}a)$.

Example 2.35 (cont.)

Let $F =$ "at least two peripheral units are functioning"

$=$"two units are functioning or three units are functioning"

$$=(B_1 \cap B_2 \cap B_3^c) \cup (B_1 \cap B_2^c \cap B_3) \cup (B_1^c \cap B_2 \cap B_3) \cup (B_1 \cap B_2 \cap B_3)$$

independent mutually exclusive

$$P[F] = P[B_1]P[B_2]P[B_3^c] + P[B_1]P[B_2^c]P[B_3] + P[B_1^c]P[B_2]P[B_3] +$$
$$P[B_1]P[B_2]P[B_3]$$
$$= (1-a)^2 a + (1-a)^2 a + a(1-a)^2 + (1-a)^3$$
$$= 3(1-a)^2 a + (1-a)^3$$

Thus, $P[\text{"system up"}] = P[A \cap F] = P[A]P[F]$

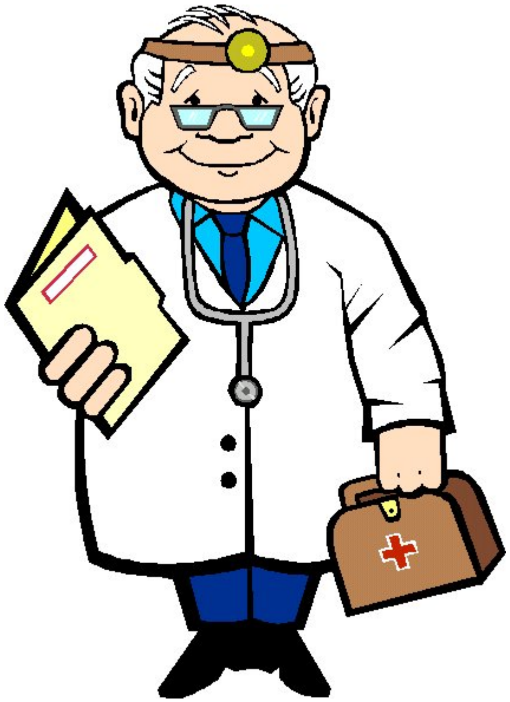
$$= (1-p)\{3(1-a)^2 a + (1-a)^3\}$$

If $a = 10\%$ and $p = 20\%$, then $P[F] = 97.2\%$

and $P[\text{"system up"}] = 77.8\%$

Elec 2600H: Lecture 3

- Conditional Probability
  - Properties
  - Total Probability Theorem
  - Bayes' Rule
- Independence
- **Example: Medical Testing**



Medical Testing

- Tests are commonly used to determine whether or not a person has a disease.
- There are two measures commonly used to quantify test accuracy: sensitivity and specificity

The Truth

Test Score:	Has the disease	Does not have the disease
Positive	True Positives (TP)	False Positives (FP)
Negative	False Negatives (FN)	True Negatives (TN)

Sensitivity

$$\frac{TP}{TP + FN}$$

Specificity

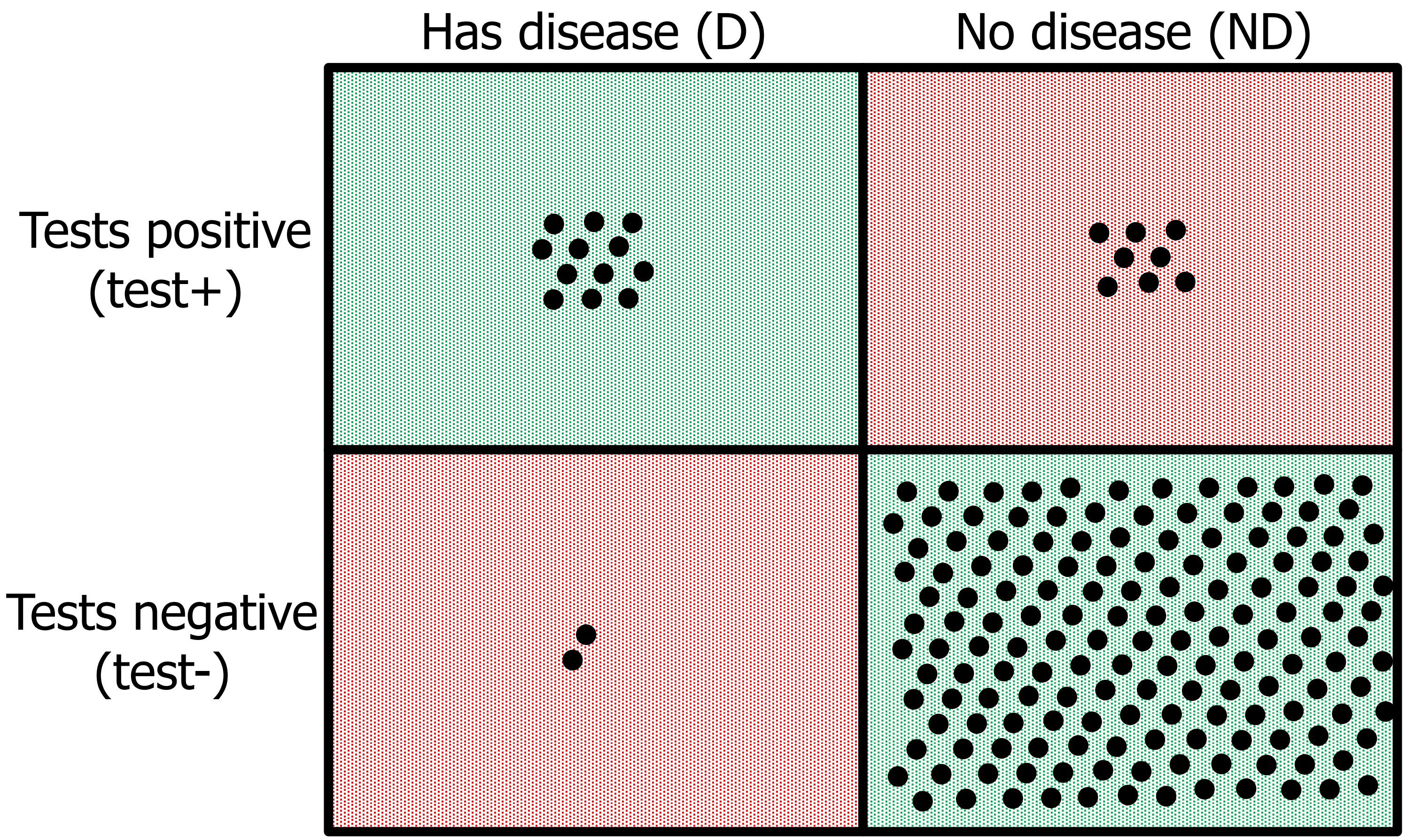
$$\frac{TN}{TN + FP}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

Sensitivity = P[test+ | has disease]

Specificity = P[test- | no disease]

The usual case



Example: Probability of a positive test result

- Suppose that 1% of the population has a disease.
- Suppose there is a test for the disease, which has
  - Sensitivity = P[test+ | has disease] = 99%
  - Specificity = P[test- | no disease] = 99%
- If random person is tested, what is P[test+]? (<1%, 1%, >1%)

Solution:

What if the sensitivity increases? Does $P[\text{test}+]$ increase or decrease?

What if the specificity increases? Does $P[\text{test}+]$ increase or decrease?

