CARNEGIE MELLON UNIVERSITY

Department of Electrical and Computer Engineering

18-751: Applied Stochastic Processes

Instructor: Prof. Ozan Tonguz

Lecture 2

I. Review of Probability Theory

In the previous lecture, the following concepts and notations are introduced:

- Probability Space S: The set has all the outcomes of an experiment; hence, $\mathbf{P}[S] = 1$.
- Empty set $\emptyset$: The set whose event has a zero probablility for occurring; hence, $\mathbf{P}[\emptyset] = 0$
- Event: A set with one or more outcomes from an experiment. In this class, a probablility value assigned to an outcome or an event is based on the relative frequency interpretation $-$ a probability value for an event or an outcome is assumed to be the ratio of the number of occurrences of the event or the outcome and the total number of the experiments conducted. (Later in this class, this relative-frequency assignment will be shown to be accurate as the number of the experiments is large. This is called Law of Large Number (LLN).)
- The probability of Event A *and* Event B occurring *together* is denoted as $\mathbf{P}[AB]$. Based on Bayes Rule, $\mathbf{P}[AB] = \mathbf{P}[A/B]\mathbf{P}[B] = \mathbf{P}[B/A]\mathbf{P}[A]$.
  where $\mathbf{P}[A/B]$ is the probability of the occurrence of Event A given that Event B has occurred. This is called a conditional probability, and usually its effect is to change the sample space from S to Event B.
- The probability of the occurrence of *either* Event A *or* Event B is denoted as $\mathbf{P}[A \cup B]$. This probability can be written as:

$$\begin{aligned} \mathbf{P}[A \cup B] &= \mathbf{P}[A] + \mathbf{P}[B] - \mathbf{P}[AB] \\ &= \mathbf{P}\left[A\bar{B}\right] + \mathbf{P}\left[\bar{A}B\right] + \mathbf{P}[AB], \end{aligned}$$

  where $\bar{\square}$ denotes the complement of Event $\square$.

II. Chance Experiments

A chance experiment is a random experiment that consists of three components: (i) experiment, denoted as E; (ii) outcomes, denoted as ζ; and (iii) events, denoted as $\{A\}, \{B\}, \{C\}$, *etc.*

Example 1: A Coin Toss

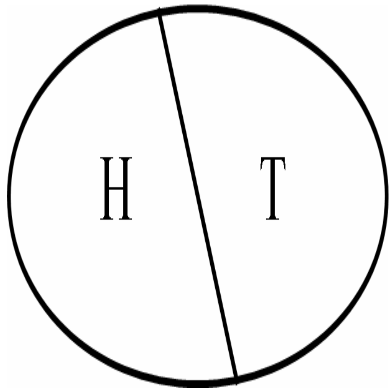

Fig. 1. Venn diagram for a coin tossing experiment with two possible outcomes: H (Head) and T (Tail).

Let E_a denote the experiment: A coin is tossed. Then, the outcomes of this experiment are $\zeta_1 = H$ and $\zeta_2 = T$:

$$\text{Event } A \triangleq \text{A Head results after the coin is tossed} = \{H\}\ ;$$

$$\text{Event } B \triangleq \text{A Tail results after the coin is tossed} = \{T\};$$

$$\text{Sample Space } S = \{H, T\}.$$

Note that the probability that neither H nor T has occurred in E_a is $\mathbf{P}[\phi] = 0$.

Example 2: Multiple Coin Tosses

Let E_b denote a new experiment: A coin is tossed 3 times in succession. Based on a simple permutation, we know that there are 2^3 outcomes in this experiment, and each of the outcomes can be written in triplets as shown below:

$$\begin{array}{cccccccc} \zeta_1 & \zeta_2 & \zeta_3 & \zeta_4 & \zeta_5 & \zeta_6 & \zeta_7 & \zeta_8 \\ HHH & HHT & HTH & THH & TTH & THT & HTT & TTT \end{array}$$

As an example of an event, suppose we are interested in $A \triangleq$ "the coin first comes up heads" $=$ $\{\zeta_1, \zeta_2, \zeta_3, \zeta_7\}$.

