Introduction: Probability and Stochastic Processes

Henry Pfister

ECE 581: Random Signals and Noise

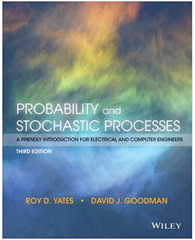
Lecture 1

Slides courtesy of Christ Richmond with slight modifications.

Administration / Staff

Instructor: Henry Pfister
Email: henry.pfister@duke.edu
Office: Room 305, Gross Hall
Class: Monday & Wednesday, 1:25-2:40PM, Wilkinson 132

Office hours: TBA
Zoom access possible at request:
`https://duke.zoom.us/my/henry.pfister`



Course Description and Reading Materials

Class Website: `https://canvas.duke.edu/login/saml`

Course Description:

- Graduate level introduction to probability, statistics, and stochastic processes.
- Goal is to develop mathematical methods for describing and analyzing engineering systems including electrical signals and systems corrupted by noise.

Required Textbook: Course and slides based on PSP (available online):

Yates and Goodman, *Probability and Stochastic Processes*, Third Edition, Wiley, 2014.

Useful References:

O. C. Ibe, *Fundamentals of Applied Probability and Random Processes*, 2014.

Prerequisites: (STA 130L/240L or Math 230/340 or ECE 380/555 or EGR 238L) or grad standing. By topic: (i) Calculus, (ii) linear algebra, (iii) linear systems, (iv) basic probability; Some basic familiarity with Python or Matlab.

ECE 581 Overview / Outlook

- Review of basic undergraduate probability (first ∼4-5 weeks)
  - Introduction, probability basics: experiments, models, probabilities
  - Sequential experiments and counting
  - Discrete (univariate) / Continuous (univariate)
  - Bivariate random variables, correlation, and conditional probability
  - Derived random variables and Jacobians
  - [Some videos assigned / suggested]
  - [Advanced optional HW assignments made available]
- Random vectors (multivariate) / Gaussians and Properties
- Sums of random variables and Central Limit Theorem
- Inequalities, convergence, and Law of Large Numbers
- Stochastic processes and Power Spectra
- Markov discrete stochastic process

Deterministic versus Random Systems

- Many science and engineering principles have specific deterministic relationships.
  - A deterministic system is a system that evolves without any randomness
    - a deterministic model produces same output from a given initial state and input [Wikipedia]
  - For example,
    - object moving with speed v over time duration t traverses distance $s = v \cdot t$
    - voltage V created by current I through load of resistance R is $V = I \cdot R$
- ECE 581 explores systems with some *uncertainty / randomness / unpredictability*.
  - For example, when measured by an observer we might say
    - distance s for moving object

$$s = v \cdot t + \varepsilon_s$$

    - voltage V due to current I through resistance load R is

$$V = I \cdot R + \varepsilon_V$$

  where ε in both cases is some random measurement error.

Probability Theory: Frequentist vs Bayesian View

- Many examples of experiments whose outcomes cannot be predicted perfectly in advance.
  - For example, lottery numbers, hand dealt in card game, final score of sports event, value of a stock next week, daily weather, etc. Can you think of other examples?
- Probability theory is a formal way of analzying random (uncertain or unpredictable) events by assigning events a number between 0 and 1 which has has two predominant interpretations:
  1. (**Frequentist View**) Probability is seen as a measure of the relative frequency (or fraction) of an event.
     - For example, if a coin flip results in "heads" with a probability of $\frac{1}{4}$, and "tails" with a probability of $\frac{3}{4}$, then if flipped 100 times, we would expect that approximately a quarter of those flips, i.e. 25 would result in heads, and 3 quarters, i.e. 75, in tails.
  2. (**Bayesian View**) Probability is interpreted as a measure of confidence regarding one's knowledge or belief about something.
     - For example, weather forcast may say: Low 51° F with a 70% chance of precipitation / 0.14 inches. Showers early with winds south at 20–30 mph.
     - Here 70 percent is measure of confidence associated with prediction; we may interpret this as meaning there's a probability of 0.7 that it will rain. Would you bring your umbrella?

Some Remarks[1]

- The Frequentist and Bayesian views of probability have both proven to be useful in practice.
- The mathematics of probability theory uses set theory to provide a flexible framework for describing and organizing a wide host of nondeterministic processes.
- The sample space for a random process/phenomena provides a global picture of the "experiment" and all its possible outcomes.
  - This picture is vital to understanding probability theory.
  - The sooner one learns to form such a picture, the quicker one will grasp the concepts of probability theory needed to solve problems.
  - Drake [5] "...stresses the sample space of representation of probabilistic processes and ... the need for explicit modeling of nondeterministic processes."
- Probability measure is the relevant metric defined on this sample space that we care about.
  - Three axioms (described in following sections) establish the nature and uniqueness of this measure.

[1] Some remarks from [5].

Some Basic Definitions

It is now useful to provide some formal definitions [with analogues from set theory].

- Experiment: a procedure yielding an observation
  - e.g. any nondeterministic process.
  - Observer cannot anticipate/predict outcome because of intrinsic randomness
- Outcome [Element]: any possible observation of an experiment
- Event [Set]: any collection of outcomes of an experiment
- Sample space [Universal Set]: "finest-grain" complete listing of all possible experiment outcomes
- Probability of an event: relative frequency of occurence that performance of the experiment will result in;
- Model: a specific assignment of probabilities for all events in sample space.

