

# MATH 4720 / MSCS 5720

---

**Instructor: Mehdi Maadooliat**

## **Chapter 4 (Part A)**



**Department of Mathematics, Statistics and Computer Science**

## CHAPTER 4 (PART A)

- **Probability**
- **Random process, Event, Sample space**
- **Set theory review**
- **Conditional Probability and Independence**
- **Bayes' Theorem**
- **Law of total probability**
- **Random Variables**
  - **Discrete**
    - **Binomial**
    - **Poisson**
  - **Continuous**
    - **Normal**
- **Normal Approximation to  $Binomial(n, \pi)$**
- **Sampling Distribution**



# PROBABILITY

## Common misconceptions about probability

- Paulos (1988) tells the story of a weather forecaster on American TV who reported that there was *a 50% chance of rain on Saturday*, and *a 50% chance of rain on Sunday*, from which he concluded that there was *a 100% chance of rain on the weekend*.
- *New Scientist* reported a story about an inspector in the Food and Drug Administration who visited a restaurant in Salt Lake City famous for its quiches made from four fresh eggs. The inspector told the owner that according to FDA research *every fourth egg has salmonella bacteria*, so the restaurant should *only use three eggs* in a quiche.

(Source: [http://www.edge.org/3rd\\_culture/gigerenzer03/gigerenzer\\_print.html](http://www.edge.org/3rd_culture/gigerenzer03/gigerenzer_print.html))

Paulos JA (1988) *Innumeracy: Mathematical illiteracy and its consequences*. Vintage Books, New York

# RANDOMNESS AND PROBABILITY

- We call a **phenomenon random** if individual outcomes are uncertain but there is nonetheless a regular distribution of outcomes in a large number of repetitions
- The **probability** of any outcome of a random phenomenon is the proportion of times the outcome would occur in a very long series of repetitions. If a random experiment is repeated  $n$  times then,

$$\frac{\text{number of times } A \text{ occurs}}{\text{number of trials, } n} \rightarrow P(A), \text{ the probability that event } A \text{ occurs}$$

- Example: Tossing coins

# PROBABILITY DEFINITIONS

- **Random process:** a process whose outcome can not be predicted with certainty
- **Sample space ( $S$ ):** the collection of all possible outcomes to a random process
- **Event ( $A, B$ ):** a collection of possible outcomes
  - **Simple:** one outcome
  - **Compound:** more than one outcome
- **Probability:** a number between 0 and 1 (inclusive) indicating the likelihood that the event will occur

# SET THEORY REVIEW

- **Union** ( $A \cup B$ ) - all elements in  $A$  or  $B$
- **Intersection** ( $A \cap B$ ) - all elements in  $A$  and  $B$
- **Complement** ( $\bar{A}$ ) - all elements not in the set
- **Mutually exclusive** – two events are mutually exclusive if they have no outcomes in common
- **Venn Diagram**
- **The proportion of times that an event  $A$  occurs converges to the probability of  $A$ ,  $P(A)$ , as the number of repetitions becomes large.**

# PROBABILITY WITH EQUALLY LIKELY OUTCOMES

- If a sample space is composed of  $k$  equally likely outcomes with  $m$  outcomes contained in  $A$ , then  $P(A) = m/k$ .
- Two way classification for the salary data:

	Education	Engineering	Total
Female	856	232	1088
Male	220	924	1144
Total	1076	1156	2232

- Randomly select a student from the survey group, what is the probability that
  - the student is female?
  - the student is an engineer?
  - the student is a female engineer?
  - the student is a female or an engineer?

# PROPERTIES OF PROBABILITY

- **Axioms:**

- $P(A) \geq 0$
- $P(S) = 1$
- If  $A$  and  $B$  are mutually exclusive,  $P(A \cup B) = P(A) + P(B)$

- **Properties that follow:**

- $0 \leq P(A) \leq 1$
- $P(\emptyset) = 0$
- $P(\overline{A}) = 1 - P(A)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