Example: Probability

- Suppose that 1% of the population has a disease.
- Suppose there is a test for the disease, which has
  - Sensitivity = P[test+ | has disease] = 99%
  - Specificity = P[test- | no disease] = 99%
- If random person tests positive, what is P[D|test+]? (<99%, 99%, >99%)

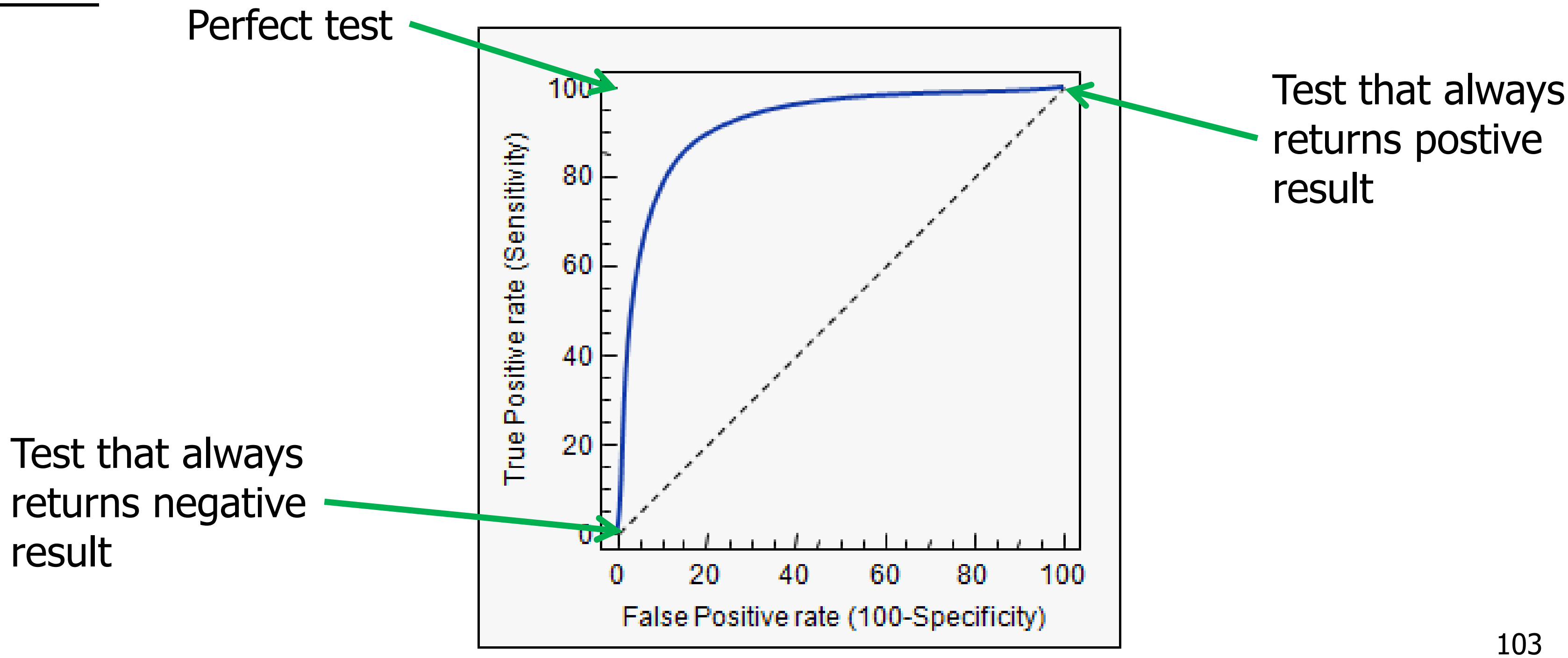
Solution:

What if the sensitivity increases? Does $P[\mathrm{D}|\mathrm{test}{+}]$ increase or decrease?

What if the specificity increases? Does $P[\mathrm{D}|\mathrm{test}{+}]$ increase or decrease?

Tradeoff between sensitivity and specificity

- Usually tests can be adjusted to increase sensitivity or specificity.
- However, if one goes up, the other usually goes down.
- This can be visualized through a curve called the receiver operating characteristic.



Example: Probability

- Suppose that 1% of the population has a disease.
- Suppose you could choose from two tests:
  - Test 1: Sensitivity = 99%, Specificity = 97%
  - Test 2: Sensitivity = 97%, Specificity = 99%
- Which test has the higher P[D|test+], or are they both the same?

Solution:

ELEC 2600H: Probability and Random Processes in Engineering

Part I: Basic Probability Theory

- Lecture 1: Course Introduction
- Lecture 2: Build a Probability Model
- Lecture 3: Conditional Probability & Independence
- Lecture 4: Sequential Experiments

- Out-of-Class Reading: Counting Method

- **Sequential Experiments**
  - Bernoulli trials
  - Binomial probability law
  - Multinomial probability law
  - Geometric probability law
  - Sequences of dependent experiments

- Example: Bean Machine Game!

Sequential Experiments

- Experiments that involve repetitions or multiple participants can often be viewed as a sequence of sub-experiments.
- The sub-experiments can be identical or non-identical, dependent or independent. The individual sample spaces can be identical or non-identical.
- If the experiments are identical, the individual sample spaces are identical but not vice versa.
- Examples:
  - Tossing a coin n times
    - *repetition*
    - *independent*
    - *identical sub-experiments*
    - *identical individual sample spaces*
  - Checking the number of students who are **sleeping in class** now
    - *multiple participants*
    - *independent?*
    - *identical individual sample spaces*
    - *non-identical sub-experiments*
      - Alice and Bob sleep with different probabilities

Sample Spaces Formed by Cartesian Products

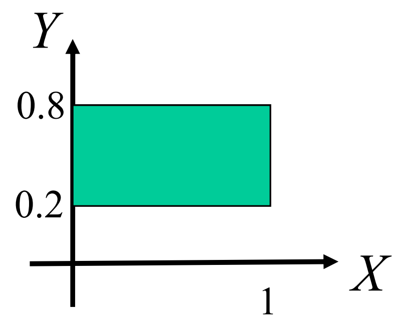
- When an experiment consists of performing sub-experiments $E_1, E_2, \ldots, E_n$. The outcome is an n-tuple $s = (s_1, s_2, \ldots. s_n)$.
- The sample space can be denoted as the Cartesian product of the individual sample spaces.

$$S_1 \times S_2 \times \ldots\ldots \times S_n$$

  - Example: Toss a coin two times. The sample space is

$$\{H,T\} \times \{H,T\} = \{HH, HT, TH, TT\}$$

  - Example: Pick a random number X from $[0,1]$ and a random number Y from $[0.2,0.8]$. The sample space is $[0,1] \times [0.2,0.8]$



Sequences of Independent Sub-Experiments

- In many cases, the sub-experiments can be assumed to be ***independent***, e.g.
  - a sequence of rolls of a die
  - a sequence of coin flips
  - a sequence of selections from a large collection of resistors.
- In this case, we can compute the probability of any event by exploiting independence.
- Let $A_1, A_2, \ldots, A_n$ be events such that A_k concerns only the outcome of the kth sub-experiment. If the sub-experiments are independent, the events A_k are independent and we have:

$$P[A_1 \cap A_2 \cap \ldots\ldots \cap A_n] = P[A_1]P[A_2]\ldots.P[A_n]$$

vice versa ?

Elec 2600H: Lecture 4

- **Sequential Experiments**
  - **Bernoulli trials**
  - Binomial probability law
  - Multinomial probability law
  - Geometric probability law
  - Sequences of dependent experiments

- Example: Bean Machine Game!