Coin Flip Experiment: Heads or Tails ?

- Single coin flip

- Two coin flips

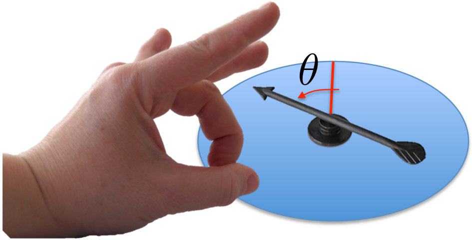
4-Sided Die Roll Experiment: Feeling Lucky ?

- Single die roll

- Two die rolls

Game Spinner: Which Direction ?

- Single spin



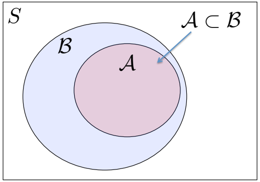
- Two spins

Set Theory: An Algebra of Events

- A set is a collection of things. The things in the set are called elements.
- There are many ways to define a set; often brackets $\{\cdot\}$ are used. For example, the following are sets: $\mathcal{A} = \{a, c, d\}$, $\mathcal{B} = \{\text{red, green, blue}\}$, or $\mathcal{F} = \{\spadesuit, \clubsuit, \heartsuit, \diamondsuit\}$.
  - notation $x \in \mathcal{X}$ means element x belongs to set $\mathcal{X}$
  - notation $x \notin \mathcal{X}$ means x is not a member of set $\mathcal{X}$
- Other example sets include $\mathcal{C} = \{x^2 | x = 1, 2, 3, 4, 5\} = \{1, 4, 9, 16, 25\}$, or $\mathcal{D} = \{x^2 | x = 1, 2, 3, \ldots\}$ that has an infinite number of elements.
  - Size of a set is sometimes called its cardinality, and denoted $|\cdot|$.
    - e.g. regarding above defined sets $|\mathcal{C}| = 5$ and $|\mathcal{D}| = \infty$.
- Set $\mathcal{A}$ is said to be a subset of $\mathcal{B}$ if every element of $\mathcal{A}$ also belongs to set $\mathcal{B}$; denoted $\mathcal{A} \subset \mathcal{B}$. For example, $\mathcal{C} \subset \mathcal{D}$ for sets defined above.

∗ Venn Diagrams (VD) are useful for visualizing sets.

∗ If set $\mathcal{A}$ can be created using *only* elements belonging to set $\mathcal{B}$, then $\mathcal{A} \subset \mathcal{B}$.



Applying Set Theory to Probability: Main Axioms

Axioms of Probability:

1. For any event $\mathcal{A}$, $\Pr(\mathcal{A}) \geq 0$ (i.e. probability is non-negative).
2. $\Pr(S) = 1$ (normalizes all probabilities to be in [0,1]).
3. If $\mathcal{A} \cap \mathcal{B} = \varnothing$, then $\Pr(\mathcal{A} \cup \mathcal{B}) = \Pr(\mathcal{A}) + \Pr(\mathcal{B})$

- Axiom 3 is motivated by behavior of relative frequencies for disjoint events
  - Consider experiment performed N times. Let the number of occurrences of disjoint events $\mathcal{A}, \mathcal{B}$ be $k(\mathcal{A}), k(\mathcal{B})$ respectively. Note the following relative frequencies ν:

$$\nu_{\mathcal{A}} = \frac{k(\mathcal{A})}{N}, \;\; \nu_{\mathcal{B}} = \frac{k(\mathcal{B})}{N}, \;\; \nu_{\mathcal{A} \cup \mathcal{B}} = \frac{k(\mathcal{A}) + k(\mathcal{B})}{N} = \nu_{\mathcal{A}} + \nu_{\mathcal{B}}.$$

  - These relative frequencies are expected to approach the true probabilities as $N \to \infty$, i.e. $\nu_{\mathcal{A}} \to \Pr(\mathcal{A})$, $\nu_{\mathcal{B}} \to \Pr(\mathcal{B})$, and $\nu_{\mathcal{A} \cup \mathcal{B}} \to \Pr(\mathcal{A} \cup \mathcal{B}) = \Pr(\mathcal{A}) + \Pr(\mathcal{B})$.
  - Clearly, axiom 3 is motivated by the behavior of these relative frequencies.
- All of conventional probability follows from these three axioms.

Some Relations of Probability

- Note by second and third axioms that

$$\Pr(\mathcal{S}) = \Pr(\mathcal{A} \cup \mathcal{A}^c) = \Pr(\mathcal{A}) + \Pr(\mathcal{A}^c) = 1 \Longrightarrow$$
$$\Pr(\mathcal{A}) = 1 - \Pr(\mathcal{A}^c) \text{ and } \Pr(\mathcal{A}^c) = 1 - \Pr(\mathcal{A})$$

- By first axiom we have that $\Pr(\mathcal{A}) \geq 0$ and $\Pr(\mathcal{A}^c) \geq 0$. Thus, it follows that $\Pr(\mathcal{A}^c) = 1 - \Pr(\mathcal{A}) \geq 0$, or simply $\Pr(\mathcal{A}) \leq 1$.
  - Hence, for any set $\mathcal{A}$ we have that $0 \leq \Pr(\mathcal{A}) \leq 1$.
- Consider three sets $\mathcal{A}, \mathcal{B}$ and $\mathcal{C}$ that are pairwise disjoint. Note that

$$\Pr(\mathcal{A} \cup \mathcal{B} \cup \mathcal{C}) = \Pr[(\mathcal{A} \cup \mathcal{B}) \cup \mathcal{C}] = \Pr(\mathcal{A} \cup \mathcal{B}) + \Pr(\mathcal{C}) = \Pr(\mathcal{A}) + \Pr(\mathcal{B}) + \Pr(\mathcal{C})$$

where last two equalities follow from third axiom.