III. Basic Set Theory and Venn Diagrams

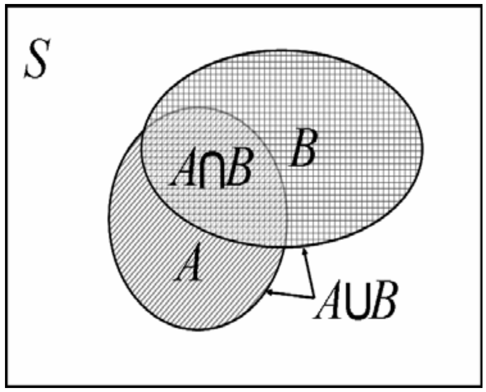

Fig. 2. A simple Venn diagram for illustrating Event A, Event B, Sample Space S, $A \cup B$, and $A \cap B$.

More Notation:

$$
\begin{aligned}
A_1 \cap A_2 \cap \cdots \cap A_N &= \bigcap_{i=1}^{N} A_i \\
A_1 \cup A_2 \cup \cdots \cup A_N &= \bigcup_{i=1}^{N} A_i .
\end{aligned}
$$

Note that the case where $N \to \infty$ is called *countably infinite*.

IV. Axioms of Probability Theory

Introduced by the Russian mathematician Kolmogorov, the following axioms form the foundations of probability theory:

- **Axiom 1**: $\mathbf{P}[A] \geq 0$.
- **Axiom 2**: $\mathbf{P}[S] = 1$.
- **Axiom 3**: if A and B are mutually exclusive *i.e.*, $A \cap B = \phi$, then $\mathbf{P}[A+B] = \mathbf{P}[A] + \mathbf{P}[B]$.
- **Axiom 4**: if $A_i \cap A_j = \phi$ for all i,j, where $i \neq j$, then $\mathbf{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbf{P}\left(A_i\right)$.

V. Statistical Independence

Statistical independence is an important concept widely used in probability theory. This concept is as important as Bayes Rule that was introduced in the previous lecture. Intuitively, two events are said to be statistically independent when the fact that the occurrence of one event does not statistically affect the occurence of the other event. The following two expressions mathematically state that Event A and Event B are statistically independent:

$$\begin{aligned} \mathbf{P}[A/B] &= P[A] \\ \mathbf{P}[AB] &= \mathbf{P}[A]\mathbf{P}[B] \end{aligned}$$

In words, the statistical independence between Event A and Event B implies that the probability of Event A and Event B occurring together equals the product of the probabilities of the two events. NB: This is a statistical property and is NOT the same thing with mutually exclusive.

VI. Total Probability Concept

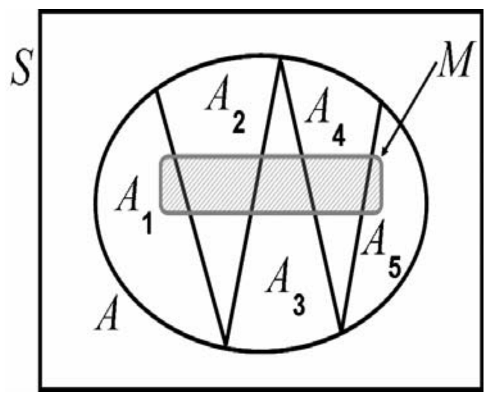

Fig. 3. A simple Venn diagram for illustrating Total Probability Theorem.

Total Probability Theorem states that the total probability of an event M equals the weighted sum of the conditional probability $\mathbf{P}[M/A_i]$ of its occuring in each event A_i of an exhaustive decomposition of S into *mutually exclusive events*, each such conditional probability being weighted by the probability $\mathbf{P}[A_i]$ that the conditioning event occurs.