Sir Jacob Bernoulli

Bernoulli Trials

- A *Bernoulli trial* is an experiment that is performed once **with two possible outcomes**

SUCCESS

Probability of success: p

FAILURE

Probability of failure: $q = 1 - p$

- A Bernoulli trial is the simplest experiment
  - Binary sample space: {success, failure}, {1, 0}, {head, tail}

- Examples
  - Tossing a coin
  - Scoring above the mean in a test
  - Receive a message in the next second
  - Typhoon signal 8 on Wednesday

Example 2.37

- Suppose that a coin is tossed three times. Assuming the outcomes of the tosses are independent and the probability of heads is always *p*, the probability for each sequence of heads and tails is

$$P[\{\mathrm{HHH}\}] = P[\{\mathrm{H}\}]P[\{\mathrm{H}\}]P[\{\mathrm{H}\}] = p^3$$
$$P[\{\mathrm{HHT}\}] = P[\{\mathrm{H}\}]P[\{\mathrm{H}\}]P[\{\mathrm{T}\}] = p^2(1-p)$$
$$P[\{\mathrm{HTH}\}] = P[\{\mathrm{H}\}]P[\{\mathrm{T}\}]P[\{\mathrm{H}\}] = p^2(1-p)$$
$$P[\{\mathrm{THH}\}] = P[\{\mathrm{T}\}]P[\{\mathrm{H}\}]P[\{\mathrm{H}\}] = p^2(1-p)$$
$$P[\{\mathrm{HTT}\}] = P[\{\mathrm{H}\}]P[\{\mathrm{T}\}]P[\{\mathrm{T}\}] = p(1-p)^2$$
$$P[\{\mathrm{THT}\}] = P[\{\mathrm{T}\}]P[\{\mathrm{H}\}]P[\{\mathrm{T}\}] = p(1-p)^2$$
$$P[\{\mathrm{TTH}\}] = P[\{\mathrm{T}\}]P[\{\mathrm{T}\}]P[\{\mathrm{H}\}] = p(1-p)^2$$
$$P[\{\mathrm{TTT}\}] = P[\{\mathrm{T}\}]P[\{\mathrm{T}\}]P[\{\mathrm{T}\}] = (1-p)^3$$

- Note that the outcomes are equiprobable only if the coin is ***fair***:

$$P[\{H\}] = P[\{T\}] \iff p = 1-p \iff p = ½$$

Example 2.37 (cont)

- Let k be the number of heads in three trials, then

$$P[k=0] = P[\{\text{TTT}\}] = (1-p)^3$$
$$P[k=1] = P[\{\text{HTT,THT,TTH}\}] = 3p(1-p)^2$$
$$P[k=2] = P[\{\text{HHT,HTH,THH}\}] = 3p^2(1-p)$$
$$P[k=3] = P[\{\text{HHH}\}] = p^3$$

Recall from previous page:

$$P[\{\text{HHH}\}] = P[\{\text{H}\}]P[\{\text{H}\}]P[\{\text{H}\}] = p^3$$
$$P[\{\text{HHT}\}] = P[\{\text{H}\}]P[\{\text{H}\}]P[\{\text{T}\}] = p^2(1-p)$$
$$P[\{\text{HTH}\}] = P[\{\text{H}\}]P[\{\text{T}\}]P[\{\text{H}\}] = p^2(1-p)$$
$$P[\{\text{THH}\}] = P[\{\text{T}\}]P[\{\text{H}\}]P[\{\text{H}\}] = p^2(1-p)$$
$$P[\{\text{HTT}\}] = P[\{\text{H}\}]P[\{\text{T}\}]P[\{\text{T}\}] = p(1-p)^2$$
$$P[\{\text{THT}\}] = P[\{\text{T}\}]P[\{\text{H}\}]P[\{\text{T}\}] = p(1-p)^2$$
$$P[\{\text{TTH}\}] = P[\{\text{T}\}]P[\{\text{T}\}]P[\{\text{H}\}] = p(1-p)^2$$
$$P[\{\text{TTT}\}] = P[\{\text{T}\}]P[\{\text{T}\}]P[\{\text{T}\}] = (1-p)^3$$

Elec 2600H: Lecture 4

- **Sequential Experiments**
  - Bernoulli trials
  - **Binomial probability law**
  - Multinomial probability law
  - Geometric probability law
  - Sequences of dependent experiments

- Example: Bean Machine Game!

Previous examples suggest that the probability of getting k successes (or failures) after performing multiple Bernoulli trials follows a set pattern.

What is this pattern for ARBITRARY numbers of trials and arbitrary *k*?

Binomial Probability Law

- If we conduct n *independent Bernoulli trials*, the **number of successes follows a *binomial probability law***: the probability of k successes in n trials is:

$$p_n(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

- Proof:

Each particular sequence of events with k successes (n-k failures) has probability $p^k (1-p)^{n-k}$

For example, the sequence of 2 successes and 3 failures: $T\,F\,T\,F\,F$ (T = success / F = failure)
has probability $p \times (1-p) \times p \times (1-p) \times (1-p) = p^2(1-p)^3$
where T denotes success and F denote failure.

The number of sequences of n trials with k successes is $\binom{n}{k} = \dfrac{n!}{k!(n-k)!} = \dfrac{n(n-1)\ldots(n-k+1)}{k(k-1)\ldots(2)(1)}$

Thus, the probability is $p_n(k) = \binom{n}{k} p^k (1-p)^{n-k}$

See notes on counting methods

Example 2.38

- Suppose that a coin is tossed $n=3$ times. Assuming the outcomes of the tosses are independent and the probability of heads is always p, find the probability of observing k heads.

- Solution:

$$P[k=0] = p_3(0) = \frac{3!}{0!3!} p^0 (1-p)^3 = (1-p)^3$$

$$P[k=1] = p_3(1) = \frac{3!}{1!2!} p^1 (1-p)^2 = 3p(1-p)^2$$

$$P[k=2] = p_3(2) = \frac{3!}{2!1!} p^2 (1-p)^1 = 3p^2(1-p)$$

$$P[k=3] = p_3(3) = \frac{3!}{3!0!} p^3 (1-p)^0 = p^3$$

These are consistent with the results of Example 2.37

Binomial Theorem

- Since the number of successes in n trials can range from 0 to n,

$$\sum_{k=0}^{n} p_n(k) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = 1$$

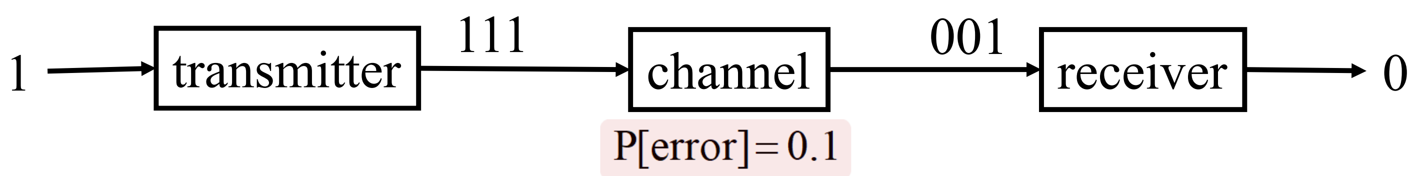
- Proof:
  This can be proven using the binomial theorem: $(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$

  Letting $a = p$ and $b = (1-p)$, $(p+1-p)^n = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k}$

$$1 = \sum_{k=0}^{n} p_n(k)$$

Example 2.40

A communication system transmits binary information over a channel that introduces random bit errors. To reduce the effect of these errors, the transmitter transmits each bit 3 times and the receiver takes a **majority vote** to decide which bit was transmitted.



Find the probability that a bit is received incorrectly.