- Clearly, this argument can be extended to n pairwise disjoint sets $\mathcal{A}_i$, $i = 1, 2, \ldots, n$ to obtain

$$\Pr(\mathcal{A}_1 \cup \mathcal{A}_2 \cup \cdots \cup \mathcal{A}_n) = \Pr(\mathcal{A}_1) + \Pr(\mathcal{A}_2) + \cdots + \Pr(\mathcal{A}_n)$$

  - Although this assumed n was finite, it is noteworthy that such also holds for an infinite set of mutually exclusive sets, i.e. as $n \to \infty$.

More Relations of Probability

- Now consider two sets $\mathcal{A}$ and $\mathcal{B}$ not necessarily disjoint, i.e. $\mathcal{A} \cap \mathcal{B} \neq \varnothing$.
  - What is $\Pr(\mathcal{A} \cup \mathcal{B}) = ?$
  - Note that their union can be expressed as

$$\mathcal{A} \cup \mathcal{B} = \mathcal{A} \cup (\mathcal{A}^c \cap \mathcal{B}), \quad \text{i.e. the union of two disjoint sets.}$$

  - Thus, by third axiom we know that

$$\Pr(\mathcal{A} \cup \mathcal{B}) = \Pr[\mathcal{A} \cup (\mathcal{A}^c \cap \mathcal{B})] = \Pr(\mathcal{A}) + \Pr(\mathcal{A}^c \cap \mathcal{B}). \tag{1}$$

  - Note that set $\mathcal{B}$ can be written as the union of two disjoint sets, i.e. $\mathcal{B} = \mathcal{S} \cap \mathcal{B} = (\mathcal{A} \cup \mathcal{A}^c) \cap \mathcal{B} = (\mathcal{A} \cap \mathcal{B}) \cup (\mathcal{A}^c \cap \mathcal{B})$ where $(\mathcal{A} \cap \mathcal{B}) \cap (\mathcal{A}^c \cap \mathcal{B}) = \varnothing$. Thus, by third axiom we have

$$\Pr(\mathcal{B}) = \Pr(\mathcal{A} \cap \mathcal{B}) + \Pr(\mathcal{A}^c \cap \mathcal{B}) \Longrightarrow \Pr(\mathcal{A}^c \cap \mathcal{B}) = \Pr(\mathcal{B}) - \Pr(\mathcal{A} \cap \mathcal{B}).$$

  - Finally, combining this with (1) we obtain the general relation

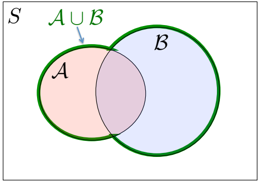
$$\Pr(\mathcal{A} \cup \mathcal{B}) = \Pr(\mathcal{A}) + \Pr(\mathcal{B}) - \Pr(\mathcal{A} \cap \mathcal{B}). \tag{2}$$

- Note that (2) is a generalization of the third axiom that holds whether or not the two sets are disjoint.
  - We've also shown that (2) is a direct consequence of the axioms of probability.

More Set Theory: An Algebra of Events

- Sets $\mathcal{A}$ and $\mathcal{B}$ are said to be equal *if and only if* $\mathcal{A} \subset \mathcal{B}$ and $\mathcal{B} \subset \mathcal{A}$; denoted $\mathcal{A} = \mathcal{B}$.
  - Recall the statement "if and only if" means both...
  1. $\mathcal{A} = \mathcal{B} \Longrightarrow \mathcal{A} \subset \mathcal{B}$ and $\mathcal{B} \subset \mathcal{A}$.
  2. $\mathcal{A} \subset \mathcal{B}$ and $\mathcal{B} \subset \mathcal{A} \Longrightarrow \mathcal{A} = \mathcal{B}$.
  - Also, note no regard is given to "order of elements," i.e. if $\mathcal{A} = \{a, c, d\}$, $\mathcal{B} = \{d, a, c\}$, then $\mathcal{A}$ and $\mathcal{B}$ represent the same set, i.e. $\mathcal{A} = \mathcal{B}$.
- The set of all possible outcomes is referred to as the universal set; denoted by S.
  - Universal set is defined by the experiment or context.
  - It represents the "universe" of all possible events.
- Union of sets $\mathcal{A}$ and $\mathcal{B}$ is set of all elements either in $\mathcal{A}$, or in $\mathcal{B}$ or in both $\mathcal{A}$ & $\mathcal{B}$.
  - union of sets is denoted by $\mathcal{A} \cup \mathcal{B}$, and sometimes as $\mathcal{A} + \mathcal{B}$.

