TE 353 Random Variables and Stochastic Processes

Lecture 1: Basics of probability

K. O.-B. Obour Agyekum

Introduction to Probability

- Probability is a mathematical model to help us study physical systems in an average sense. It provides us with mathematical tools that allow us to indicate how likely a random event is going to occur.
- We quantify the random event by attaching a number to it.
- Probabilistic models are used in real world because we do not know, cannot calculate, or cannot measure all the forces contributing to an effect. The forces may be too complicated, too numerous, or too faint.
- Probability cannot be used in any meaningful sense to answer questions such as
  - “What is the probability that there is life on other planets?”
  - “What is the probability that a comet will strike the earth tomorrow?”
- While we cannot predict the exact result of a random event, we can use probability theory to say that certain results are likely and other results are unlikely

The Different Kinds of Probability

Probability as Intuition

- This kind of probability deals with judgments based on intuition.
- It is the general way of defining probability. Common term used is "probably"
- As an example, people buying lottery tickets intuitively believe that certain number combinations like month/day/year of their birthday are more likely to win.

Probability as the Ratio of Favorable to Total Outcomes (Classical Theory)

- In this approach, which is not experimental, the probability of an event is computed *a priori* by counting the number of ways N_E an event E can occur given a total number of all possible outcomes N. Thus, forming the ratio N_E/N
- An important notion here is that all outcomes are equally likely
- Example, suppose we throw a pair of unbiased dice and ask what is the probability of getting a seven? 1/6
- *A priori* means relating to reasoning from self-evident propositions or presupposed by experience. *A posteriori* means relating to reasoning from observed facts

The Different Kinds of Probability

The classical theory suffers from at least two significant problems:

1. It cannot deal with outcomes that are not equally likely.
2. It cannot handle uncountably infinite outcomes with ambiguity

Probability as a Measure of Frequency of Occurrence

- The relative-frequency approach to defining probability of an event E is to perform an experiment n times. Then it is tempting to define the probability of E occurring by

$$P[E] = \lim_{n \to \infty} \frac{n_E}{n}$$

- Quite clearly since $n_E \leq n$ we must have $0 \leq P[E] \leq 1$. One difficulty with this approach is that we can never perform the experiment an infinite number of times so we can only estimate $P[E]$ from a finite number of trials.
- Secondly, we postulate that $\frac{n_E}{n}$ approaches a limit as n goes to infinity.

The Different Kinds of Probability

Probability based on an Axiomatic Theory

- This is the approach followed in recent times.
- To develop it we must introduce certain ideas, especially those of a random experiment/trial, a sample description space, and an event.
- If an experiment is repeated under the same conditions, any number of times, it does not give unique results but may result in any one of the several possible outcomes. Thus, an action or operation which can produce any result or outcome is called a random experiment or a trial.
- The outcomes of the experiment are known as events

Example

i. Rolling of a die is a trial and getting a 6 is an event.

ii. Tossing a coin is a trial and getting head (H) or tail (T) is an event.

iii. Drawing a ball from a basket containing black and white balls is a trial and getting a black or a white ball is an event

The Different Kinds of Probability

- An event whose occurrence is inevitable when an experiment is performed is called a certain or sure event. An event which can never occur when an experiment is performed is called an impossible event
- An event is called simple if it corresponds to a single possible outcome of the experiment otherwise it is called compound event or composite event

Example

In rolling a single die, the chance of getting 5 is a simple event and the chance of getting even numbers (2,4,6) is a compound event. Occurrence of an 8 is an impossible event

Think about this…

Consider a defendant in a murder trial who pleads not guilty to murdering his wife. The defendant has on numerous occasions beaten his wife.

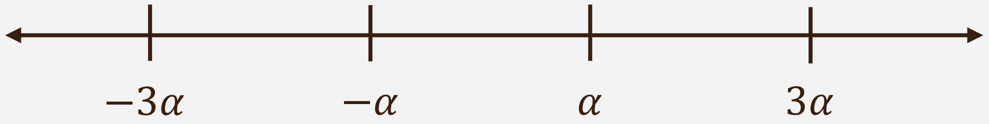
His lawyer argues that, yes, the defendant has beaten his wife but that among men who do so, the probability that one of them will actually murder his wife is only 0.001.