Solution

A bit is incorrect if two or three errors occur during transmission:

$$\begin{aligned} \mathrm{P}[\text{incorrect}] &= p_3(2) + p_3(3) \\ &= \binom{3}{2}(0.1)^2(0.9) + \binom{3}{3}(0.1)^3 = 0.028 \end{aligned}$$

Elec 2600H: Lecture 4

- **Sequential Experiments**
  - Bernoulli trials
  - Binomial probability law
  - **Multinomial probability law**
  - Geometric probability law
  - Sequences of dependent experiments

- Example: Bean Machine Game!

The Multinomial Probability Law

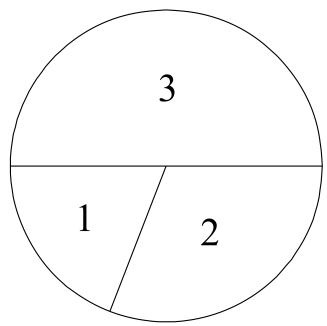
- The binomial probability law can be ***generalized***.
- Suppose that an experiment consisting of n independent repetitions of the sub-experiment are performed.
- Let $B_1, B_2, B_3, \ldots\ldots, B_M$ partition the sample space S of the sub-experiment and let $P[B_j] = p_j$
- Since the events are mutually exclusive, $p_1 + p_2 + \ldots.. + p_M = 1$
- Let k_j be the number of times that event B_j occurs.
- The probability of the vector $(k_1, k_2, k_3, \ldots, k_M)$ that specifies the number of times each event occurs is given by the multinomial probability law:

$$\boxed{P[(k_1, k_2, \ldots\ldots, k_M)] = \frac{n!}{k_1! k_2! \ldots. k_M!} p_1^{k_1} p_2^{k_2} \ldots. p_M^{k_M}}$$

Example 2.41

A dart is thrown nine times at a target consisting of three areas. Each throw has a probability of 0.2, 0.3, and 0.5 of landing in areas 1, 2, and 3. Find the probability that the dart lands three times in each area.

$$P[(3,3,3)] = \frac{9!}{3!3!3!}(0.2)^3(0.3)^3(0.5)^3 = 0.04536$$



$$p_1 = 0.2$$
$$p_2 = 0.3$$
$$p_3 = 0.5$$

Example 2.19

Suppose that we have 12 balls, each of which can fall randomly and independently into one of 12 cells. Note that each cell might contain more than one ball. What is the probability that each cell contains only one ball?

Solution:

Here each ball represents the repetition of an experiment which has 12 equally probable outcomes (the cell). Thus, n = 12 and

$$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8 = p_9 = p_{10} = p_{11} = p_{12} = \frac{1}{12}$$

The outcome that each cell contains only one ball is specified by

$$k_1 = k_2 = k_3 = k_4 = k_5 = k_6 = k_7 = k_8 = k_9 = k_{10} = k_{11} = k_{12} = 1$$

Using the multinomial probability law: $P[(1,1,.....,1)] = \frac{12!}{1!1!....1!}\left(\tfrac{1}{12}\right)^1\left(\tfrac{1}{12}\right)^1 ...\left(\tfrac{1}{12}\right)^1$

$$= \frac{12!}{12^{12}} \approx 5.37 \times 10^{-5}$$

Elec 2600H: Lecture 4

- **Sequential Experiments**
  - Bernoulli trials
  - Binomial probability law
  - Multinomial probability law
  - **Geometric probability law**
  - Sequences of dependent experiments

- Example: Bean Machine Game!

The Geometric Probability Law (i)

- Assume that we conduct an independent sequence of Bernoulli trials.

- Let M be the number of trials **until the first success.**
- The probability that $M=m$ is given by the ***geometric probability law***:

$$\boxed{p_M(m) = P[A_1^c A_2^c \ldots\ldots A_{m-1}^c A_m] = p(1-p)^{m-1} \qquad m = 1,2,\ldots.}$$

  - A_k : Event that kth trial resulted in a success
  - A_k^c : Complement of A_k

$$\underbrace{0\ 0\ 0\ 0\ 0\ 0\ 0\ 1}_{M}\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 0$$

The Geometric Probability Law (ii)

- **But....**

- Let N be the number of trials **before the first success**.
- The probability that $N=n$ is given by

$$p(n) = P[A_1^c A_2^c \ldots\ldots A_n^c A_{n+1}] = p(1-p)^n \qquad n = 0,1,2,\ldots.$$

This is ***also called the geometric probability law***!

- Be careful which version you are dealing with!!

$$\underbrace{\overbrace{0\ 0\ 0\ 0\ 0\ 0\ 0}^{N}\ 1}_{M}\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 0$$

Properties of Geometric Probability Law

- The probabilities sum to one: $\sum_{m=1}^{\infty} p(m) = p\sum_{m=1}^{\infty}(1-p)^{m-1}$

$$= p\sum_{m'=0}^{\infty}\left(1-p\right)^{m'} \qquad \text{change of variables: } m = m'+1$$

$$= \frac{p}{1-(1-p)} = 1$$

- The probability that **more than** K **trials** are required until the first success is

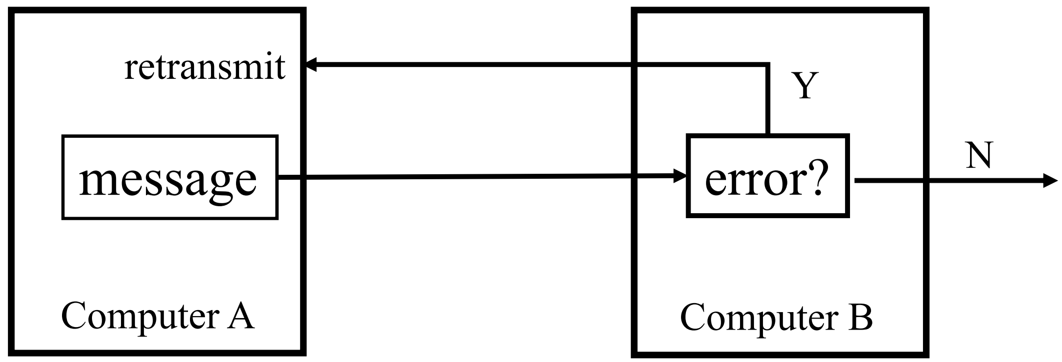
$$P[M>K] = p\sum_{m=K+1}^{\infty}(1-p)^{m-1}$$

$$= p(1-p)^{K}\sum_{m'=0}^{\infty}(1-p)^{m'} \qquad \text{change of variables: } m = m'+\left(K+1\right)$$

$$= \frac{p(1-p)^{K}}{1-(1-p)} = (1-p)^{K}$$

Example

Computer A transmits a message to computer B. If B detects an error in the message, it asks A to retransmit. If the probability of an error is 0.1, what is the probability the message needs to be sent *more than twice*?



Solution:

The probability of a successful transmission is $p = 0.9$.

