Probability and Random Process

Aimin Tang

The University of Michigan- Shanghai Jiao Tong University Joint Institute
Shanghai Jiao Tong University

Sep. 10 2019

Ve501 2020-2021 Fall

Based on Lecture Notes by Prof. Yong Long

Outline

- 1. Introduction to Probability
  - Application example
  - Review of set and functions
  - Models of random experiments
  - Axioms and properties of probability
  - Conditional probability
  - Independence of events
  - Combinatorics and probability

Models of random experiments

Random experiments

- Examples of random experiments:
  - tossing a coin/dice several times
  - draw cards
  - throw a dart to a dartboard
  - spinning a wheel/pointer
  - stirring a pan of water
  - read a letter of text, read a sequence of letters
  - record the times of occurrence of phone calls (emails arriving at a server)
  - measuring the length of queue at an airport

Probability model

- What does a probability model model?
  - It models frequency $\longrightarrow$ law of nature
- Probability models for a random experiment $(\Omega, \mathcal{A}, P, X)$
  - **probability space** $(\Omega, \mathcal{A}, P)$
    - sample space Ω (in some textbooks S is used)
    - event space $\mathcal{A}$
    - probability measure P
  - **variable name** $X/Y/Z$ (a random variable)

Sample space

- **Sample space** Ω is the set of all possible experiment outcomes.
  - Judgement is required. (distinct, indecomposable)
- Example: tossing a die
  1. Ω={1, 2, 3, 4, 5, 6}
  2. Ω= {0, 1, 2, 3, 4, 5, 6, . . .}
  3. Ω= {even, odd}
  4. Ω= {1, 2, 3, 4, 5, 6, even, odd}
  5. Ω= {2, 3, 4, 5, 6}

Which ones are correct?

- 1,3

Example

- tossing two dices
  - $\Omega = \{(1, 1), (1, 2), \ldots, (1, 6), (2, 1), \ldots, (6, 6)\}$
  - $\Omega = \{2, 3, 4, \ldots, 12\}$

- wheel spin
  - $\Omega = [0, 2\pi)$

- throw a dart to a circular dartboard
  - $\Omega = \{(x, y) : x^2 + y^2 \leq 1\}$

Event and outcome

- A subset of Ω is called an **event**
- Elements or points in the sample space Ω are called **outcomes**

Example

- Example: In a radar system, the voltage of a noise waveform at time t can be viewed as possibly being any real number. The first step in modeling such a noise voltage is to consider the sample space consisting of all real numbers, i.e.,

$$\Omega = (-\infty, \infty)$$

Is $\{1.5\}$ an event?

- Yes. Notice that singleton set, that is set consisting of a single point, are also events; e.g. $\{1.5\}$, $\{-8\}$, $\{\pi\}$. Be sure you understand the difference between the outcome -8 and the event $\{-8\}$, which is the set consisting of the single outcome -8.

Event space

- **Event space** $\mathcal{A}$ is a collection of subsets of the sample space Ω. (set of sets)
- $\mathcal{A}$ describes the "information" that
  - one has about the system that yields uncertain outcomes.
  - equivalently one has about the experiment that describes the uncertain outcomes produced by the system.

- Example: tossing a die
  - Sample space $\Omega = \{1, 2, 3, 4, 5, 6\}$
  - Subsets such as $\{1, 3, 5\}$ are events.
  - Event space $\{\{2,4,6\}, \{1,3,5\}, \Omega, \emptyset\}$
- ''

σ-algebra

- If $\mathcal{A}$ is a collection of subsets of Ω with the following properties, then $\mathcal{A}$ is called a σ-algebra or σ-field.
  - $\emptyset \in \mathcal{A}$
  - If $A \in \mathcal{A}$, then $A^c \in \mathcal{A}$
  - If $A_1, A_2, A_3, \ldots \in \mathcal{A}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$

- What are the other properties we can derive from the above ones?