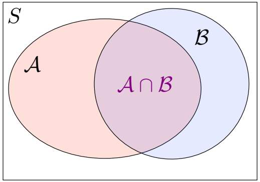
$*$ $x \in \mathcal{A} \cup \mathcal{B}$ if and only if $x \in \mathcal{A}$ or $x \in \mathcal{B}$; i.e. $\cup$ corresponds to logical "or"



More Set Theory: An Algebra of Events

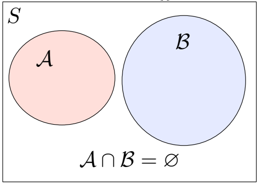
- Intersection of sets $\mathcal{A}$ and $\mathcal{B}$ is set of all elements contained in both $\mathcal{A}$ and $\mathcal{B}$.
  - intersection of sets is denoted by $\mathcal{A} \cap \mathcal{B}$, and we will use the shorthand $\mathcal{AB}$.

$*$ $x \in \mathcal{A} \cap \mathcal{B}$ if and only if $x \in \mathcal{A}$ and $x \in \mathcal{B}$;
i.e. $\cap$ corresponds to logical "and"



- If sets $\mathcal{A}$ and $\mathcal{B}$ have no elements in common, then how do we represent this?
- The null set is defined as the set having no elements, i.e. $\{\}$;
  - also called the empty set and denoted $\varnothing$.

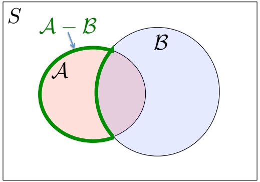
$*$If $\mathcal{A}$ and $\mathcal{B}$ have no elements in common,
then we say $\mathcal{A} \cap \mathcal{B} = \varnothing$.



  - The null set is unique among sets and necessary for completeness.

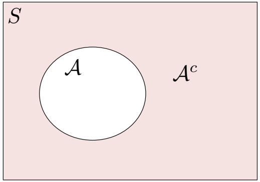
More Set Theory: An Algebra of Events

- The null set is a subset of every set, i.e. $\varnothing \subset \mathcal{A}$ for any set $\mathcal{A}$.
  - The null set $\varnothing$ has no elements and thus any element in $\varnothing$ must in $\mathcal{A}$
- Similarly, for any set $\mathcal{A}$ we have $\mathcal{A} \cap \varnothing = \varnothing$, and $\mathcal{A} \cup \varnothing = \mathcal{A}$.
- Set difference $\mathcal{A} - \mathcal{B}$ is set of all elements in $\mathcal{A}$ that are not in $\mathcal{B}$.



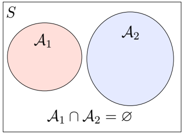
- Complement of set $\mathcal{A}$, denoted $\mathcal{A}^c$, is set of all elements in S that are not in $\mathcal{A}$.

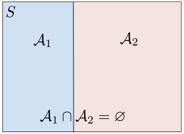
∗ $x \in \mathcal{A}^c$ if and only if $x \notin \mathcal{A}$; thus, $\mathcal{A} \cap \mathcal{A}^c = \varnothing$.
∗ Clearly $\mathcal{A}^c = S - \mathcal{A}$; thus, $S = A \cup A^c$;
∗ If we let $\mathcal{A} = S$, then we see that $S^c = S - S = \varnothing$; also, $\varnothing^c = S - \varnothing = S$.



More Set Theory: An Algebra of Events

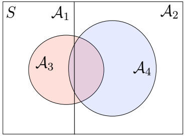
- Collection of sets $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n$ is mutuallay exclusive if and only if $\mathcal{A}_i \cap \mathcal{A}_j = \varnothing$ for $i \neq j$. Such sets are said to be pairwise disjoint.





- Collection of sets $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n$ is collectively exhaustive if and only if $\mathcal{A}_1 \cup \mathcal{A}_2 \cup \cdots \cup \mathcal{A}_n = S$.

  * Which example above is also collectively exhaustive?

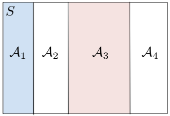


- Some useful shorthand notation for set unions and intersections:

$$\bigcup_{i=1}^{n} \mathcal{A}_i = \mathcal{A}_1 \cup \mathcal{A}_2 \cup \cdots \cup \mathcal{A}_n, \qquad \bigcap_{i=1}^{n} \mathcal{A}_i = \mathcal{A}_1 \cap \mathcal{A}_2 \cap \cdots \cap \mathcal{A}_n$$

More Set Theory: An Algebra of Events

- A collection of sets $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n$ that is both mutually exclusive and collectively exhaustive is said to be a partition of the universal set.



- Recall, set unions are analogous to a logical "or" operation, and set intersection are analogous to a logical "and" operation.
  - Logic and reasoning have long history of formalism
    - Represents the basis for boolean algebra or binary arithmetic
    - Indeed, unions and intersections can define a type of "algebra" for sets
    - There are several axioms/laws for this algebra. Six are listed below:

$$
\begin{array}{ll}
\mathcal{A} \cup \mathcal{B} = \mathcal{B} \cup \mathcal{A} & \text{Commutative} \\
\mathcal{A} \cap \mathcal{B} = \mathcal{B} \cap \mathcal{A} & \\
\mathcal{A} \cup (\mathcal{B} \cup \mathcal{C}) = (\mathcal{A} \cup \mathcal{B}) \cup \mathcal{C} & \text{Associative} \\
\mathcal{A} \cap (\mathcal{B} \cap \mathcal{C}) = (\mathcal{A} \cap \mathcal{B}) \cap \mathcal{C} & \\
\mathcal{A} \cap (\mathcal{B} \cup \mathcal{C}) = (\mathcal{A} \cap \mathcal{B}) \cup (\mathcal{A} \cap \mathcal{C}) & \text{Distributive} \\
\mathcal{A} \cup (\mathcal{B} \cap \mathcal{C}) = (\mathcal{A} \cup \mathcal{B}) \cap (\mathcal{A} \cup \mathcal{C}) & \\
(\mathcal{A}^c)^c = \mathcal{A} & \\
\mathcal{A} \cap \mathcal{A}^c = \varnothing & \text{Compliment} \\
\mathcal{A} \cap S = \mathcal{A} & \text{Identity}
\end{array}
$$

    - Most axioms can be established graphically via a Venn Diagram.