The probability that more than 2 transmissions are required until a success is

$$P[M > 2] = (1 - p)^2 = 0.1^2 = 0.01$$

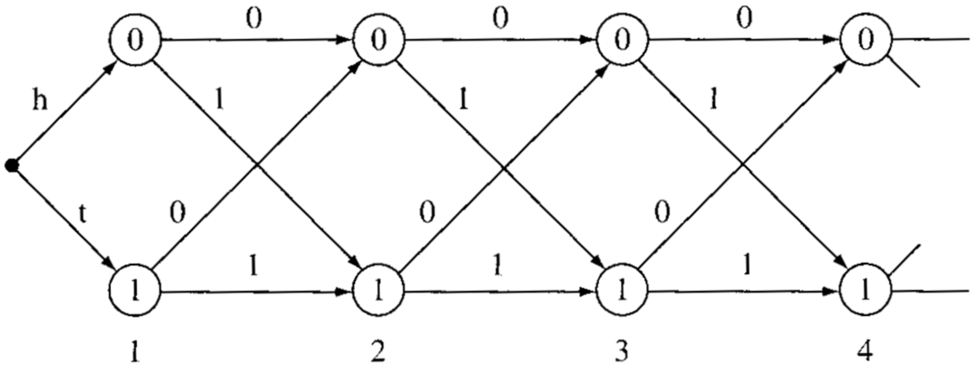
Elec 2600H: Lecture 4

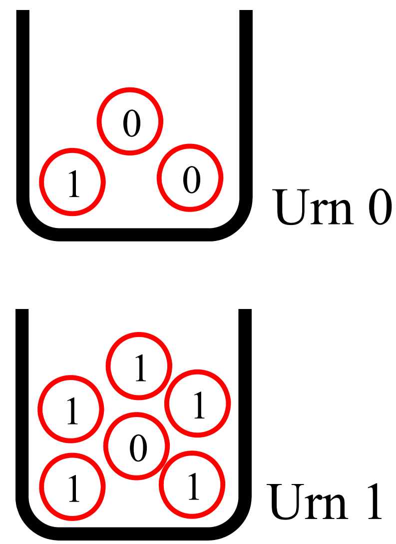
- **Sequential Experiments**
  - Bernoulli trials
  - Binomial probability law
  - Multinomial probability law
  - Geometric probability law
  - **Sequences of dependent experiments**

- Example: Bean Machine Game!

Sequences of Dependent Experiments

- In some cases, the **outcome of a sub-experiment in a sequence may depend upon the outcome of a previous sub-experiment**.
- Example 2.44: We have two urns labeled 0 and 1, each containing balls also labeled 0 and 1. To start, we pick one of the urns at random (flip a coin) and pick a ball from it at random. The label of the ball is used to determine the urn which is picked from in the next sub-experiment, and so on…
- The sample space consists of sequences of 0’s and 1’s.
- Any outcome corresponds to a path through the **“trellis”** diagram shown below:





Computing Probabilities

- With sequences of dependent experiments, we rely heavily upon conditional probabilities.
- Let s_n be the outcome of the nth sub-experiment. Since $P[A \cap B] = P[A|B] \times P[B]$,

$$\begin{aligned} P[s_0 \cap s_1 \cap s_2] &= P[s_2 \mid s_1 \cap s_0] \times P[s_1 \cap s_0] \\ &= P[s_2 \mid s_1 \cap s_0] \times P[s_1 \mid s_0] \times P[s_0] \end{aligned}$$

Always true!

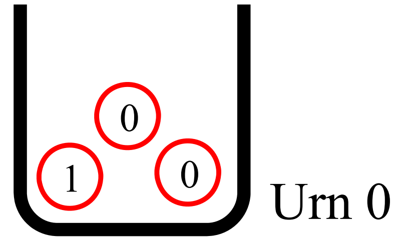
- In this case, the urn we pick from depends **only upon the last sub experiment**. This implies that $P[s_2 \mid s_1 \cap s_0] = P[s_2 \mid s_1]$
- More generally, $P[s_n \mid s_{n-1} \cap \ldots \cap s_0] = P[s_n \mid s_{n-1}]$

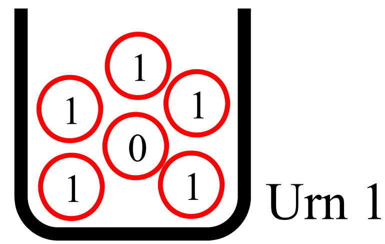
If this is true $\forall n$, we have a Markov Chain

- Thus, $P[s_0 \cap s_1 \cap s_2] = P[s_2 \mid s_1] \times P[s_1 \mid s_0] \times P[s_0]$

Example 2.45

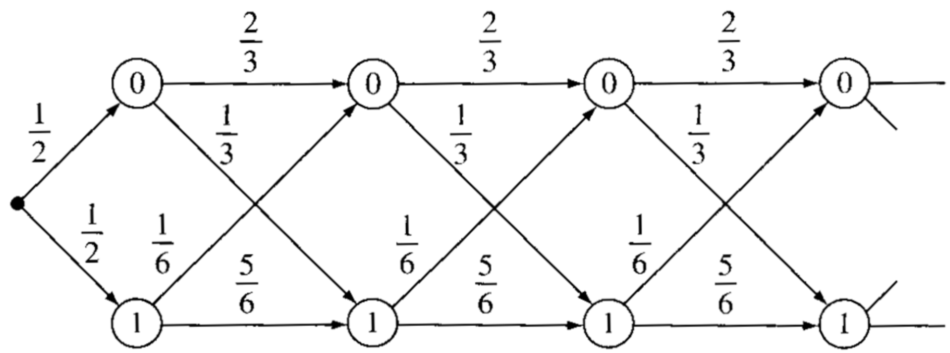
Find the probability of the urn sequence 0011 for the urn experiment introduced in the previous example.





Solution:

We can label the branches of the trellis diagram with the conditional probabilities:

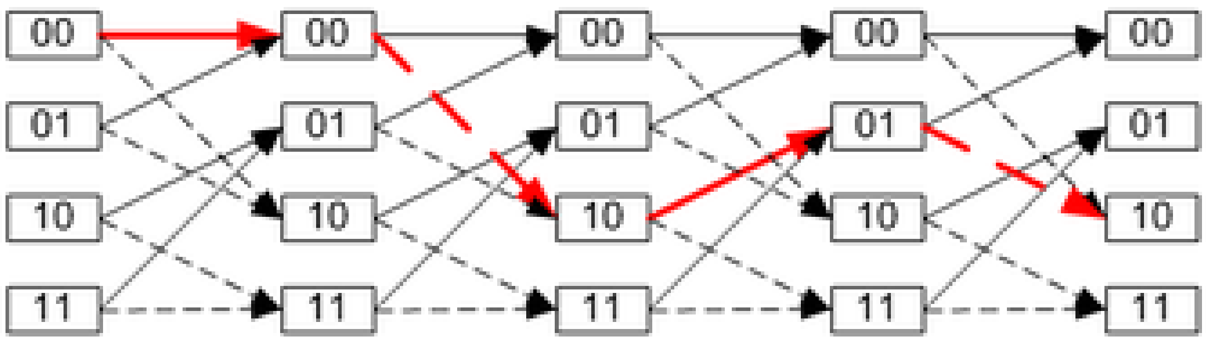


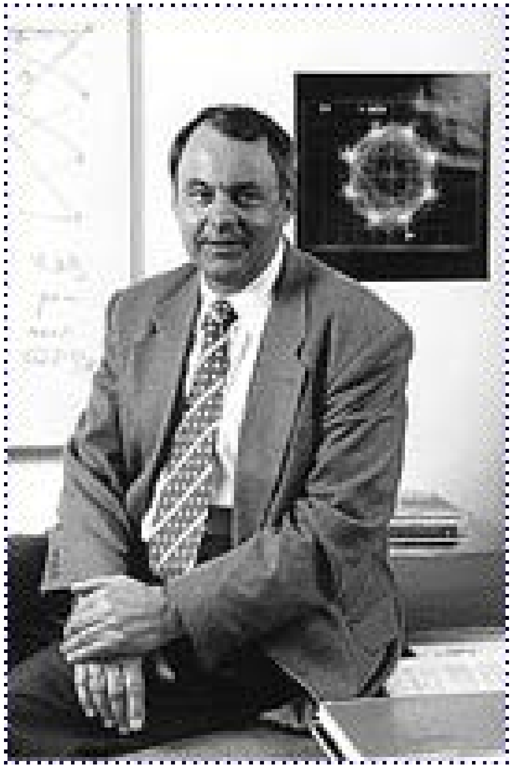
Using the conditional probability, we obtain $P[0011] = P[1\,|\,1] \times P[1\,|\,0] \times P[0\,|\,0] \times P[0]$

$$= \frac{5}{6} \times \frac{1}{3} \times \frac{2}{3} \times \frac{1}{2} = \frac{5}{54}$$

Trellis Coded Modulation

Trellis modulation was invented by Gottfried Ungerboeck working for IBM in the 1970s.