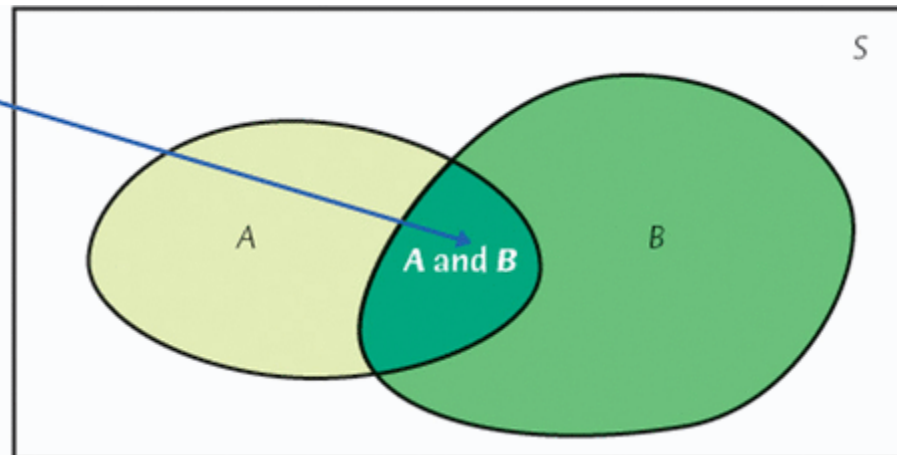


## ADDITION RULE IN GENERAL

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

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

Outcomes here are double-counted by  $P(A) + P(B)$ .



# A RAINY WEEKEND?

- **Recall the TV weather forecaster who announced that if the probability of rain on Saturday is 50% and the probability of rain on Sunday is 50%, then the probability of rain over the weekend is 100%.**
- **If the probability that it rains on both Saturday and Sunday is 35%, what is the probability of rain over the weekend?**

**Events:  $A$  = it rains on Saturday,  $B$  = it rains on Sunday.**

## BACK TO THE SALARY EXAMPLE

- **Consider the following two way classification table between major and gender for the salary data:**

	Education	Engineering	Total
Female	856	232	1088
Male	220	924	1144
Total	1076	1156	2232

- **If we randomly select a student from this group, what is the probability that**
  - **the student is not a male educator?**
  - **the student is a female or an engineer?**

# CONDITIONAL PROBABILITY AND INDEPENDENCE

- The **conditional probability** of  $A$  given  $B$  is

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

if  $P(B) > 0$ .

- $A$  and  $B$  are **independent** if  $P(A|B) = P(A)$ .
- If the events are independent, knowing  $B$  occurs does not change the probability that  $A$  occurs.
- If  $P(A) > 0$  and  $P(B) > 0$ , then  $A$  independent of  $B$  means  $P(A \cap B) = P(A)P(B)$ .
- What are some examples of events that are independent?

## BACK TO THE EXAMPLES

- **For the salary example, is the gender of the student selected independent of the student's major?**

	Education	Engineering	Total
Female	856	232	1088
Male	220	924	1144
Total	1076	1156	2232



# APPLYING THE MULTIPLICATION RULE TO INDEPENDENT EVENTS: THE BAD EGG?

- *New Scientist* reported a story about an inspector who visited a restaurant in Salt Lake City famous for its quiches made from four fresh eggs. The inspector told the owner that according to FDA research *every fourth egg has salmonella bacteria*, so the restaurant should only *use three eggs* in a quiche.

Now, what is the probability

- a. use 1 egg and it is bad
- b. use 2 eggs and at least one is bad
- c. use 3 eggs and at least one is bad
- d. use 4 eggs and at least one is bad

# BAYES' THEOREM

- In some situations, we know the conditional probability  $P(B|A)$  but are much more interested in  $P(A|B)$ .
- For example, diagnostic tests provide  $P(\text{Test} + | \text{disease})$  but we are interested in  $P(\text{disease} | \text{Test} +)$

		True health status	
		Disease	No disease
Test result	Positive	✓	<b>False positive</b>
	Negative	<b>False negative</b>	✓

(a)

- Bayes' Theorem provides a method for finding  $P(A|B)$  from  $P(B|A)$

# BAYES' FORMULA

- If  $A$  and  $B$  are any events whose probabilities are not 0 or 1, then

- $$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\bar{A})P(\bar{A})}$$

- How do we get this?