Let us assume that this statement is true. It is meant to sway the jury by implying that the fact of beating one's wife is no indicator of murdering one's wife.

Given that a battered wife is murdered, what is the probability that the husband is the murderer?

Sets and Events

- A set is a collection of objects, called elements, either concrete or abstract. An example of a concrete set is the set of all Telecom 3 students whose height equals or exceeds 6 feet.
- The number of elements in a set is its cardinality and is denoted $|A|$.
- A subset of a set is a collection that is contained within the larger set. Thus, the set of all Telecom 3 students whose height is between 6 and 6.5 feet is a subset of the previous set. In the study of probability, we are particularly interested in the set of all outcomes of a random experiment and subsets of this set
- We therefore denote the experiment as outcomes (random if they cannot be exactly predicted)
- Sample space – the set of all possible outcomes of an experiment
  - Considering an $M-ary$ [Pulse Amplitude Modulation]() (PAM) where $M=4$, the sample space $S=\{-3\alpha,-\alpha,\alpha,3\alpha\}$



  - PAM is the selection from one of M possible amplitude levels to convey information. The selection is done randomly.
  - Each element in the sample space is known as the sample point. Since any set is a subset of itself, S is itself an event. In particular, S is called the certain event.
  - Thus, S is used to denote two objects: the set of all elementary outcomes of a random event and the certain event.

Sets and Events: Examples of Sample Spaces

Example 1.1

The experiment consists of flipping a coin once. Then $S = \{H, T\}$

Example 1.2

The experiment consists of choosing a person at random and counting the hairs on his or her head. Then $S = \{0, 1, 2, \dots, 10^7\}$, that is, the set of all nonnegative integers up to 10^7, it being assumed that no human head has more than 10^7 hairs

Example 1.3

The experiment consists of determining the age to the nearest year of each member of a married couple chosen at random. Then with x denoting the age of the man and y denoting the age of the woman, $S = \{$2 tuples (x, y): x any integer in 10 – 200; y any integer in 10 – 200$\}$

We have assumed that no human lives beyond 200 years and that no married person is ever less than ten years old. Similarly, we have assumed that the coin never lands on the edge. These are certain events.

Event – any subset of the sample space

- Example $A = \{-3\alpha, \alpha\}$; $A \subset S$, from PAM example

Complement – sample points in S but not in event A; denoted $\bar{A}$ or A^c

- $\bar{A} = \{-\alpha, 3\alpha\}$

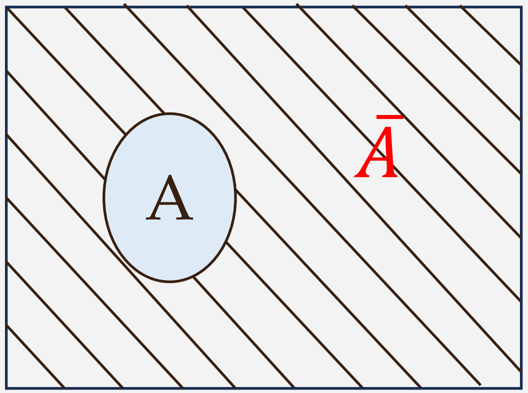
Sets and Events

- Union – sample points either in A or in B, denoted as $A \cup B$ or $A + B$
  - Example: If $A = \{1, 3, 5\}$ and $B = \{3, 6\}$, then $A \cup B = \{1, 3, 5, 6\}$.
  - The operation is commutative and associative: $A \cup B = B \cup A$; $(A \cup B) \cup C = A \cup (B \cup C)$
  - If $B \subseteq A$, then $B \cup A = A$ so that $A \cup A = A$; $A \cup \{\emptyset\} = A$; $A \cup S = S$

- Intersection – sample points in both A and B, denoted as $A \cap B$ or AB
  - From the example above, $AB = \{3\}$
  - Intersection is commutative, associative and distributive
  - $AB = BA$; $(AB)C = A(BC)$; $A(B \cup C) = AB \cup AC$
  - If $A \subseteq B$, then $AB = A$ and $AA = A$; $\{\emptyset\}A = \{\emptyset\}$; $AS = A$