Properties of event space

1. $\Omega \in \mathcal{A}$ because $A \cup A^c \in \mathcal{A}$
2. $A \cap B \in \mathcal{A}$ when $A, B \in \mathcal{A}$
   – Because $A \cap B = (A^c \cup B^c)^c$
3. $\bigcap_{i=1}^{\infty} A_i \in \mathcal{A}$, when $A_1, A_2, A_3, \ldots \in \mathcal{A}$
4. $A - B \in \mathcal{A}$ when $A, B \in \mathcal{A}$
   – Because $A - B \triangleq A \cap B^c$ is the set of samples that belong to A but not belong to B

σ-algebra example

- Dice toss
  - $\Omega = \{1, 2, 3, 4, 5, 6\}$
  - $\mathcal{A}=\{\{2, 4, 6\} , \{1, 3, 5\}\}$

Which sets are missing?

  - $\Omega, \emptyset$

Event space & σ-algebra

- We will require event space $\mathcal{A}$ to be a σ-algebra. $\mathcal{A}$ is a set of subsets of Ω that is closed under union and complements
- Meaning of $\mathcal{A}$
  - No matter what the outcome ω of the experiment is, $\omega \in \Omega$, so Ω always occurs and since we know Ω always occurs this is part of our information about the experiment and $\Omega \in \mathcal{A}$ always
  - Since we know Ω always occurs, we know $\emptyset$ never occurs, so this is part of our information about the experiment and $\emptyset \in \mathcal{A}$ always
  - If $A \in \mathcal{A}$, that is, we know that A occurs when $\omega \in A$, we also know that if A occurs A^c does not occur, so this is part of our information about the experiment, therefore, $A^c \in \mathcal{A}$ whenever $A \in \mathcal{A}$
  - Suppose $A_1, A_2, A_3, \ldots \in \mathcal{A}$, i.e., we "observe" A_k, $k = 1, 2, \ldots$ whenever it occurs, then, whenever one A_k, $k = 1, 2, \ldots$ occurs, $\bigcup_{i=1}^{\infty} A_i$ occurs

Examples of $\mathcal{A}$

- I toss a coin twice
  - I tell you whether or not the outcomes of the coin tosses are the same or different. Then
    - $\Omega = \{HH, HT, TH, TT\}$, H ⇒ "Heads", T ⇒ "Tails"
    - $\mathcal{A} = \{\Omega, \emptyset, \{HH, TT\}, \{TH, HT\}\}$
  - I allow you to see the outcome of each coin toss. Then
    - $\Omega = \{HH, HT, TH, TT\}$, H ⇒ "Heads", T ⇒ "Tails"
    - $\mathcal{A} = 2^{\Omega}$ = power set of Ω = set of all subsets of Ω
- Note that the "information" we have about the above two cases is different

Event space & σ-algebra

- We will require event space $\mathcal{A}$ to be a σ-algebra, but not all σ-algebra's are acceptable event spaces.

- Why not make $\mathcal{A}$ the set of all subsets of Ω?
  - Tossing a dice $\mathcal{A}$ ={∅, {1}, {2},…, {6}, {1, 2},…,Ω}
    - fine
  - **does not always work!** → the sample space Ω is continuous or uncountably infinite

Event occurring

- In probability theory, we are interested in **probabilities of events** (not outcomes)
- We say that an event A occurs iff the outcome of the random experiment is $\omega \in \Omega$ belongs to A, i.e., $\omega \in A$

Probability measure P

- **Probability measure** P assigns probability to events in $\mathcal{A}$ and only to events in $\mathcal{A}$
- Example: tossing a dice
  - Sample space $\Omega = \{1, 2, 3, 4, 5, 6\}$
  - Event space $\{\{2, 4, 6\}, \{1, 3, 5\}, \Omega, \emptyset\}$
  - Probability measure
    - $P(\{2, 4, 6\}) = 1/2$
    - $P(\{1, 3, 5\}) = 1/2$
    - $P(\{1, 2\}) = ?$
      - **does not exist.**