More Set Theory: De Morgan's Law

- **De Morgan's Law**: Given sets $\mathcal{A}$ and $\mathcal{B}$, the following equality holds

$$(\mathcal{A} \cup \mathcal{B})^c = \mathcal{A}^c \cap \mathcal{B}^c. \tag{3}$$

  Proof: See Venn Diagram.

  Suppose $x \in (\mathcal{A} \cup \mathcal{B})^c$; then $x \notin \mathcal{A} \cup \mathcal{B}$, i.e. $x \notin \mathcal{A}$ and $x \notin \mathcal{B} \Longrightarrow x \in \mathcal{A}^c$ and $x \in \mathcal{B}^c$. Thus, $x \in \mathcal{A}^c \cap \mathcal{B}^c$.

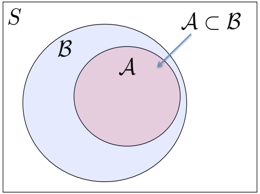
  Now suppose $x \in \mathcal{A}^c \cap \mathcal{B}^c$; then $x \in \mathcal{A}^c$ and $x \in \mathcal{B}^c$, i.e. $x \notin \mathcal{A}$ and $x \notin \mathcal{B} \Longrightarrow x \notin \mathcal{A} \cup \mathcal{B}$. Thus, $x \in (\mathcal{A} \cup \mathcal{B})^c$. ■

- Since (3) is true for any sets $\mathcal{A}, \mathcal{B}$, replace each set with its compliment, i.e.

$$(\mathcal{A}^c \cup \mathcal{B}^c)^c = \mathcal{A} \cap \mathcal{B} \Longrightarrow \mathcal{A}^c \cup \mathcal{B}^c = (\mathcal{A} \cap \mathcal{B})^c$$

  thus, we have an alternative yet equivalent form of De Morgan's law.

More Relations of Probability



- Now consider two sets such that $\mathcal{A} \subset \mathcal{B}$. Clearly, $\mathcal{A} = \mathcal{A} \cap \mathcal{B}$. Note that

$$\mathcal{B} = (\mathcal{A} \cap \mathcal{B}) \cup (\mathcal{A}^c \cap \mathcal{B}) \Longrightarrow \Pr(\mathcal{B}) = \Pr(\mathcal{A} \cap \mathcal{B}) + \Pr(\mathcal{A}^c \cap \mathcal{B}) \Longrightarrow$$

$$\Pr(\mathcal{B}) = \Pr(\mathcal{A}) + \Pr(\mathcal{A}^c \cap \mathcal{B}) \geq \Pr(\mathcal{A}) \Longrightarrow \Pr(\mathcal{B}) \geq \Pr(\mathcal{A}).$$

- Hence, if $\mathcal{A} \subset \mathcal{B}$, then $\Pr(\mathcal{A}) \leq \Pr(\mathcal{B})$.

Event Probabilities and the Sample Space

- Recall the sample space is an exhaustive list of all possible experimental outcomes.
  - Note each element can be interpreted as a set containing only one element.
  - These single element sets are collectively exhaustive and pairwise disjoint.
- Let the sample space be given by a finite list of elements, i.e. $S = \{s_1, s_2, \ldots, s_n\}$ where s_i represents individual elements (outcomes).
- If an event is given by any subset of these elements, e.g. $\mathcal{B} = \{s_{j_1}, s_{j_2}, \ldots, s_{j_m}\}$ where $j_k \in \{1, 2, \ldots, n\}$, $k = 1, 2, \ldots, m$, then by the third axiom the probability of event $\mathcal{B}$ is

$$\Pr(\mathcal{B}) = \Pr(\{s_{j_1}\}) + \Pr(\{s_{j_2}\}) + \cdots + \Pr(\{s_{j_m}\}) \tag{4}$$

  where we noted that $\mathcal{B} = \{s_{j_1}\} \cup \{s_{j_2}\} \cup \cdots \cup \{s_{j_m}\}$ and $s_{j_k} \cap s_{j_l} = \varnothing$ for $k \neq l$.
- Thus, the probability of any event is obtained by summing the individual probabilities of each element of the sample space making up the event.

Examples for Event Probabilities and the Sample Space

Revisit sample space examples:

- Examples discussed in class:
  - Two coin flips
  - Single 4-sided die roll
  - Two 4-sided die rolls
  - Single game spin

Fixed Volume of Unity Probability

- Consider a universal set $S = \{s_1, s_2, \ldots, s_n\}$ made up of a finite set of equally likely (fair) outcomes s_i, i.e. such that $\Pr(s_i) = p$, $i = 1, 2, \ldots, n$.
  - By second axiom of probability and arguments justifying (4), we note that

$$\Pr(S) = \Pr(s_1) + \Pr(s_2) + \cdots + \Pr(s_n) = np = 1 \Longrightarrow p = \frac{1}{n}.$$

  - Note as number of possible outcomes n increases (i.e. size of universal set) the probability of any particular/specific outcome decreases;
    - Indeed, $p \to 0$ as $n \to \infty$.
  - There's a total fixed "volume" of unity probability that gets distributed among all outcomes of the sample space.
  - As the total number of possible outcomes increases, the total probability remains fixed at unity but it gets spread among more elements.