Mathematically, Total Probability Theorem says that when the following conditions are satisfied: (i) $A = A_1 + A_2 + A_3 + \cdots + A_N$; and (ii) $\forall i, j; i \neq j, \mathbf{P}[A_iA_j] = 0$; then, $\mathbf{P}[A] = \mathbf{P}\left[A_1 + A_2 + \cdots + A_N\right] = \mathbf{P}\left[A_1\right] + \mathbf{P}\left[A_2\right] + \cdots + \mathbf{P}\left[A_N\right]$. As a result, one can observe that:

$$\begin{aligned} &\mathbf{P}[AM] = \mathbf{P}\left[\left(A_1 + A_2 + \cdots + A_N\right) M\right] \\ &= \mathbf{P}\left[A_1M\right] + \mathbf{P}\left[A_2M\right] + \cdots + \mathbf{P}\left[A_NM\right] \end{aligned}$$

$$= \mathbf{P}\left[A_1 \cap M\right] + \mathbf{P}\left[A_2 \cap M\right] + \cdots + \mathbf{P}\left[A_N \cap M\right] = \mathbf{P}[M]. \tag{1}$$

Dividing through by $\mathbf{P}[M]$, we have:

$$\frac{\mathbf{P}[AM]}{\mathbf{P}[M]} = \frac{\mathbf{P}\left[A_1 M\right]}{\mathbf{P}[M]} + \frac{\mathbf{P}\left[A_2 M\right]}{\mathbf{P}[M]} + \cdots + \frac{\mathbf{P}\left[A_N M\right]}{\mathbf{P}[M]}$$

Note that the previous equation admits the inverse of Bayes Rule, then:

$$\mathbf{P}[M] = \mathbf{P}\left[A_1/M\right] + \mathbf{P}\left[A_2/M\right] + \cdots + \mathbf{P}\left[A_N/M\right] = 1. \tag{2}$$

Using the fact that $\mathbf{P}\left[A_i \cap M\right] = \mathbf{P}\left[M/A_i\right]\mathbf{P}\left[A_i\right]$ (Bayes Rule) in eqn. (1), one gets:

$$\mathbf{P}[M] = \mathbf{P}\left[M/A_1\right]\mathbf{P}\left[A_1\right] + \mathbf{P}\left[M/A_2\right]\mathbf{P}\left[A_2\right] + \cdots + \mathbf{P}\left[M/A_N\right]\mathbf{P}\left[A_N\right] \tag{3}$$

Eqn. (3) is known as **the principle of total probability**.

Example from Telecommunications: Binary Asymmetric Channel

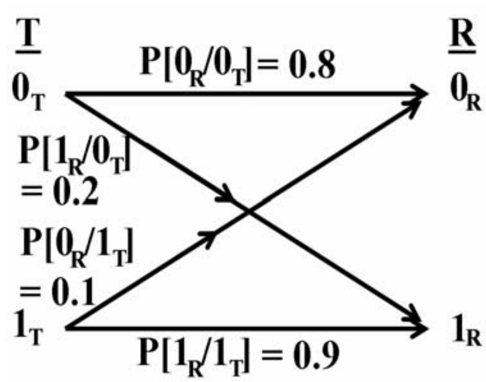

Fig. 4. Binary Asymmetric Channel.

In a binary communication channel, a transmitted bit and a received bit can take a value of either 0 or 1. However, due to external factors such as "white noise" and interference, erroneous transmission occurs, *i.e.*, a received bit may not be the same as the corresponding transmitted bit. Note that this example is called "Binary *Asymmetric* Channel" because $\mathbf{P}\left[0_{\mathrm{R}}/1_{\mathrm{T}}\right] \neq \mathbf{P}\left[1_{\mathrm{R}}/0_{\mathrm{T}}\right]$. If the two probabilities are equal, this channel model would be called "Binary Symmetric Channel."

Channel Matrix:

$$\begin{array}{cc} & \begin{array}{cc} 0_{\mathrm{R}} & 1_{\mathrm{R}} \end{array} \\ \begin{array}{c} 0_{\mathrm{T}} \\ 1_{\mathrm{T}} \end{array} & \left[\begin{array}{cc} 0.8 & 0.2 \\ 0.1 & 0.9 \end{array}\right] \end{array}$$

NB: $\mathbf{P}\left[0_{\mathrm{R}}/0_{\mathrm{T}}\right] \neq \mathbf{P}\left[0_{\mathrm{T}}/0_{\mathrm{R}}\right]$.