Improvement: Sharing a floppy disk via a BBS could be done in just a few minutes, instead of an hour. (How lucky we are now!)

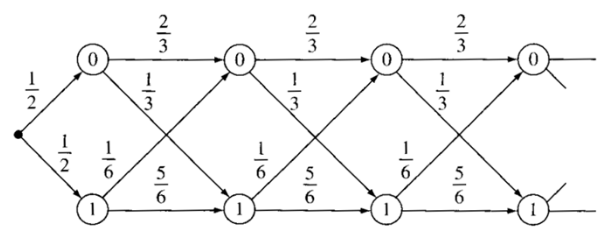
Summary

Bernoulli trials 

Binomial probability law 

*Multinomial probability law 

Geometric probability law 

Sequences of dependent experiments 

Elec 2600H: Lecture 4

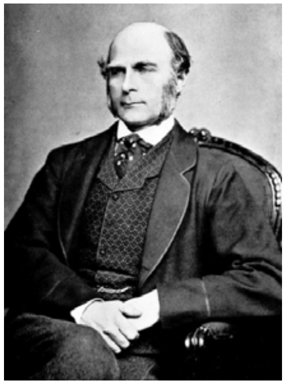
- Sequential Experiments
  - Bernoulli trials
  - Binomial probability law
  - Multinomial probability law
  - Geometric probability law
  - Sequences of dependent experiments

- **Example: Bean Machine Game!**

Bean Machine (Quincunx or Galton Board)

The **bean machine**, also known as the **quincunx** or **Galton board**, is a device invented by Sir **Francis Galton**.

The machine consists of a vertical board with interleaved rows of pins. Balls are dropped from the top, and bounce left and right as they hit the pins.



Sir Francis Galton

Bean Machine (Quincunx or Galton Board)

美國冒險樂園 JUMPIN GYM U.S.A.

For every 10$, you get 6 Balls.

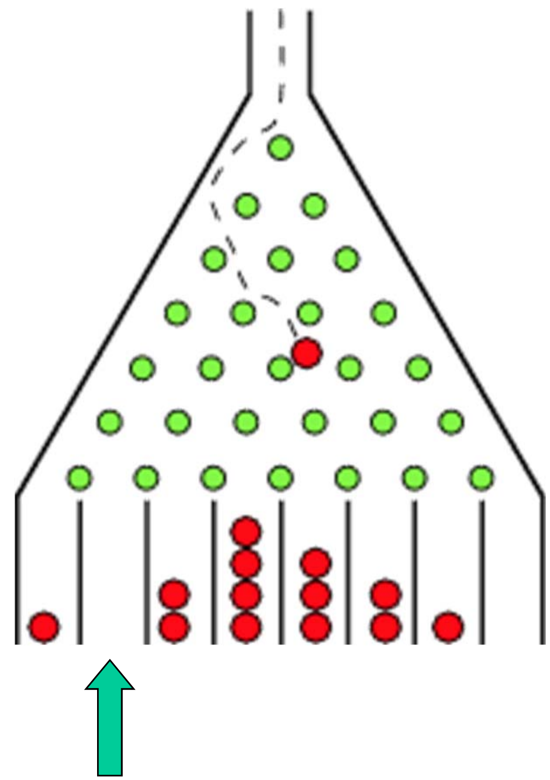
If you can hit the four bins, you win 100$!

What is the chance to win?

Bean Machine (Quincunx or Galton Board)

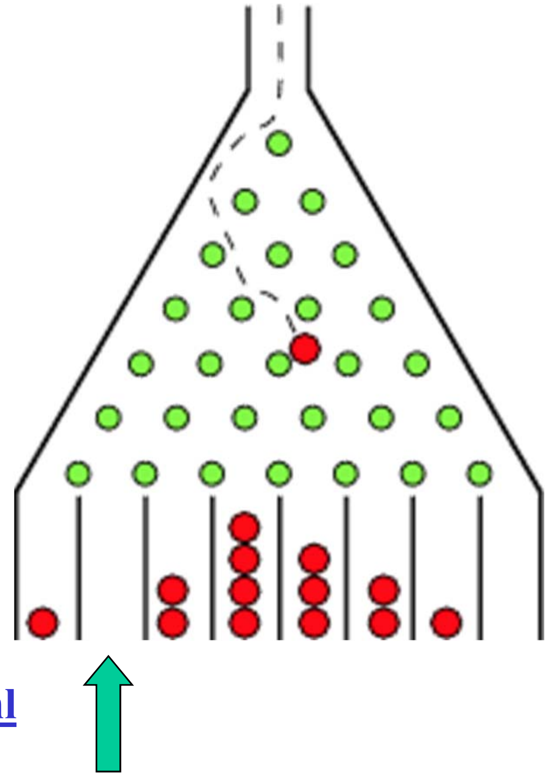
A simpler question: For a Galton board with eight rows of pins, what is the probability that the ball will fall into the bin located second to the left?

Assumption: the ball will bounce to each side with equal probability.



Bean Machine (Quincunx or Galton Board)

- Hints
  - *How many paths are there?*
  - *What is the probability for each path?*
  - *Are different paths independent?*
  - *Are different bounce independent?*



- Other conclusions
  - What are the probabilities for all bins?
  - What is the shape of the distribution curve?

http://www.mathsisfun.com/data/quincunx.html

ELEC 2600H: Probability and Random Processes in Engineering

Part I: Basic Probability Theory

- Lecture 1: Course Introduction
- Lecture 2: Build a Probability Model
- Lecture 3: Conditional Probability & Independence
- Lecture 4: Sequential Experiments

- Out-of-Class Reading: Counting Method

Counting Methods (Self-Reading)

- **Computing Probabilities using Counting Methods**

- Sample Size Computation and Examples

- Probabilities and Poker!

Computing Probabilities Using Counting Methods

- In experiments where the outcomes are equiprobable, we can compute the probability of any event by **counting the number of outcomes** in the event and dividing by the total number of outcomes in the sample space.

$$P[A] = \frac{\text{number of outcomes in A}}{\text{number of outcomes in sample space}}$$

- Thus, $P[A]$ is a **measure** of the size of the set A.
- Here we develop several formulas that are useful for counting problems which are posed as sampling (choosing) problems
  - Balls from an urn
  - Cards from a deck
  - Objects from a population
  - Answers to a multiple-choice question

Number of ordered 2-tuples

- Assume a multiple choice exam with 2 questions.
  - Question 1 has n_1 possible answers: $a_1, a_2, \ldots, a_{n1}$
  - Question 2 has n_2 possible answers: $b_1, b_2, \ldots, b_{n2}$
- Let x_1 represent the answer to question 1, and similarly for x_2.
- Each possible pair is called a 2-tuple, e.g., $\left(x_1, x_2\right) = \left(a_3, b_8\right)$
- The number of possible ways to answer the test is $n_1 n_2$.

$$
\begin{array}{cc|cccc|}
 & & & x_1 & & \\
 & & a_1 & a_2 & \cdots & a_{n_1} \\
\hline
 & b_1 & (a_1,b_1) & (a_2,b_1) & \cdots & (a_{n_1},b_1) \\
 & b_2 & (a_1,b_2) & (a_2,b_2) & \cdots & (a_{n_1},b_2) \\
x_2 & \vdots & \vdots & & \ddots & \vdots \\
 & b_{n_2} & (a_1,b_{n_2}) & (a_2,b_{n_2}) & \cdots & (a_{n_1},b_{n_2}) \\
\hline
\end{array}
$$

Number of ordered k-tuples

- Assume a multiple choice exam with k questions. Question i has n_i possible answers.
- Each possible way of answering the test is called a k-tuple:

$$\left(x_1, x_2, x_3, \ldots, x_k\right)$$

where x_i represents the chosen answer to question i.