Probability measure P

- Why assign probability to sets/events, i.e., elements of $\mathcal{A}$, rather than to outcomes, i.e., elements of Ω?
  - If Ω is finite, there is no problems
  - But if Ω is infinite (uncountable infinite), then there are problems with assigning to every outcome.
- Example: dice toss
  - $\Omega = \{1, 2, 3, 4, 5, 6\}$
  - $\mathcal{A}$= all subsets of Ω
  - Pr(3 occurs) = P({3})

Probability measure P

- Example: wheel spin (continuous outcomes) $\Omega = [0, 2\pi)$
- Suppose we want
  - $\Pr(\text{wheel} = 1.6) = \alpha \ (0 \le \alpha \le 1)$
  - $\Pr(\text{wheel} = 1.7) = \alpha$
  - $\Pr(\text{wheel} = 1.6 \text{ or } 1.7) = 2\alpha$
  - $\Pr(\text{wheel is } 0.001, 0.002, 0.003, \ldots) = (\infty) * \alpha < 1 \Rightarrow \alpha = 0$
  - $\Pr(\pi/2 < \text{wheel} < \pi) = 1/4$

Event space & σ-algebra

- Why not make $\mathcal{A}$ the set of all subsets of Ω?
- Example: Wheel spin (continuous outcomes)
  - $\Omega = [0, 2\pi)$
  - There is no way to assign probabilities to all subsets of Ω in a way that handles frequencies.

Set operations and events

- **Set** operations or relations have an interpretation in terms of **events** in random experiment.
  1. $A \subset B$, the occurrence of $A \Rightarrow$ occurrence of B, B occurs whenever A occurs
  2. $C = A \cup B$: C occurs whenever A or B occurs. (union$\cup\Rightarrow$ "or")
  3. $C = A \cap B$: C occurs if both A and B occur. (intersection $\cap$ $\Rightarrow$"and")
  4. A^c: A does not occur
  5. $A - B$: A occurs but B does not
  6. $C = A \times B$ (two experiments): if C occurs means $(X, Y) \in C = A \times B \Rightarrow X \in A$ and $Y \in B$ (A occurs and B occurs.) (cross product$\times\Rightarrow$"and")

Probability measure

- If $A \in \mathcal{A}$, $\mathrm{P}(A)$ denotes the probability of A
  - $\mathrm{P}(A) \in [0, 1]$
  - P(A) models for the frequency with which the event A occurs when experiments run many times
- Example: Dice toss, $\mathcal{A}$ all possible subsets of Ω

$$\mathrm{P}(A) = \frac{\#\text{ of element in } A}{6}$$

$$\mathrm{P}(\{3\}) = \frac{1}{6}, \qquad \mathrm{P}(\{2,3\}) = \frac{1}{3}$$

Kolmogorov's axioms

- Requirements of probability measure P
  - If $A \in \mathcal{A}$, $P(A) \geq 0$
  - $P(\Omega) = 1$
  - If $A_1, A_2, \ldots$ are pairwise disjoint, i.e., $A_k \cap A_l = \emptyset$, for all $k \neq l$, then $P(A_1 \cup A_2 \ldots) = \sum_{i=1}^{\infty} P(A_i)$

Probability measure: examples

- Throw a dart to a circular dartboard

$$\mathrm{P}(A) = \frac{\text{area of } A}{\text{area of } \Omega}$$

- Dice toss, $\mathcal{A}$ all possible subsets of Ω

$$\mathrm{P}(\{2,4,6\}) = \mathrm{P}(\{2\}) + \mathrm{P}(\{4\}) + \mathrm{P}(\{6\}) = 3 * \frac{1}{6} = \frac{1}{2}$$

Variable name

- We use **capital letters** for **random variables** $X, Y, Z, \ldots, X_1, X_2, X_3, \ldots$
  - X = voltage at some point in a circuit
  - $A \in \mathcal{A}$, $A = \{\text{voltage} \geq 1\} = [1,\infty)$, $\mathrm{P}(A) =?$
  - $\Pr(X \in A)$: probability of the event that the experiment outcome X lies in A.
  - $\Pr(X = 3) = \mathrm{P}(\{3\})$, $\Pr(X \geq 1) = \mathrm{P}([1,\infty))$