Note: In Binary Asymmetric Channel, calculations always make use of Bayes Rule. In this context, the rule can be stated as:

$$\begin{aligned} \mathbf{P}\left[0_{\mathrm{T}}0_{\mathrm{R}}\right] &= \mathbf{P}[0_{\mathrm{T}}/0_{\mathrm{R}}]\mathbf{P}[0_{\mathrm{R}}] \\ &= \mathbf{P}[0_{\mathrm{R}}/0_{\mathrm{T}}]\mathbf{P}[0_{\mathrm{T}}] \end{aligned}$$

Note: As will be shown, in order to solve a problem, you will need to use your judgement or even creativity to identify which formula to use.

Notation: $\mathbf{P}\left[0_{\mathrm{T}}\right]$ and $\mathbf{P}\left[1_{\mathrm{T}}\right]$ are called *a priori* (transmission) probabilities, and $\mathbf{P}\left[0_{\mathrm{R}}/0_{\mathrm{T}}\right]$, $\mathbf{P}\left[0_{\mathrm{R}}/1_{\mathrm{T}}\right]$, $\mathbf{P}\left[1_{\mathrm{R}}/0_{\mathrm{T}}\right]$, and $\mathbf{P}\left[1_{\mathrm{R}}/1_{\mathrm{T}}\right]$ are called *a posteriori* probabilities.

What is the value of $\mathbf{P}\left[0_{\mathrm{R}}\right]$? It can be shown that Bayes Rule and Total Probability Theorem can be used to find the answer as follows:

$$\begin{aligned} \mathbf{P}\left[0_{\mathrm{R}}\right] &= \mathbf{P}\left[0_{\mathrm{R}}\left(S_{\mathrm{T}}\right)\right] \\ &= \mathbf{P}\left[0_{\mathrm{R}}\left(0_{\mathrm{T}}+1_{\mathrm{T}}\right)\right] \\ &= \mathbf{P}\left[0_{\mathrm{R}}0_{\mathrm{T}}\right]+\mathbf{P}\left[0_{\mathrm{R}}1_{\mathrm{T}}\right] \\ &= \mathbf{P}\left[0_{\mathrm{R}}/0_{\mathrm{T}}\right]\mathbf{P}\left[0_{\mathrm{T}}\right]+\mathbf{P}\left[0_{\mathrm{R}}/1_{\mathrm{T}}\right]\mathbf{P}\left[1_{\mathrm{T}}\right] \end{aligned}$$

Similarly, one can compute $\mathbf{P}\left[1_{\mathrm{R}}\right]$ as follows:

$$\mathbf{P}\left[1_{\mathrm{R}}\right] = \mathbf{P}\left[1_{\mathrm{R}}/0_{\mathrm{T}}\right]\mathbf{P}\left[0_{\mathrm{T}}\right]+\mathbf{P}\left[1_{\mathrm{R}}/1_{\mathrm{T}}\right]\mathbf{P}\left[1_{\mathrm{T}}\right]$$

Let $\mathbf{P}[e]$ denote the probability of erroneous transmission. Note that e is a *compound event* because in this case, there are two possibilities for erroneous transmission. Thus, we can compute the probability of an erroneous transmission as:

$$\begin{aligned} \mathbf{P}[e] &= \mathbf{P}\left[e/0_{\mathrm{T}}\right]\mathbf{P}\left[0_{\mathrm{T}}\right]+\mathbf{P}\left[e/1_{\mathrm{T}}\right]\mathbf{P}\left[1_{\mathrm{T}}\right] \\ &= 0.2\times 0.5+0.1\times 0.5 = 0.5(0.2+0.1) \\ &= 0.15. \end{aligned}$$

VII. Bernoulli Trials

Every experiment that has only two outcomes (*e.g.* Head/Tail, or Success/Failure) can be called "Bernoulli Trials." Assume that an experiment with two outcomes *e.g.*, success or failure, is made. Let S and F denote sucess and failure, respectively, and then assign probability values to these outcomes: $\mathbf{P}[S] = p$, and $\mathbf{P}[F] = 1 - p = q$.