- Null event, $\emptyset$ – does not contain any sample point. Two events A, B are mutually exclusive if $A \cap B = \emptyset$

- Mutually exclusive means the two events cannot happen simultaneously in a single trial



S

$A \cup \bar{A} = S$

$A \cap \bar{A} = \emptyset$

$\bar{\bar{A}} = A$

Axioms (fundamental rules) of probability

- The probability measure of any set A, satisfies
  - Axiom 1: probability measure is non-negative. That is $P(A) \geq 0$

From $S = A \cup \bar{A}$ and $A \cap \bar{A} = \emptyset$

$$1 = P(S) = P(A \cup \bar{A}) = P(A) + P(\bar{A})$$

$$\Rightarrow P(A) = 1 - P(\bar{A})$$

For any event A, $0 \leq P(A) \leq 1$

  - Axiom 2: probability measure of the sample space is unity. That is $P(S) = 1$

$S \cup \emptyset = S$ and $S \cap \emptyset = \emptyset$

$$1 = P(S) = P(S \cup \emptyset) = P(S) + P(\emptyset)$$

$$\Rightarrow P(\emptyset) = 0$$

This is known as an impossible event

Axioms of probability

Axiom 3: if two events are mutually exclusive, then the probability measure of the union is the sum of the probability of each individual event. That is $P(A \cup B) = P(A) + P(B)$ if $A \cap B = \emptyset$

- $P(A) \leq P(B)$ if $A \subset B$

- If A, B are two events that are not mutually exclusive, then
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Joint Probability

- Assume that we perform the following experiment: We wish to collect call data for MTN at Ayeduase. In particular we are interested in three events; A, B, C, where,

A is the event that on any particular day, the number of calls equals or exceeds 500;

B is the event that on any particular day, the number of calls from MTN to Telecel equals or exceeds 120;

C is the event that on any particular day A and B both occur, that is, $\mathrm{C} \triangleq AB$

- Since C is an event, $\mathrm{P[C] = P[AB]}$
- We call $\mathrm{P[AB]}$ the joint probability of the events A and B
- This notion can obviously be extended to more than two events, that is, $\mathrm{P[EFG]}$ is the joint probability of events E, F and G.
- Now let $\mathrm{n_i}$ denote the number of days on which event i occurred. Over a 1000-day period ($\mathrm{n} = 1000$), the following observations are made: $\mathrm{n_A} = 811$, $\mathrm{n_B} = 306$, $\mathrm{n}_{AB} = 283$. By the relative frequency interpretation of probability

$$\mathrm{P[A]} = \frac{\mathrm{n_A}}{\mathrm{n}} = \frac{811}{1000} = 0.811; \quad \mathrm{P[B]} = \frac{\mathrm{n_B}}{\mathrm{n}} = \frac{306}{1000} = 0.306; \quad \mathrm{P[AB]} = \frac{\mathrm{n_{AB}}}{\mathrm{n}} = \frac{283}{1000} = 0.283$$

Conditional Probability

- Consider now the ratio n_{AB}/n_A. This is the relative frequency with which event AB occurs when event A has occurred.
- Put into words, it is the fraction of time that the number of calls from MTN to Telecel equals or exceeds 120 on those days when the number of calls made equals or exceeds 500. Thus, we are dealing with the frequency of an event, given that or *conditioned upon the fact that another event has occurred*.
- Note that

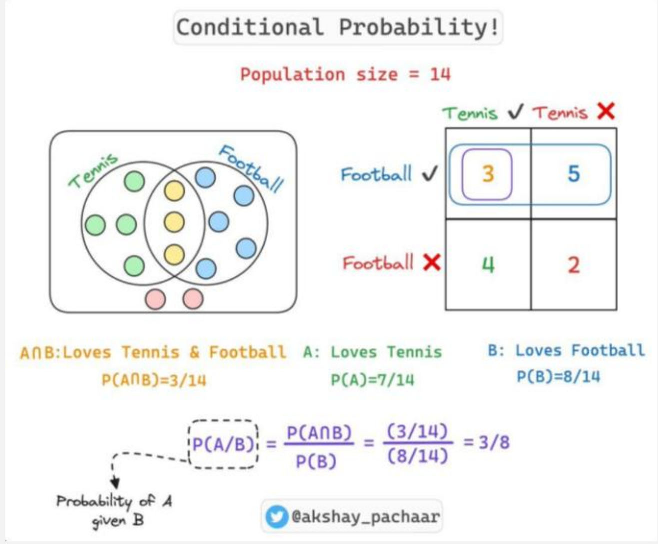
$$\frac{n_{AB}}{n_A} = \frac{n_{AB}/n}{n_A/n} \simeq \frac{P(AB)}{P(A)}$$