- **Different textbook may use different notation:**
  - $\mathbf{x}$ for **random variables,** $\mathbf{x}(\xi)$ indicate the number assigned to the specific outcome ξ
  - $\mathrm{P}\{\mathbf{x} > 1\}$

Axioms and properties of probability

Recall Kolmogorov's axioms

- Requirements of probability measure P (Kolmogorov's axioms)
  1. If A $\in \mathcal{A}$, P(A) $\geq$ 0.
  2. P(Ω) = 1.
  3. If $A_1, A_2, \ldots$ are pairwise disjoint, i.e., $A_k \cap A_l = \emptyset$, whenever $k \neq l$, then $P(A_1 \cup A_2 \cup A_3 \ldots) = \sum_{i=1}^{\infty} P(A_i)$

- Which properties can be derived from these axioms?

Consequences of the axioms

1. Probability of a complement: $\mathrm{P}(A^c) = 1 - \mathrm{P}(A)$
2. Probability of the impossible event: $\mathrm{P}(\emptyset) = 0$
3. Monotonicity: $A \subset B \Rightarrow \mathrm{P}(A) \leq \mathrm{P}(B)$
4. Inclusion-exclusion: $\mathrm{P}(A \cup B) = \mathrm{P}(A)+\mathrm{P}(B)-\mathrm{P}(A \cap B)$
5. Finite disjoint unions: $\mathrm{P}(\bigcup_{i=1}^{N} A_i)=\sum_{i=1}^{N} \mathrm{P}(A_i)$, A_i pairwise disjoint
6. Union bound: $\mathrm{P}(A \cup B)\leq\mathrm{P}(A)+\mathrm{P}(B)$; $\mathrm{P}(\bigcup_{i=1}^{N} A_i)\leq \sum_{i=1}^{N} \mathrm{P}(A_i)$

Proof (1)

- Probability of a complement: $\mathrm{P}(A^c) = 1 - \mathrm{P}(A)$
- Proof: $P(A) + P(A^c) = P(A \cup A^c)$ (Axiom 3)

$$= P(\Omega) = 1 \quad \text{(Axiom 2)}$$

Proof (2)

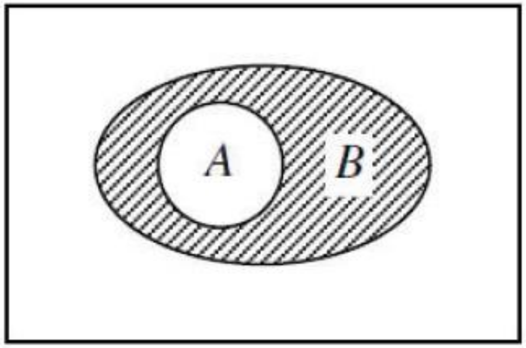
- Probability of the impossible event: $P(\emptyset) = 0$
- Proof: $P(A) + P(\emptyset) = P(A \cup \emptyset)$ (Axiom 3)

$$= P(\mathrm{A})$$

So that $P(\emptyset) = 0$

Proof (3)

- Monotonicity: $A \subset B \Rightarrow \mathrm{P}(A) \leq \mathrm{P}(B)$
- Proof: $A \subset B \Rightarrow B = A \cup (B \cap A^c)$