Quiz 1.3

A student's test score T is an integer between 0 and 100 corresponding to the experimental outcomes $s_0, \ldots, s_{100}$. A score of 90 to 100 is an A, 80 to 89 is a B, 70 to 79 is a C, 60 to 69 is a D, and below 60 is a failing grade of F. If all scores between 51 and 100 are equally likely and a score of 50 or less never occurs, find the following probabilities:

(a) P[$\{s_{100}\}$]

(b) P[A]

(c) P[F]

(d) P[$T < 90$]

(e) P[a C grade or better]

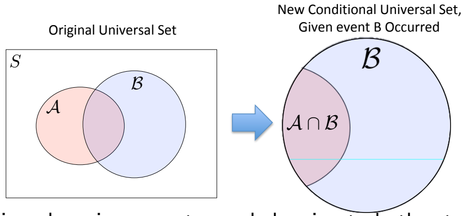
(f) P[student passes]

Conditional Probability

- When one first observes an experiment/process there may be very little known about the potential chances of outcomes; thus, it is reasonable to adopt a probability model that reflects this uncertainty.
- As one continues to observe the process, however, additional information may become available that changes the likelihood of certain outcomes.
  - For example, assume yesterday the weatherman said that there's a 50% chance of rain today, i.e. he had no idea whether or not it will rain today. Early this morning, however, after observing cloud cover, temperature, humidity, windspeed, and other meteorological data the weatherman proclaimed that there's a 75% chance of rain today.
    - Additional information from a side event lead to a change in his model.
- The notion of conditional probability addresses the adjustment of probabilities based on additional information.

Conditional Probability

- Consider the [original] universal set S shown in figure that includes events $\mathcal{A}$ and $\mathcal{B}$:



  - Clearly, $\mathcal{A} \cap \mathcal{B} \neq \varnothing$, i.e. observing an outcome belonging to both sets is possible.

- Assume that the probabilities assigned to events $\mathcal{A}, \mathcal{B}$ are given by $\Pr(\mathcal{A})$, $\Pr(\mathcal{B}) \neq 0$, and $\Pr(\mathcal{A} \cap \mathcal{B})$.

- Now assume we were unable to directly observe the experiment, but we are informed that event $\mathcal{B}$ has occurred. How should we adjust our model to reflect this new information?
  - Specifically, what can we say about the probability of event $\mathcal{A}$, given this additional information about the occurrence of $\mathcal{B}$?

Conditional Probability

- Conditional probability characterizes our knowledge (i.e. updates the model) of $\mathcal{A}$ when we know that $\mathcal{B}$ has occurred, and is denoted $\Pr(\mathcal{A}|\mathcal{B})$.
- As illustrated in figure, once informed that event $\mathcal{B}$ has occurred, then effectively set $\mathcal{B}$ is treated as the new universal set.
  - Note we have received no information, however, leading us to alter the relative probabilities within $\mathcal{B}$, but we can renormalize the probabilities for all elements in $\mathcal{B}$ to obtain an updated model for this new conditional universal set.
- Conditional probability of event $\mathcal{A}$ given that $\mathcal{B}$ has occurred is defined as

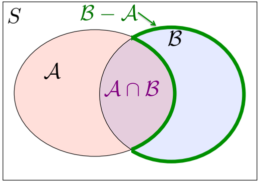
$$\Pr(\mathcal{A}|\mathcal{B}) = \frac{\substack{\text{Original probability} \\ \text{of } \mathcal{A} \text{ within set } \mathcal{B}}}{\substack{\text{Total probability} \\ \text{of set } \mathcal{B}}} = \frac{\Pr(\mathcal{A} \cap \mathcal{B})}{\Pr(\mathcal{B})}.$$

  - Definition makes sense only if $\Pr(\mathcal{B}) \neq 0$: if $\Pr(\mathcal{B}) = 0$, then event $\mathcal{B}$ is an impossible event, and the conditional probability becomes nonsensical and undefined.
  - Note, it follows that:

$$\Pr(\mathcal{A} \cap \mathcal{B}) = \Pr(\mathcal{B}) \Pr(\mathcal{A}|\mathcal{B}) \tag{5}$$

Conditional Probability

- Note from figure that $\mathcal{B} = (\mathcal{A} \cap \mathcal{B}) \cup (\mathcal{B} - \mathcal{A})$, i.e. set $\mathcal{B}$ is decomposable as the union of two disjoint sets:

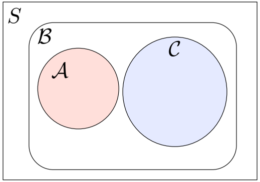


Thus, by third axiom and then renormalizing:

$$\begin{aligned} \Pr(\mathcal{B}) &= \Pr(\mathcal{A} \cap \mathcal{B}) + \Pr(\mathcal{B} - \mathcal{A}) \implies \\ \frac{\Pr(\mathcal{B})}{\Pr(\mathcal{B})} &= \frac{\Pr(\mathcal{A} \cap \mathcal{B})}{\Pr(\mathcal{B})} + \frac{\Pr(\mathcal{B} - \mathcal{A})}{\Pr(\mathcal{B})} \implies \\ 1 = \Pr(\mathcal{B}|\mathcal{B}) &= \Pr(\mathcal{A}|\mathcal{B}) + \Pr(\mathcal{B} - \mathcal{A}|\mathcal{B}). \end{aligned} \tag{6}$$