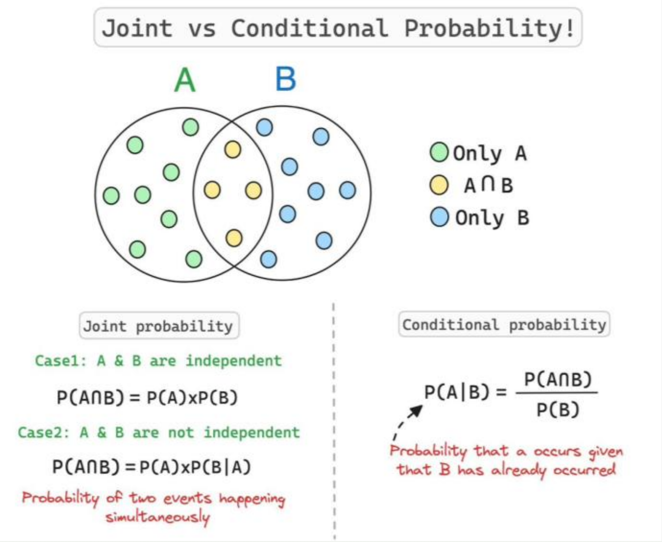
- This concept suggests that we introduce in our theory a conditional probability measure $P(B|A)$ defined by

$$P(B|A) \triangleq \frac{P(AB)}{P(A)}, \qquad P(A) > 0$$

- Similarly,

$$P(A|B) \triangleq \frac{P(AB)}{P(B)}, \qquad P(B) > 0$$





Independent Events

- Two events A, B are statistically independent if $P(A|B) = P(A)$ or $P(B|A) = P(B)$

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

Example 2

Let $S = \{-3\alpha, -\alpha, \alpha, 3\alpha\}$, A $= \{-3\alpha, \alpha\}$ and B $= \{\alpha\}$

Are A and B independent, given that $\mathrm{P}(-3\alpha) = P(-\alpha) = \frac{1}{8}, P(\alpha) = \frac{1}{4}, P(3\alpha) = \frac{1}{2}$?

Example 3

If A and B are independent, show that A and $\bar{B}$ are also independent.

- In wireless communication systems, the symbols generated at different times are frequently independent.

$$\boxed{\text{Information source (S)}} \longrightarrow \begin{array}{c} x_1, x_2, \dots, x_n \\ \text{symbols} \end{array}$$

- Each x_i is generated, and is *independent identically distributed (IID)* from S

Independent Events

Example 4: Block Transmission Symbols

Let $S = \{-3\alpha, -\alpha, \alpha, 3\alpha\}$, $\mathrm{P}(-3\alpha) = P(-\alpha) = \frac{1}{8}, P(\alpha) = \frac{1}{4}, P(3\alpha) = \frac{1}{2}$.

Consider a block of 10 symbols.

a. What is the probability that all symbols in the block belong to $\mathrm{A} = \{-3\alpha, \alpha\}$?

Ans: 5.499×10^{-5}

b. What is the probability that none of the symbols in the block belong to $\mathrm{A} = \{-3\alpha, \alpha\}$?

Ans: 9.095×10^{-3}

c. What is the probability that at least one of the symbols in the block belongs to $\mathrm{A} = \{-3\alpha, \alpha\}$?

Ans: 0.991

Independent Events

Example 5a

The probability that a company director will travel by train is 1/5 and by plane is 2/3. What is the probability of his traveling by train or plane?

Ans: $\frac{13}{15}$

Example 5b

If we draw a card from a well-shuffled deck of cards, what is the probability that the card is either an ace or a king?