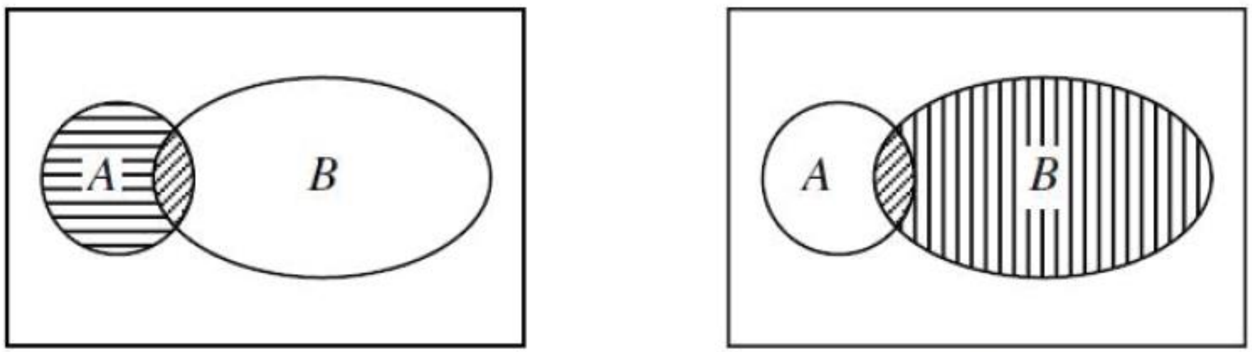
So $P(B) = P(A) + P(B \cap A^c)$, we know that

$P(B \cap A^c) \geq 0$ (Axiom 1)

Thus, $P(A) \leq P(B)$

Proof (4)

- Inclusion-exclusion: $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)$
- Proof:



(a) (b)

$$A = (A \cap B^c) \cup (A \cap B); B = (A \cap B) \cup (A^c \cap B)$$

$$A \cup B = (A \cap B^c) \cup (A \cap B) \cup (A^c \cap B)$$

$$P(A) = P(A \cap B^c) + P(A \cap B) \text{ (Axiom 3)}$$

$$P(B) = P(A \cap B) + P(A^c \cap B) \text{ (Axiom 3)}$$

$$P(A \cup B) = P(A \cap B^c) + P(A \cap B) + P(A^c \cap B)$$

(Axiom 3)

Proof (5) and (6)

- Leave for homework

Limit property

- The following limit properties are essential to answer questions about the probability that something (event) ever happens or never happens
- For any sequence of events A_i

$$P(\cup_{i=1}^{\infty} A_i) = \lim_{N\to\infty} P(\cup_{i=1}^{N} A_i)$$

$$P(\cap_{i=1}^{\infty} A_i) = \lim_{N\to\infty} P(\cap_{i=1}^{N} A_i)$$

Continuity of probability

- **Continuity of probability for monotonic sequences**
- A sequence of events $A_1, A_2, \ldots$ satisfies
$$\lim_{i\to\infty} \mathrm{P}(A_i) = \mathrm{P}(\lim_{i\to\infty} A_i)$$
- if A_i's are increasing $A_1 \subset A_2 \subset \cdots$, in which case
$$\lim_{i\to\infty} \mathrm{P}(A_i) = \mathrm{P}(\cup_{i=1}^{\infty} A_i)$$
- if A_i's are decreasing $A_1 \supset A_2 \supset \cdots$, in which case
$$\lim_{i\to\infty} \mathrm{P}(A_i) = \mathrm{P}(\cap_{i=1}^{\infty} A_i)$$

Continuity of probability: example

- $\lim_{n\to\infty} \mathrm{P}\left(\left(1,2+\frac{1}{n}\right)\right)$
  - $A_n = \left(1,2+\frac{1}{n}\right)$ decreasing
  - $\lim_{n\to\infty} \mathrm{P}\left(\left(1,2+\frac{1}{n}\right)\right) = \mathrm{P}\left(\lim_{n\to\infty}\left(1,2+\frac{1}{n}\right)\right)$
  - $= \mathrm{P}\left(\bigcap_{n=1}^{\infty}\left(1,2+\frac{1}{n}\right)\right) = \mathrm{P}((1,2])$

More about event space and probability measures

- Event space $\mathcal{A}$ is a collection of subsets of the sample space Ω. (set of sets)
- Event space $\mathcal{A}$ must be a σ-algebra
- Probability measure P assigns probability to events in $\mathcal{A}$ and only to events in $\mathcal{A}$
- Requirements of P (Kolmogorov′s axioms)

- Why don't we simplify our lives and always take the event space $\mathcal{A}$ to be the power set 2^{Ω} of Ω?