- The total number of possible ways to answer the test is

$$\prod_{i=1}^{k} n_i = n_1 n_2 n_3 \ldots n_k$$

Example

- Suppose the test has 5 questions, and each question has 4 possible answers.
- One student's answer can be represented by the 5-tuple:

$$\left(x_1, x_2, x_3, x_4, x_5\right) = (4, 3, 1, 4, 2)$$

i.e., for the first question the student chose answer 4, for the second question the student chose answer 3, and so on...

- Since each question has 4 possible answers, $n_1 = n_2 = n_3 = n_4 = n_5 = 4$

Thus, the total number of possible ways to answer the test is $n_1 n_2 n_3 n_4 n_5 = 4^5$

Sampling (Replacement and Ordering)

- We can choose k objects from a set A that has n members in different ways
- Replacement
  - ***With replacement*:** after selecting an object and noting its identity, the object is *put back* before the next selection.
  - ***Without replacement*:** the object is *not put back* before the next selection.
- Ordering
  - ***With ordering:*** the *order* in which we draw the objects is recorded
  - ***Without ordering*:** only the identity and number of times each object is drawn is important

- Computing Probabilities using Counting Methods

- **Sample Size Computation and Examples**

- Probabilities and Poker!

Size of the Sample Space

- Suppose we choose k objects from a set A that has n members, how many ways can we do this?
  - ***With replacement and with ordering***: $\boxed{\text{number of outcomes} = n^k}$
  - ***Without replacement and with ordering***

$$\boxed{\text{number of outcomes} = \underbrace{n(n-1)(n-2)\ldots(n-k+1)}_{k\ \text{terms}}}$$

  - ***Without replacement and without ordering***

$$\boxed{\text{number of outcomes} = \underbrace{\frac{n(n-1)(n-2)\ldots(n-k+1)}{k(k-1)(k-2)\ldots(2)(1)}}_{k\ \text{terms in numerator and denominator}}}$$

  - The outcomes when choosing *with replacement and without ordering* should **not** be assumed equally probable.

Example (with replacement/with ordering)

- An urn contains five balls numbered 1 to 5, suppose we select two balls from the urn **with replacement**. What is the probability that the first ball has a number larger than the second?

- Solution:

  Figure (a) shows the size of the sample space is $5^2 = 25$.

  10 out of 25 have the first ball larger than the second.

$$P = \frac{10}{25} = 0.4$$

	ball 2				
ball 1	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)
	(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)
	(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)
	(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)
	(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)

(a) Ordered pairs for sampling with replacement.

	(1, 2)	(1, 3)	(1, 4)	(1, 5)
(2, 1)		(2, 3)	(2, 4)	(2, 5)
(3, 1)	(3, 2)		(3, 4)	(3, 5)
(4, 1)	(4, 2)	(4, 3)		(4, 5)
(5, 1)	(5, 2)	(5, 3)	(5, 4)	

(b) Ordered pairs for sampling without replacement.

	(1, 2)	(1, 3)	(1, 4)	(1, 5)
		(2, 3)	(2, 4)	(2, 5)
			(3, 4)	(3, 5)
				(4, 5)

(c) Pairs for sampling without replacement or ordering.

Example 2.16 (without replacement/with ordering)

- An urn contains five balls numbered 1 to 5, suppose we select two balls from the urn **without replacement**. What is the probability that the first ball has a number larger than the second?

- Solution:

  Figure (b) shows the size of the sample space is (5)(4) = 20.

  10 out of 20 have the first ball larger than the second.

  Thus, $P = \dfrac{10}{20} = 0.5$

ball 2

(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)
(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)
(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)
(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)
(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)

(a) Ordered pairs for sampling with replacement.

ball 1

	(1, 2)	(1, 3)	(1, 4)	(1, 5)
(2, 1)		(2, 3)	(2, 4)	(2, 5)
(3, 1)	(3, 2)		(3, 4)	(3, 5)
(4, 1)	(4, 2)	(4, 3)		(4, 5)
(5, 1)	(5, 2)	(5, 3)	(5, 4)	

(b) Ordered pairs for sampling without replacement.

	(1, 2)	(1, 3)	(1, 4)	(1, 5)
		(2, 3)	(2, 4)	(2, 5)
			(3, 4)	(3, 5)
				(4, 5)

(c) Pairs for sampling without replacement or ordering.

Example 2.17

- An urn contains five balls numbered 1 to 5, suppose we select three balls from the urn **with replacement**. What is the probability that the three balls are different?

- **Solution:**

  The number of ways of choosing the three balls is $5^3 = 125$.

  The number of outcomes for which the balls are different are given by the number of ways we can choose three balls *without replacement*: (5)(4)(3) = 60.

  Thus, the probability is $\frac{60}{125} = 0.48$

Permutations of n distinct objects

- Consider sampling without replacement where the number of objects chosen is equal to the number of objects in the urn, i.e. $k = n$.
- Applying the previous formula, the total number of possible orderings (arrangements or *permutations*) is

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

  where we refer to $n! = (n)(n-1)\ldots(2)(1)$ as n **factorial**, the number of possible orderings of n objects.
- For large n, this is expensive to compute and **Stirling's formula** is sometimes used:

$$n! \sim \sqrt{2\pi}\, n^{n+1/2} e^{-n}$$

Example 2.18

- Find the number of permutations of three distinct objects {1,2,3}.

- **Solution**

Using the factorial formula, we obtain (3)(2)(1) = 6 permutations.

These are given by

$$123 \quad 132 \quad 213 \quad 231 \quad 312 \quad 321$$

Example 2.19

- Suppose that 12 balls are placed at random into 12 cells, where more than 1 ball can occupy a cell. What is the probability that **all cells** are occupied?

- **Solution**

There are 12^{12} possible placements of the 12 balls in the 12 cells.

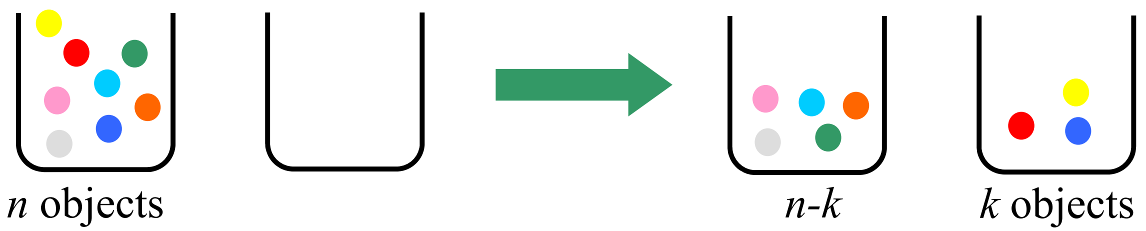
Of these, 12! have one ball in each cell.

$$\text{Thus, the probability is } \frac{12!}{12^{12}} = 5.36 \times 10^{-5}$$

Even if the balls are randomly assigned, it is very unlikely that they will be uniformly distributed.