## POLYGRAPH TEST

- **Some employers use lie detector tests to screen applicants. Lie detector tests are not completely reliable. Suppose that in a lie detector test, 64% of lies are identifies as lies and that 17% of true statements are also identified as lies. A company gives job applicants a polygraph test, asking, "Did you tell the truth on your job application? Suppose that 94% of the job applicants tell the truth during the polygraph test. What is the probability that a person who fails the test was actually telling the truth?**

# LAW OF TOTAL PROBABILITY AND BAYES RULE

- If  $A_1, \dots, A_n$  are mutually exclusive, each event has positive probability and one of the events must occur, then

$$P(B) = \sum_{j=1}^n P(B \mid A_j)P(A_j)$$

- If  $P(B) > 0$ , the above result implies Bayes rule:

$$P(A_k \mid B) = \frac{P(B \mid A_k)P(A_k)}{\sum_{j=1}^n P(B \mid A_j)P(A_j)}$$



## VOLTAGE REGULATOR EXAMPLE

- **In a batch of voltage regulators, 60% came from supplier 1, 30% from supplier 2 and 10% from supplier 3.**
- **95% of regulators from supplier 1 work**
- **60% of regulators from supplier 2 work**
- **50% of regulators from supplier 3 work**
  
- **If a regulator randomly selected from the batch works, what is the probability it came from supplier 1?**

# RANDOM VARIABLES

- **Random Variable:**  **$Y$**  is a random variable if it assumes values randomly
- **Examples:**
  - Toss a coin 2 times.  $Y = \#$  of heads
  - $Y = \#$  of accidents in Wisconsin Ave. per day
  - $Y = \text{time(in minutes) until next accident in Wisconsin Ave.}$
- **A discrete random variable takes on a finite or countable number of values.**

# DISCRETE RANDOM VARIABLE

- We will identify the distribution of a discrete random variable  $Y$  by its **probability distribution function**,  $P(Y = y)$ .
- The probability associated with every value of  $Y$  lies between 0 and 1.
- The sum of the probabilities for all values of  $Y$  is equal to 1.
- The probabilities for a discrete random variable are additive. Hence, the probability that  $Y = 1$  or 2 is equal to  $P(Y = 1) + P(Y = 2)$ .

# BINOMIAL DISTRIBUTION

- **Binomial Experiment:**
  - The experiment consists of  $n$  identical trials.
  - Each trial can result in one of two outcomes ( $S$  or  $F$ )
  - The probability of success,  $P(S)$ , is a constant  $\pi$  for all trials
  - Trials are independent
  - $X$  counts the number of successes observed in  $n$  trials
- The random variable  $X$  is said to have a **Binomial distribution** with parameters  $n$  and  $\pi$ .
- $X \sim \text{Binomial}(n, \pi)$

## BINOMIAL DISTRIBUTION CONT...

- **Let**  $Y \sim \text{Binomial}(n, \pi)$
- **Note that  $Y$  is discrete with possible values  $0, 1, \dots, n$**
- $$P(Y = k) = \frac{n!}{k!(n-k)!} \pi^k (1 - \pi)^{n-k}, \quad k = 0, 1, \dots, n$$
- **Example: Toss a coin two times.**
  - $Y = \# \text{ of heads}$
  - $Y \sim \text{Binomial}(2, 1/2)$

## ANOTHER EXAMPLE

- **Suppose there are 10 multiple choice questions each with 4 multiple choices.**

- Q.1:      A B C D
- Q.2:      A B C D
- .
- .
- .
- Q.10:    A B C D



## MULTIPLE CHOICE EXAMPLE

- **Suppose a student answer each question by random guessing.**
- *$Y = \# \text{ of correct answers.}$*
- **What is the distribution of  $Y$ ?**
  1. **What is the probability that the student gets all correct answers?**
  2. **What is the probability that the students gets at least 7 correct answers?**

## EXAMPLE CONT.

- $Y \sim \text{Binomial}(10, 1/4)$

- $P(Y = k) = \frac{n!}{k!(n-k)!} \pi^k (1 - \pi)^{n-k}$

1. 
$$P(Y = 10) = \frac{10!}{10!(10-10)!} \left(\frac{1}{4}\right)^{10} \left(1 - \frac{1}{4}\right)^{10-10}$$
$$= 9.5 \times 10^{-7}$$

2. 
$$P(Y \geq 7) = P(Y = 7) + P(Y = 8)$$
$$+ P(Y = 9) + P(Y = 10)$$
$$= 0.0035$$

- [Binomial Calculator](#)
- **Minitab or Calculator**

## EXAMPLE CONT.

- **Using Calculator:**
- **Binomialpdf** $(n, \pi, k) = P(Y = k)$
- **Binomialcdf** $(n, \pi, k) = P(Y \leq k)$ 
  - $P(Y = 10) = \text{binomialpdf}(10, 0.25, 10)$   
 $= 9.5 \times 10^{-7}$
  - $P(Y \geq 7) = 1 - P(Y \leq 6)$   
 $= 1 - \text{binomialcdf}(10, 0.25, 6)$   
 $= 0.0035$
- **This implies that, if you only make a random guess, it is highly unlikely that you will get all 10 answers correct, or even 7 or more answers correct.**