Ans: $\frac{2}{13}$

Example 5c

If we draw a card from a well-shuffled deck of cards, what is the probability that the card is either a spade or an ace?

Ans: $\frac{4}{13}$

Independent Events

Example 6: Multiantenna Fading

- Deep Fade is a significant drop in the received signal power in a fading wireless channel.
- Deep Fade occurs when $a < \frac{1}{\sqrt{SNR}}$, where a is the amplitude of the fading channel coefficient and SNR is the signal to noise ratio
- One way to reduce probability of deep fade is to use multiple antennas

Consider a single-input, multiple-output (SIMO) system

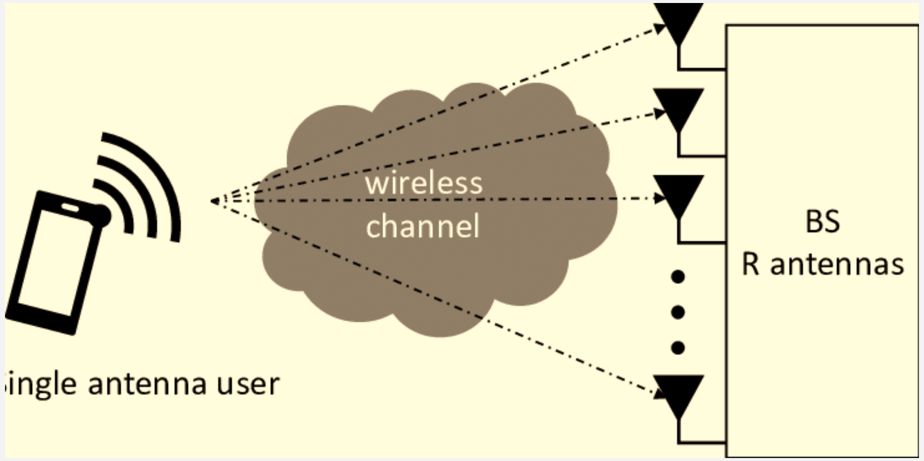
Let E_i denote the event that link i is in Deep Fade

$$P(E_i) = \frac{1}{SNR}$$

Probability that the entire system is in Deep Fade is given as

$$P(E_1 \cap E_2 \cap E_3 \cap \dots \cap E_L) = P(E_1) \times P(E_2) \times \dots \times P(E_L) = \left(\frac{1}{SNR^L}\right)$$

- Having larger number of antennas increases communication efficiency.
- $\frac{1}{SNR^L}$ is known as diversity [Diversity-Wireless Communication] (a technique used to decrease the effect of fading), and diversity decreases the BER



$$y = hx + n$$
$$h = ae^{j\phi}$$

Received power = $P \cdot |h|^2 = a^2 P$, where P is the power of the transmit symbol x

Noise power = σ^2

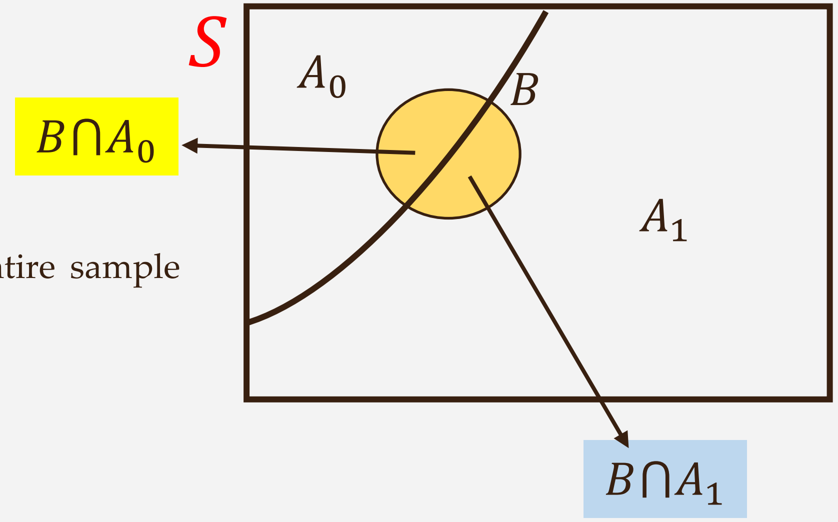
Deep Fade = received power <<< noise power