Sampling without replacement and without ordering

- Suppose we choose k objects from a set A that has n members **without replacement and without ordering**:



- The number of possible outcomes for this experiment is

$$C_k^n = \frac{\text{number of ordered ways to choose } k \text{ objects from } n}{\text{number of ways to order } k \text{ objects}}$$

$$= \frac{n(n\text{-}1)(n\text{-}2)\ldots(n-k+1)}{k!} = \frac{n!}{(n-k)!\,k!} = \binom{n}{k}$$

This is called the **binomial coefficient** and read as "n choose k"

Example 2.20 (without replacement/without order)

- An urn contains five balls numbered 1 to 5, suppose we select two balls from the urn **without replacement and without ordering**. What is the total number of ways we can do this?

- Solution:

  According to the equation on the previous page:

$$\binom{5}{2} = \frac{5!}{(5-2)!\bullet 2!} = \frac{(5)(4)(3)(2)(1)}{(3)(2)(1)\bullet(2)(1)} = \frac{(5)(4)}{(2)(1)} = 10$$

ball 2

ball 1

(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)
(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)
(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)
(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)
(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)

(a) Ordered pairs for sampling with replacement.

	(1, 2)	(1, 3)	(1, 4)	(1, 5)
(2, 1)		(2, 3)	(2, 4)	(2, 5)
(3, 1)	(3, 2)		(3, 4)	(3, 5)
(4, 1)	(4, 2)	(4, 3)		(4, 5)
(5, 1)	(5, 2)	(5, 3)	(5, 4)	

(b) Ordered pairs for sampling without replacement.

	(1, 2)	(1, 3)	(1, 4)	(1, 5)
		(2, 3)	(2, 4)	(2, 5)
			(3, 4)	(3, 5)
				(4, 5)

(c) Pairs for sampling without replacement or ordering.

Example 2.21

- What is the number of distinct permutations of k white balls and n-k black balls (n balls in total)

- **Solution**

  Note that each permutation is uniquely specified by the locations of the white balls.

  Thus, the number of permutations is equal to the number of ways to choose k positions out of n possible positions, C_k^n.

  For example, if there are two white balls and two black balls, the six combinations are:

$$\text{WWBB} \quad \text{WBWB} \quad \text{WBBW} \quad \text{BWWB} \quad \text{BWBW} \quad \text{BBWW}$$

$$C_2^4 = \frac{4!}{(4-2)!2!} = \frac{4\cdot 3\cdot 2\cdot 1}{2\cdot 1\cdot 2\cdot 1} = 6$$

Example 2.22 (modified)

- A batch of 50 items contains 10 defective items. Suppose that 12 items are selected at random and tested. What is the probability that *exactly* 5 of the items tested are defective.

- **Solution**

The number of ways of selecting 12 items from 50 is $\binom{50}{12}$

The number of ways of choosing exactly 5 items is $N_1 N_2$ where

N_1 = # of ways to select 5 defective items from 10 = $\binom{10}{5}$

N_2 = # of ways to select 7 non-defective items from 40 = $\binom{40}{7}$

Thus, the probability is $$\frac{\binom{10}{5}\binom{40}{7}}{\binom{50}{12}} = 0.0387$$

Example 2.22 – More Questions

- Can you determine the probability that:
  - ***At least* 7 selected items are defective?**
  - ***None* of the selected items are defective?**

- Computing Probabilities using Counting Methods
- Sample Size Computation and Examples
- **Probabilities and Poker!**

Poker

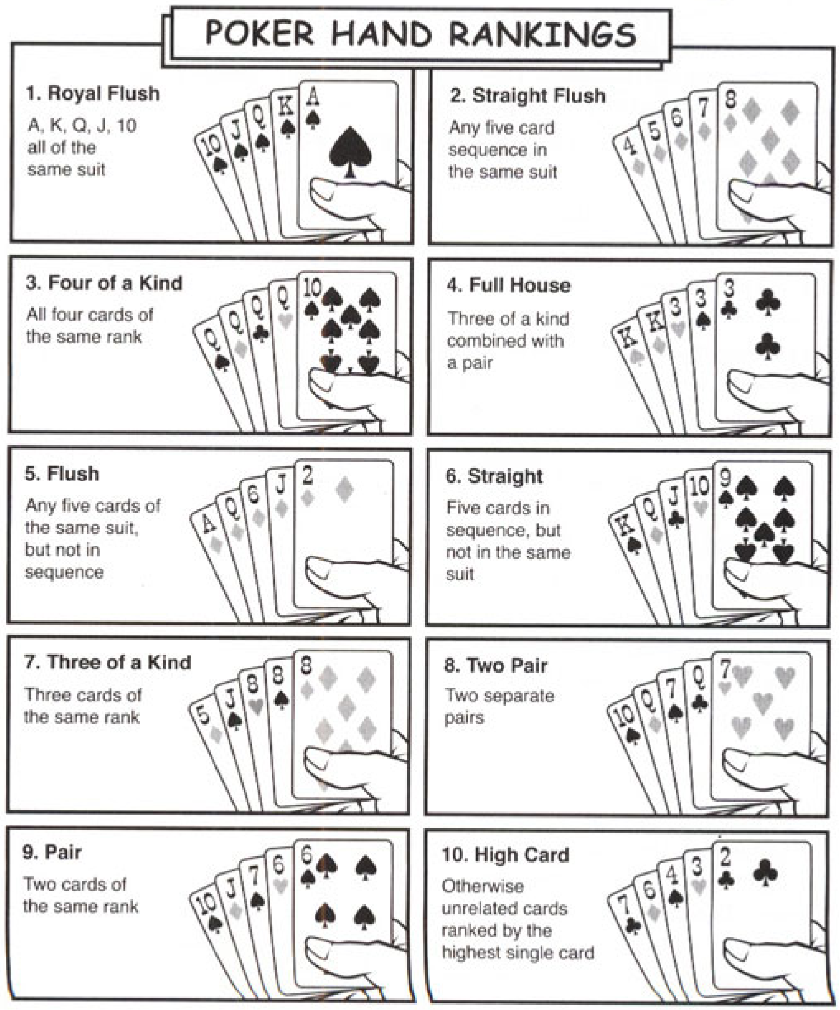
- Played with pack of 52 cards with 4 **suits**

13 hearts	♥ :	2 3 4 5 6 7 8 9 10 J Q K A
13 diamonds	♦ :	2 3 4 5 6 7 8 9 10 J Q K A
13 clubs:	♣ :	2 3 4 5 6 7 8 9 10 J Q K A
13 spades:	♠ :	2 3 4 5 6 7 8 9 10 J Q K A

- Very Basic Rules **(no bets)**:
  - Each player dealt 5 cards
  - Each player looks at cards, and "gives back" to dealer n cards that they don't want
  - Dealer then gives n new cards to the player
  - Player with highest ranking hand wins!

Poker

What is the probability of being dealt each type of poker hand?



Poker: Computing Probabilities

- **Size of sample space:**
  - Number of different possible poker hands
  - **Can you compute this?**

- **Size of event: “A particular type of hand is dealt”**
  - Number of possibilities (possible outcomes) for this specific hand
    - E.g., for a Royal Flush, the size is 4
  - **Can you compute this for:**
    - A straight flush
    - Four of a kind
    - Full house
    - …

Probability of hand being dealt = Size of event / Size of sample space

Poker: Final Note…

- **We are considering here standard 5-card poker**

- **There are different variations, each with different probabilities:**
  - **Texas Hold Em**
  - **7 Card Poker**
  - **Lowball Poker**
  - **….**

- **http://en.wikipedia.org/wiki/Poker_probability**