- The new conditional universal set likewise obeys the three axioms of probability:

1. $\Pr(\mathcal{A}|\mathcal{B}) \geq 0$.
2. $\Pr(\mathcal{B}|\mathcal{B}) = 1$.
3. For $\mathcal{A}, \mathcal{C} \subset \mathcal{B}$ where $\mathcal{A} \cap \mathcal{C} = \varnothing$,
   then $\Pr(\mathcal{A} \cup \mathcal{C}|\mathcal{B}) = \Pr(\mathcal{A}|\mathcal{B}) + \Pr(\mathcal{C}|\mathcal{B})$.



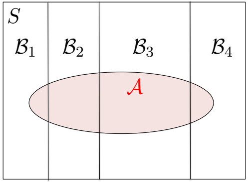
Law of Total Probability

- Given a partition $B = \{\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_m\}$, any event $\mathcal{A}$ can be written as the union of a set of mutually exclusive events, i.e. $\mathcal{A} = \mathcal{A} \cap S = \mathcal{A} \cap (\mathcal{B}_1 \cup \mathcal{B}_2 \cup \cdots \cup \mathcal{B}_m)$ implies

$$\mathcal{A} = (\mathcal{A} \cap \mathcal{B}_1) \cup (\mathcal{A} \cap \mathcal{B}_2) \cup \cdots \cup (\mathcal{A} \cap \mathcal{B}_m).$$

By the third axiom,

$$\Pr(\mathcal{A}) = \Pr(\mathcal{A} \cap \mathcal{B}_1) + \Pr(\mathcal{A} \cap \mathcal{B}_2) + \cdots + \Pr(\mathcal{A} \cap \mathcal{B}_m).$$



- This is called law of total probability since $\sum_{i=1}^{m} \Pr(\mathcal{B}_i) = 1$.
- Assuming $\Pr(\mathcal{B}_i) > 0$, $i = 1, 2, \ldots, m$, the relation (5) can also be written as

$$\Pr(\mathcal{A}) = \Pr(\mathcal{B}_1)\Pr(\mathcal{A}|\mathcal{B}_1) + \Pr(\mathcal{B}_2)\Pr(\mathcal{A}|\mathcal{B}_2) + \cdots + \Pr(\mathcal{B}_m)\Pr(\mathcal{A}|\mathcal{B}_m) \quad (7)$$

- If *only* conditional probabilities are available, then (7) gives unconditional probabilities.

Bayes Theorem

- Some applications naturally conditional probability $\Pr(\mathcal{A}|\mathcal{B})$ when one wants $\Pr(\mathcal{B}|\mathcal{A})$.
- As long as $\Pr(\mathcal{A}) \neq 0$ and $\Pr(\mathcal{B}) \neq 0$, Bayes theorem provides a relation between the two:

$$\Pr(\mathcal{B}|\mathcal{A}) = \frac{\Pr(\mathcal{A} \cap \mathcal{B})}{\Pr(\mathcal{A})} = \frac{\Pr(\mathcal{B}) \cdot \Pr(\mathcal{A}|\mathcal{B})}{\Pr(\mathcal{A})} \implies$$

$$\Pr(\mathcal{B}_i|\mathcal{A}) = \frac{\Pr(\mathcal{B}_i) \cdot \Pr(\mathcal{A}|\mathcal{B}_i)}{\Pr(\mathcal{A})} = \frac{\Pr(\mathcal{B}_i) \cdot \Pr(\mathcal{A}|\mathcal{B}_i)}{\sum_{i=1}^{m} \Pr(\mathcal{B}_i) \Pr(\mathcal{A}|\mathcal{B}_i)} \tag{8}$$

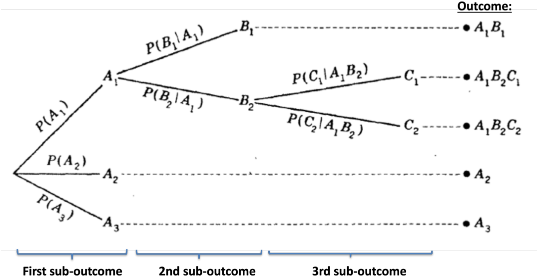
where the first equality follows from

$$\Pr(\mathcal{A}\mathcal{B}) = \Pr(\mathcal{B}) \cdot \Pr(\mathcal{A}|\mathcal{B}) = \Pr(\mathcal{A}) \cdot \Pr(\mathcal{B}|\mathcal{A}) \tag{9}$$

and last equality uses (7) to express the prior for $\mathcal{A}$ in terms of conditional probabilities.

Sample Space for Sequential Experiments

- Probability tree diagrams capture sequential nature (causality) of experiments.
- Figure illustrates such a tree diagram of the sample space



Sum of branch probabilities emerging from any node must sum to unity!