Event space (1)

- If the sample space Ω is finite or countably infinite, this is always a possibility.
- The probability measure is usually defined as

$$P(A) \triangleq \sum_{\omega \in A} P(\omega), P(\omega) \geq 0 \ \& \sum_{\omega \in \Omega} P(\omega) = 1$$

- It is easy to check P satisfies the axioms of a probability measure.

Event space (2)

- However, it might be **wasteful** in the sense that if we are only interested in a small set of events, then it forces us to assign probability to all subsets rather than just to a **minimal set of interest**

- If we do choose an event space smaller than the power set, it must be a σ-algebra. Why?

Event space (3)

- Because *unions*, *intersections*, *complements of interesting events* are also interesting events and need to be in the event space so they are assigned probabilities.

- Note that the power set itself is a σ-algebra.

Event space (4)

- P(A) assigns probability to events A ∈ $\mathcal{A}$. $\mathcal{A}$ being a σ-algebra, the axioms of probability will be satisfied.
  1. $\Omega \in \mathcal{A}$
     – so it makes sense in axiom 2 to talk about P(Ω).
  2. $A_1, A_2, \ldots \in \mathcal{A} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$
     – so it makes sense in axiom 3 to talk about $P(\bigcup_{i=1}^{\infty} A_i)$

Event space (5)

- If Ω is uncountably infinite, the power set usually "does not work "
- Example: $\Omega = [0, 1]$
  - Surprising, non-intuitive fact:
  - There is no function $P : 2^{\Omega} \to [0, 1]$ such that the following two conditions hold
    - the probability of an interval to be proportional to its length $P([a, b]) = b - a, \forall a, b, 0 \leq a < b \leq 1$
    - the axioms of probability are satisfied
- What to do?

The smallest σ-algebra (1)

- We must choose the event space $\mathcal{A}$ to be a smaller collection of events (smaller than the power set of Ω.)
- It should contain all the events of interest, e.g., intervals
- It should be a σ-algebra

- Let $\mathcal{C}$ be any collection of subsets of Ω. We do not assume $\mathcal{C}$ is a σ-algebra. Define $\sigma(\mathcal{C})$ to be the **smallest σ-algebra that contains** $\mathcal{C}$. By this we mean that if $\mathcal{D}$ is any σ-algebra with $\mathcal{C} \subset \mathcal{D}$ then $\sigma(\mathcal{C}) \subset \mathcal{D}$.

The smallest σ-algebra (2)

- For general collections $\mathcal{C}$ of subsets of Ω, the smallest σ-algebra containing $\mathcal{C}$ is the intersection of all σ-algebra containing $\mathcal{C}$, i.e.,

$$\sigma(\mathcal{C}) = \bigcap_{\mathcal{A}:\mathcal{C}\subset\mathcal{A}} \mathcal{A}$$

- (Proof, textbook 2 Problem 1.45)
- Note that there is at least one σ-algebra containing $\mathcal{C}$, namely the power set.

Borel σ-algebra

- Let $\mathcal{B}$ denote the smallest σ-algebra containing all the open subsets of $R = (-\infty,\infty)$. This collection $\mathcal{B}$ is called the Borel σ-algebra. The sets in $\mathcal{B}$ are called Borel sets
  - All the closed subsets, semi-open subsets and singletons are Borel sets.

- Now, how to choose P on Borel σ-algebra $\mathcal{A}$?

Lebesgue measure

- Lebesgue measure is the standard way of assigning a probability measure to (open, closed, semi-open) intervals

  $$P\big((a,b)\big) = P([a,b]) = P\big((a,b]\big) = P\big([a,b)\big) = b - a$$ where $0 \le a < b \le 1$.

Generalization

- $\Omega = [u, v], -\infty \le u < v \le \infty$(or any form of interval from u to v)
- $\mathcal{A}$ = Borel σ-algebra on Ω

for all intervals, define probability by

$$P((a,b)) = P([a,b]) = P((a,b]) = P([a,b))$$
$$= \int_a^b f(x)\mathrm{d}x$$

where $u \le a < b \le v$ and f is any nonnegative function s.t. $\int_u^v f(x)\mathrm{d}x = 1$

Thank You!