## EXAMPLE

- Genetic theory suggest that 1 in 1000 adult population have a particular genetic disorder. What is the probability that in a sample of 25 adults at least one adult have the genetic disorder.
  - $Y \sim \text{Binomial}(25, 0.001)$
  - $P(Y \geq 1) = 1 - P(Y = 0)$   
 $= 1 - \text{binomialpdf}(25, 0.001, 0)$   
 $= 1 - 0.975 = 0.025$
- This implies that it is **NOT** likely that in a sample of 25 adults, at least one has genetic disorder.
- [Binomial Calculator](#)

## NURSE EMPLOYMENT CASE

- **Contract requires 90% of records handled timely**
- **32 of 36 sample records handled timely, she was fired!**
- **If the proportion of all records handled timely is 0.9, what is the probability that 32 or fewer would be handled timely in a sample of 36?**
  - $Y \sim \text{Binomial}(36, 0.90)$
  - $P(Y \leq 32) = \text{binomialcdf}(36, 0.90, 32)$
  - $= 0.4915$
- **[Binomial Calculator](#)**

# POISSON DISTRIBUTION

- Let  $Y$  be the number of accidents at a particular intersection during a year.
- $Y$  is discrete with possible values  $0, 1, 2, \dots$ , (we cannot assign a maximum value).
- Binomial distribution does not make sense in this case.
- For such situation we use *Poisson* distribution.

# POISSON DISTRIBUTION

- Let  $Y$  be the number of occurrence of an event during a time period (or in a given region), then under certain condition, the distribution of  $Y$  can be described as Poisson.
- Notation:  $Poisson(\mu)$ .
- $\mu$  — average number of events during time period (or in a given region)
- $$P(Y = y) = \frac{\mu^y e^{-\mu}}{y!}, \quad y = 0, 1, 2, \dots$$
- [Poisson Calculator](#)

## EXAMPLE 1

- **Suppose the average number of accidents at a particular intersection is 20 per year. What is the probability that during next year more than 12 accidents would occur?**

- $Y \sim \text{Poisson}(\mu = 20)$

$$P(Y > 12) = 1 - P(Y \leq 12)$$

$$= 1 - \text{Poissoncdf}(0, 12, 20)$$

$$= 1 - 0.0390 = 0.961$$

- **What is the probability that 60 or more accidents would occur during the next 5 years?**

- $Y \sim \text{Poisson}(\mu = 20 * 5) = \text{Poisson}(\mu = 100)$

$$P(Y \geq 60) = 1 - P(Y \leq 59)$$

$$= 1 - \text{Poissoncdf}(0, 59, 100)$$

$$\approx 1.000$$



## EXAMPLE 2

- It is generally observed that the average number of infected trees per acre in a forest is 0.1. It was observed that a 100 acre forest has 17 infected trees. What is the probability that 17 or more trees are infected in the forest if the average per acre is 0.1?

- $Y \sim \text{Poisson}(\mu = 0.1 * 100) = \text{Poisson}(\mu = 10)$
- $$\begin{aligned} P(Y \geq 17) &= 1 - P(Y \leq 16) \\ &= 1 - \text{Poissoncdf}(0, 16, 10) \\ &= 1 - 0.9730 = 0.0270 \end{aligned}$$

- [Poisson Calculator](#)



# POISSON APPROXIMATION OF THE *Binomial*( $n, \pi$ ) DISTRIBUTION

- **Let**  $Y \sim \text{Binomial}(n, \pi)$ .

**If  $n$  is very large, and  $\pi$  is very small, then**

$$Y \approx \text{Poisson}(\mu = n\pi)$$

- **Example: Suppose the prevalence of a disease is 1 in 10,000 among adults. What is the probability that a city with 100,000 adults has 15 or more adults carrying the disease?**

$$Y \approx \text{Poisson}\left(\mu = 100000 * \frac{1}{10000}\right) = \text{Poisson}(\mu = 10)$$

$$\begin{aligned} P(Y \geq 15) &= 1 - \text{Poissoncdf}(0, 14, 10) \\ &= 1 - 0.9165 = 0.0835 \end{aligned}$$

- **[Poisson Calculator](#)**