- Probabilities for first branches are prior probabilities.
  - i.e. $\Pr(A_1), \Pr(A_2)$, and $\Pr(A_3)$ where $\Pr(A_1) + \Pr(A_2) + \Pr(A_3) = 1$
- Probabilities of all other subsequent branches are conditional probabilities
  - $\Pr(B_1|A_1), \Pr(B_2|A_1), \Pr(C_1|A_1 B_2)$ and $\Pr(C_2|A_1 B_2)$
- Probability of an outcome (element) given by all product of tree probabilities, e.g. $\Pr(A_1 B_2 C_2) = \Pr(A_1) \Pr(B_2|A_1) \Pr(C_2|A_1 B_2) = \Pr(A_1 B_2) \Pr(C_2|A_1 B_2)$.

Quiz 1.5

Monitor customer behavior in the Phonesmart store. Classify the behavior as buying (B) if a customer purchases a smartphone. Otherwise the behavior is no purchase (N). Classify the time a customer is in the store as long (L) if the customer stays more than three minutes; otherwise classify the amount of time as rapid (R). Based on experience with many customers, we use the probability model $\mathrm{P}[N] = 0.7$, $\mathrm{P}[L] = 0.6$, $\mathrm{P}[NL] = 0.35$. Find the following probabilities:

(a) $\mathrm{P}[B \cup L]$

(b) $\mathrm{P}[N \cup L]$

(c) $\mathrm{P}[N \cup B]$

(d) $\mathrm{P}[LR]$

Independence

- Events $\mathcal{A}$ and $\mathcal{B}$ are said to be independent if and only if

$$\Pr(\mathcal{AB}) = \Pr(\mathcal{A}) \cdot \Pr(\mathcal{B}). \tag{10}$$

  - Recalling that $\Pr(\mathcal{AB}) = \Pr(\mathcal{A})\Pr(\mathcal{B}|\mathcal{A}) = \Pr(\mathcal{B})\Pr(\mathcal{A}|\mathcal{B})$, if events $\mathcal{A}, \mathcal{B}$ are independent, then $\Pr(\mathcal{B}|\mathcal{A}) = \Pr(\mathcal{B})$ and $\Pr(\mathcal{A}|\mathcal{B}) = \Pr(\mathcal{A})$.
    - It is intuitive that if events $\mathcal{A}$ and $\mathcal{B}$ are independent then knowledge of $\mathcal{B}$ does not change my initial model for $\mathcal{A}$, i.e. $\Pr(\mathcal{A}|\mathcal{B}) = \Pr(\mathcal{A})$; and likewise knowledge of $\mathcal{A}$ does not change my initial model for $\mathcal{B}$, i.e. $\Pr(\mathcal{B}|\mathcal{A}) = \Pr(\mathcal{B})$.
  - Independence means events $\mathcal{A}, \mathcal{B}$ do not influence each other, i.e. no cause and effect happening; knowledge of one does not inform of the other.
- Sets $\mathcal{A}, \mathcal{B}$ cannot be independent if they are mutually exclusive. Why is this?
- It is seldom obvious that two events are independent; in most cases one must compute the required probabilities and test for the validity of relation (10).

More on Independence

- Three events $\mathcal{A}, \mathcal{B}, \mathcal{C}$ are said to be mutually independent if and only if

$$\begin{aligned} \Pr(\mathcal{AB}) &= \Pr(\mathcal{A}) \Pr(\mathcal{B}) \\ \Pr(\mathcal{AC}) &= \Pr(\mathcal{A}) \Pr(\mathcal{C}) \\ \Pr(\mathcal{BC}) &= \Pr(\mathcal{B}) \Pr(\mathcal{C}) \\ \Pr(\mathcal{ABC}) &= \Pr(\mathcal{A}) \Pr(\mathcal{B}) \Pr(\mathcal{C}). \end{aligned}$$

- n events $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n$ are said to be mutually independent if and only if for any $k = 2, 3, \ldots, n$ the following relation holds:

$$\Pr(\mathcal{A}_{i_1} \mathcal{A}_{i_2} \cdots \mathcal{A}_{i_k}) = \Pr(\mathcal{A}_{i_1}) \Pr(\mathcal{A}_{i_2}) \cdots \Pr(\mathcal{A}_{i_k})$$

  for any combination $(i_1, i_2, \ldots, i_k)$ of k objects chosen from set $\{2, \ldots, n\}$.
  - Pairwise independence is weaker condition than this mutual independence.
- Two events $\mathcal{A}, \mathcal{B}$ are said to be conditionally independent given $\mathcal{C}$ if it is true that

$$\Pr(\mathcal{AB}|\mathcal{C}) = \Pr(\mathcal{A}|\mathcal{C}) \Pr(\mathcal{B}|\mathcal{C})$$

  assuming $\Pr(\mathcal{C}) > 0$.

References

[1] Yates and Goodman, Probability and Stochastic Processes, Third Edition, Wiley, 2014.

[2] O. C. Ibe, Fundamentals of Applied Probability and Random Processes, Elsevier Inc., 2014.

[3] S. M. Kay, Intuitive Probability and Random Processes Using Matlab, Springer, 2006.

[4] A. Papoulis and S. U. Pillai, Probability, Random Variables, and Stochastic Processes, 4th Edition, McGraw Hill, 2002.

[5] A. W. Drake, *Fundamentals of Applied Probability Theory*, McGraw-Hill Inc., 1967.

[6] D. P. Bertsekas and J. N. Tsitsiklis, *Introduction to Probability*, Athena Sci., 2002.