Suppose that the experiment is repeated 3 times, and the outcomes of all trials are independent. As an example, we can compute the probability that the first two experiments are successes and the last experiment is a failure as follows:

$$\mathbf{P}[SSF] = \mathbf{P}\left[S\right]\mathbf{P}\left[F\right]\mathbf{P}\left[F\right] = pq^2.$$

Now we can generalize this result by asking the following question:
What is the value of $\mathbf{P}\left[K \text{ successes in } N \text{ experiments of a particular order}\right]$?
Ans: $p^K(1-p)^{(N-K)}$.
What is the value of $\mathbf{P}[K \text{ successes}]$? This question tacitly implies that order is *not* important.
Ans: $\begin{pmatrix} N \\ K \end{pmatrix} p^K q^{(N-K)}$.
Note that $\begin{pmatrix} N \\ K \end{pmatrix}$ is read as "N choose K, " and $\begin{pmatrix} N \\ K \end{pmatrix} = \frac{N(N-1)\cdots(N-(K-1))}{K!}$.
Example: Assume that $N = 3$ and $K = 1$. One way to find $\mathbf{P}[1 \text{ success out of 3 trials}]$ is to list all the possible outcomes in Event "1 success out of 3 trials" as follows:

The 1^{st} outcome:	S	F	F	occurs with the probability of pq^2;
the 2^{nd} outcome:	F	S	F	occurs with the probability of pq^2;
and the 3^{rd} outcome:	F	F	S	occurs with the probability of pq^2.

As a result, the total probability of have 1 success out of 3 trials is $3pq^2$.

Based on the aforementioned answer for $\mathbf{P}[K \text{ successes}]$, we can then find the value of $\mathbf{P}[\textit{At least } K \text{ successes in } N \text{ experiments}]$ as: $\sum_{r=K}^{N} \begin{pmatrix} N \\ r \end{pmatrix} p^r q^{(N-r)} = 1 - \sum_{r=0}^{K-1} \begin{pmatrix} N \\ r \end{pmatrix} p^r q^{(N-r)}$.
Recall Binomial Coefficient Theorem:

$$\sum_{r=0}^{N} \begin{pmatrix} N \\ r \end{pmatrix} p^r q^{(N-r)} = 1$$

VIII. Connections among Bernoulli Trials, Binomial, and Poisson Law

As previously shown, the probability of having K successes out of N Bernoulli trials is $\begin{pmatrix} N \\ K \end{pmatrix} p^k(1-p)^k$, which will later be defined in a future lecture as Binomial Distribution.

Next, an important approximation of the binomial distribution can be obtained when the number of trials is very large *i.e.*, $N \gg 1$, and the probability of success is very small *i.e.*, $p \ll 1$. With the two aforementioned conditions, one has:

$$\begin{pmatrix} N \\ K \end{pmatrix} = \frac{N(N-1)\cdots(N-(K-1))}{K!} \approx \frac{N^K}{K!},$$

and

$$1-p \approx e^{-p} \ \ \text{[since } p \ll 1\text{]}.$$

Consequently, we can approximate the binomial distribution as:

$$\begin{aligned}
\begin{pmatrix} N \\ K \end{pmatrix} p^K(1-p)^{(N-K)} &\approx \frac{(pN)^K}{K!}e^{-p(N-K)} \\
&= \frac{a^K}{K!}e^{-pN}e^{pK} \\
&\approx \frac{(Np)^K}{K!}e^{-(Np)} \qquad [\because pK \to 0, \text{ and } \therefore e^{pK} \to 1]
\end{aligned}$$

Note that it can be shown the average number of successes of Bernoulli Trials is $a = Np$. Then the approximation just obtained can be restated as:

$$\begin{pmatrix} N \\ K \end{pmatrix} p^K(1-p)^{(N-K)} \approx \frac{a^K}{K!}e^{-a} \qquad \text{[This is called “Poisson Law”].}$$

Later in this course, the last expression will be defined as Poisson Distribution. We have thus demonstrated the inner connections among Bernoulli Trials, Binomial Distribution, and Poisson Distribution.