$$a^2 P < \sigma^2$$

$$a < \sqrt{\frac{\sigma^2}{P}} < \frac{1}{\sqrt{SNR}}$$

Bayes' Theorem

- Consider the events A_0, A_1, B such that
  - $A_0, A_1 \in S$
  - $A_0 \cup A_1 = S$
  - $A_0 \cap A_1 = \emptyset$, A_0, A_1 are mutually exclusive and exhaustive (union spans the entire sample space)
  - $B \cap A_0, B \cap A_1$ are mutually exclusive
  - $(B \cap A_0) \cup (B \cap A_1) = B$



- From the conditional probability principle,

$$P(B) = P(B \cap A_0) + P(B \cap A_1)$$

$$= P(B|A_0)P(A_0) + P(B|A_1)P(A_1)$$

- Similarly,

$$P(A_0|B) = \frac{P(B|A_0)P(A_0)}{P(B|A_0)P(A_0) + P(B|A_1)P(A_1)}$$

$$P(A_1|B) = \frac{P(B|A_1)P(A_1)}{P(B|A_0)P(A_0) + P(B|A_1)P(A_1)}$$

$$P(A_0|B) + P(A_1|B) = 1$$

- $P(A_0)$ and $P(A_1)$ are known as prior probabilities because they exist before we gain any information from the experiment itself
- $P(A_0|B)$ and $P(A_1|B)$ are known as Aposteriori probabilities because they are determined after the results of the experiment are known
- $P(B|A_0)$ and $P(B|A_1)$ are known as likelihoods
- Bayes Theorem is merely an application of the definition of conditional probability

Application of Bayes' Theorem: Maximum Aposteriori Probability (MAP)

- A MAP is a technique in digital/wireless communication channels that can be used to obtain a point estimate of an unobserved quantity on the basis of empirical data
- Consider a digital communication system with binary information symbols where $S = \{0,1\}$
- Let $A_0 = \{0\}$ and $A_1 = \{1\}$. We can clearly see that $A_0 \cap A_1 = \emptyset$

Assume a Binary Symmetric Channel, a simple channel that has a binary input and binary output

$P(A_0) = P(A_1)$ are the prior probabilities

$P(B|A_1) = P(\tilde{B}|A_0) = 0.2$ and $P(B|A_0) = P(\tilde{B}|A_1) = 0.8$ are the likelihoods

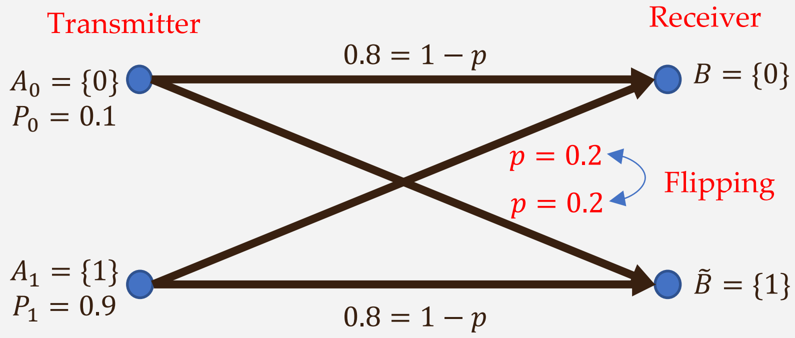
The objective is to find the aposteriori probability, $P(A_0|B)$. That is, given $B = \{0\}$ has been received, what is the probability that $A_0 = \{0\}$ has been transmitted?

Ans: 0.308

Similarly, $P(A_1|B) = 0.692$

Based on this, we can conclude that the actual transmitted symbol must have been a 1 even though a 0 was observed.

MAP receiver minimizes the probability of error at the receiver



- Since $P(B|A_0) = 0.8 > 0.2 = P(B|A_1)$, the maximum likelihood (ML) receiver decides a 0 was transmitted
- It does not minimize the probability of error.

THANK YOU
Kwame Nkrumah University of Science and Technology, Kumasi | Leaders In Change
Visit us at www.knust.edu.gh
uro@knust.edu.gh
Follow KNUST on: