

Probability Theory

Chapter 2—Axioms of Probability

Nan Liu¹

¹These slides are based on Prof. Mui Li's. Any mistakes are my own. If you find an error, please send an email at nanliu@xmu.edu.cn

2.2 Sample Space and Events

- **Definition.** The *sample space* S of an experiment (whose outcome is uncertain) is the set of all possible outcomes of the experiment.
- *Example* (child): Determining the sex of a newborn child in which case $S = \{boy, girl\}$.
- *Example* (horse race): Assume you have an horse race with 12 horses. If the experiment is the order of finish in a race, then

$$S = \{\text{all } 12! \text{ permutations of } (1, 2, 3, \dots, 11, 12)\}.$$

- *Example* (coins): If the experiment consists of flipping two coins, then the sample space is

$$S = \{(H, H), (H, T), (T, H), (T, T)\}.$$

- *Example* (lifetime): If the experiment consists of measuring the lifetime (in years) of your pet then the sample space consists of all nonnegative real numbers: $S = \{x; 0 \leq x < \infty\}$.

2.2 Sample Space and Events

- Any *subset* E of the sample space S is known as an *event*; i.e. an event is a set consisting of possible outcomes of the experiment.
- If the outcome of the experiment is in E , then we say that E has occurred.
- *Example* (child): The event $E = \{\text{boy}\}$ is the event that the child is a boy.
- *Example* (horse race): The event $E = \{\text{all outcomes in } S \text{ starting with a } 7\}$ is the event that the race was won by horse 7.
- *Example* (coins): The event $E = \{(H, T), (T, T)\}$ is the event that a tail appears on the second coin.
- *Example* (lifetime): The event $E = \{x : 3 \leq x \leq 5.5\}$ is the event that your pet will live more than 3 years but won't live more than 5 years and 6 months.

Set Theory: Union of Events

- Given events E and F , $E \cup F$ is the set of all outcomes *either* in E or F or in *both* E and F . $E \cup F$ occurs if *either* E or F occurs. $E \cup F$ is the **union** of events E and F
- Example* (coins): If we have $E = \{(H, T)\}$ and $F = \{(T, H)\}$ then $E \cup F = \{(H, T), (T, H)\}$ is the event that one coin is head and the other is tail.
- Example* (horse race): If we have $E = \{\text{all outcomes in } S \text{ starting with a 7}\}$ and $F = \{\text{all outcomes in } S \text{ finishing with a 3}\}$ then $E \cup F$ is the event that the race was won by horse 7 or/and the last horse was horse 3.
- Example* (lifetime): If $E = \{x : 0 \leq x \leq 5\}$ and $F = \{x : 10 \leq x < \infty\}$ then $E \cup F$ is the event that your pet will die before 5 or will die after 10.
- If $\{E_i\}_{i \geq 1}$ are events then the union is denoted $\bigcup_{i=1}^{\infty} E_i$: it is the event which consists of all the outcomes in $\{E_i\}_{i \geq 1}$.

Set Theory: Intersection of Events

- Given events E and F , $E \cap F$ is the set of all outcomes which are *both* in E and F . $E \cap F$ is also denoted EF .
- *Example* (coins): If we have $E = \{(H, H), (H, T), (T, H)\}$ (event that one H at least occurs) and $F = \{(H, T), (T, H), (T, T)\}$ (event that one T at least occurs) then $E \cap F = \{(H, T), (T, H)\}$ is the event that one H and one T occur.
- *Example* (horse race): If we have $E = \{\text{all outcomes in } S \text{ starting with a 7}\}$ and $F = \{\text{all outcomes in } S \text{ starting with a 8}\}$ then $E \cap F$ does not contain any outcome and is denoted by $\emptyset$.
- *Example* (lifetime): If we have $E = \{x : 0 \leq x \leq 5\}$ and $F = \{x : 3 \leq x < 7\}$ then $E \cap F = \{x : 3 \leq x \leq 5\}$ is the event that your pet will die between 3 and 5.
- If $\{E_i\}_{i \geq 1}$ are events then the intersection is denoted $\bigcap_{i=1}^{\infty} E_i$: it is the event which consists of the outcomes which are in all of the events $\{E_i\}_{i \geq 1}$.

Set Theory: More Notations

- For any event E , E^c denote the *complement* set of all outcomes in S which are not in E . Hence we have $E \cup E^c = S$ and $E \cap E^c = \emptyset$.
- For any two events E and F , we write $E \subset F$ is all the outcomes of E are in F .

Set Theory: Some “Algebra”

- *Commutative laws*

$$E \cup F = F \cup E \text{ and } E \cap F = F \cap E.$$

- *Associative laws*

$$\begin{aligned}(E \cup F) \cup G &= E \cup (F \cup G), \\ (E \cap F) \cap G &= E \cap (F \cap G).\end{aligned}$$

- *Distributive laws*

$$\begin{aligned}(E \cup F) \cap G &= (E \cap G) \cup (F \cap G), \\ (E \cap F) \cup G &= (E \cup G) \cap (F \cup G).\end{aligned}$$

Set Theory: DeMorgan's Laws

- We have

$$(\cup_{i=1}^n E_i)^c = \cap_{i=1}^n E_i^c$$

$$(\cap_{i=1}^n E_i)^c = \cup_{i=1}^n E_i^c$$

- *Proof.* Suppose x is an outcome of $(\cup_{i=1}^n E_i)^c$, then x is not in $(\cup_{i=1}^n E_i)$. Thus it is not in any of the event E_i , $i = 1, \dots, n$. Hence it is in E_i^c for all $i = 1, \dots, n$. Hence this proves $(\cup_{i=1}^n E_i)^c \subset \cap_{i=1}^n E_i^c$. Suppose x is an outcome of $\cap_{i=1}^n E_i^c$, hence it is in E_i^c for all $i = 1, \dots, n$. Hence it is in none of the event E_i , $i = 1, \dots, n$. Thus it is $(\cup_{i=1}^n E_i)^c$ and we have proven that $\cap_{i=1}^n E_i^c \subset (\cup_{i=1}^n E_i)^c$. The result $(\cup_{i=1}^n E_i)^c = \cap_{i=1}^n E_i^c$ follows.
- To prove $(\cap_{i=1}^n E_i)^c = \cup_{i=1}^n E_i^c$, we remark that

$$\underbrace{(\cup_{i=1}^n E_i^c)^c}_{\text{previous result applied to events } \{E_i^c\}} = \cap_{i=1}^n (E_i^c)^c = \cap_{i=1}^n E_i \text{ as } (E_i^c)^c = E_i.$$

Thus by taking the complement on both sides, we obtain the result.

2.3 Axioms of Probability

- Consider an experiment with sample space S . For each event E , we assume that a number $P(E)$, the *probability* of the event E , is defined and satisfies the following 3 axioms.

- Axiom 1**

$$0 \leq P(E) \leq 1$$

- Axiom 2**

$$P(S) = 1$$

- Axiom 3.** For any sequence of mutually exclusive events $\{E_i\}_{i \geq 1}$, i.e. $E_i \cap E_j = \emptyset$ when $i \neq j$, then

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$$

- Direct consequences include $P(\emptyset) = 0$ and for mutually exclusive events $\{E_i\}_{i \geq 1}$

$$P\left(\bigcup_{i=1}^n E_i\right) = \sum_{i=1}^n P(E_i).$$

Examples

- *Example (coins):* Assume both coins are unbiased; i.e. a head is as likely to appear as a tail, then

$$P(\{H, H\}) = P(\{H, T\}) = P(\{T, H\}) = P(\{T, T\}) = \frac{1}{4}.$$

If the first coin is unbiased but the second coin is such that $P(H) = 1 - P(T) = 2/3$ then

$$\begin{aligned}P(\{H, H\}) &= P(\{T, H\}) = \frac{1}{2} \frac{2}{3} = \frac{1}{3}, \\P(\{H, T\}) &= P(\{T, T\}) = \frac{1}{2} \frac{1}{3} = \frac{1}{6}.\end{aligned}$$

- *Example:* A die is rolled and we assume $P(\{1\}) = P(\{2\}) = \dots = P(\{6\}) = 1/6$. Hence as a consequence from axiom 3, the probability of having an even or odd number is equal to

$$\begin{aligned}P(\{1, 3, 5\}) &= P(\{1\}) + P(\{3\}) + P(\{5\}) = 1/2, \\P(\{2, 4, 6\}) &= P(\{2\}) + P(\{4\}) + P(\{6\}) = 1/2.\end{aligned}$$

2.4 Some Simple Propositions

- **Proposition:** $P(E^c) = 1 - P(E)$.
- We have $S = E \cup E^c$ and $E \cap E^c = \emptyset$ so

$$\underbrace{P(S) = 1}_{\text{axiom 2}} = \underbrace{P(E \cup E^c) = P(E) + P(E^c)}_{\text{axiom 3}}.$$

- **Proposition:** If $E \subset F$ then $P(E) \leq P(F)$.
- We have $F = E \cup (E^c \cap F)$ and $E \cap ((E^c \cap F)) = \emptyset$ so

$$P(F) = P(E) + \underbrace{P(E^c \cap F)}_{\geq 0 \text{ by axiom 1}} \geq P(E).$$

- **Proposition:** We have $P(E \cup F) = P(E) + P(F) - P(E \cap F)$.
- We have $E \cup F = E \cup (E^c \cap F)$ and $E \cap ((E^c \cap F)) = \emptyset$ so

$$P(E \cup F) = P(E \cup (E^c \cap F)) = P(E) + P(E^c \cap F)$$

but $F = (E^c \cap F) \cup (E \cap F)$ with $(E^c \cap F) \cap (E \cap F) = \emptyset$ so

$$P(F) = P(E^c \cap F) + P(E \cap F) \Rightarrow P(E^c \cap F) = P(F) - P(E \cap F).$$

Example

- You are in a restaurant and ordering 2 dishes. With proba 0.6, you will like the first dish; with proba 0.4, you will like the second dish. With proba 0.3, you will like both of them. What is the proba, you will like neither dish?
- Let A_i the event: "You like dish i ". Then the proba you like at least one is

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) = 0.6 + 0.4 - 0.3 = 0.7.$$

- The event that you like neither dish is the complement of liking at least one, so

$$\begin{aligned} P(\text{"you will like neither dish"}) &= P((A_1 \cup A_2)^c) \\ &= 1 - P(A_1 \cup A_2) \\ &= 0.3 \end{aligned}$$

2.4 Some Simple Propositions: Inclusion-Exclusion Identity

- For events $E_1, E_2, \dots, E_n$,

$$\begin{aligned} P\left(\bigcup_{i=1}^n E_i\right) &= \sum_{i=1}^n P(E_i) - \sum_{i < j} P(E_i E_j) \\ &\quad + \sum_{i < j < k} P(E_i E_j E_k) + \dots + (-1)^{n+1} P(E_1 \dots E_n). \end{aligned}$$

- Implication: matching problem.

2.5 Sample Spaces having Equally Likely Outcomes

- Assume $S = \{1, 2, \dots, N\}$ then it is often natural to assume $P(\{i\}) = 1/N$ so, for any event E ,

$$P(E) = \frac{\# \text{ outcomes in } E}{\# \text{ outcomes in } S}.$$

- Example:** If 5 balls are “randomly drawn” from a bowl containing 10 white and 7 black balls, what is the probability that 3 of the balls are white and the 2 other black?
- Answer:** There are $\binom{17}{5}$ possible outcomes of the experiment which are assumed equally likely. We have $\binom{10}{3}$ ways to select 3 white balls among 10 and $\binom{7}{2}$ ways to select 2 black balls among 7. So the probability is given by

$$\binom{10}{3} \binom{7}{2} / \binom{17}{5} = 0.4072$$

Example: Chevalier de Méré, Fermat and Pascal

- **Example:** Chevalier de Méré, who was a big gambler, realized empirically that he could earn money by betting that, if you throw a dice 4 times, at least one “6” would appear. What is the probability of this event?
- **Answer:** The complement event is to never obtain a “6” when you throw the dice 4 times, this has a probability $\left(\frac{5}{6}\right)^4$. Hence the probability of interest is given by $1 - \left(\frac{5}{6}\right)^4 \approx 0.5177$.
- **Example:** Chevalier de Méré decided to extend this “trick” to two dices and conjectured that he would still make a money by betting that, if you throw two dices 24 times, at least one double “6” would appear. He started losing money and asked Fermat and Pascal to explain him why.
- **Answer:** The complement event is to never obtain a double “6” when you throw the dice 24 times, it has a probability $\left(\frac{35}{36}\right)^{24}$. Hence the probability of interest is given by $1 - \left(\frac{35}{36}\right)^{24} \approx 0.4914$.

Example: Wall Street Journal

- “The renowned Harvard scholar Graham Allison has posited that there is greater than a 50% likelihood of a nuclear terrorist attack in the next decade, which he says could kill upward of 500,000 people. If we accept Mr. Allison’s estimates—a 5% chance per year of a 500,000-fatality event in a Western country (25,000 casualties per year)—the risk from such incidents is some 150 times greater than that from conventional terrorist attacks.” , Nate Silver, 11th January 2010.
- Similar type of mistake as Chevalier de Méré: a 5% chance a year does NOT imply a 50% chance over 10 years. This would suggest that in 20 years, we have a $5 \times 20 = 100\%$ chance of having a nuclear attack.
- Assuming the probability of at least one attack over the next 10 years is 0.5, the probability P of an attack per year (assuming independence between years!) is

$$(1 - P)^{10} = 0.5 \Rightarrow P = 1 - \exp(\ln(0.5)/10) = 0.0670 > 0.05.$$

Example: Portable Music Player

- **Example:** You are having 4,000 songs on your portable music player and are using an option that picks the songs randomly (after each song, it picks a new song, each with proba $1/4000$). You then realize that the 75th song played by your player is the same as the 30th (or the 12th or whatever) you have listened to. Should you complain to the manufacturer?
- **Answer:** If the songs are picked randomly, then the probability of not listening twice to the same song among the 75 first songs is

$$\frac{4000.3999 \dots (4000 - 75 + 1)}{4000^{75}} \approx 0.4975.$$

- Your portable music player works just fine.

Example: Birthday Problem

- **Example:** If n people are present in a room, what is the probability that at least two of them celebrate their birthday on the same day of the year? (We assume that there are 365 days and proba of being born a given day is $1/365$).
- **Answer:** The complementary event is that no two of them celebrate their birthday on the same day which is given by

$$\bar{P}_n = \frac{365 \cdot 364 \cdot \dots \cdot (365 - n + 1)}{365^n}$$

so the probability is $P_n = 1 - \bar{P}_n$.

- For $n = 23$, we have $P_n \geq 0.5$ and for $n = 70$, we have $P_n \approx 0.9992$.
- This might appear surprising but there are $\binom{n}{2}$ possible pairs of people. For $n = 23$, $\binom{23}{2} = 253$ and for $n = 70$, $\binom{70}{2} = 2415$.

2.6 Probability as a Continuous Set Function

- A sequence of events $\{E_n\}_{n \geq 1}$ is said to be increasing if

$$E_1 \subset E_2 \subset \cdots \subset E_n \subset E_{n+1} \subset \cdots$$

and we define a new event, denoted $\lim_{n \rightarrow \infty} E_n$, by

$$\lim_{n \rightarrow \infty} E_n = \bigcup_{i=1}^{\infty} E_i$$

- Similarly, $\{E_n\}_{n \geq 1}$ is said to be decreasing if

$$E_1 \supset E_2 \supset \cdots \supset E_n \supset E_{n+1} \supset \cdots$$

and we define a new event, denoted $\lim_{n \rightarrow \infty} E_n$, by

$$\lim_{n \rightarrow \infty} E_n = \bigcap_{i=1}^{\infty} E_i$$

- **Proposition:** If $\{E_n\}_{n \geq 1}$ is either an increasing or decreasing sequence of events, then

$$\lim_{n \rightarrow \infty} P(E_n) = P\left(\lim_{n \rightarrow \infty} E_n\right).$$

- We prove it for a sequence of increasing events. The idea is to use Axiom 3 to establish the result.
- We have as $\{E_n\}_{n \geq 1}$ is increasing

$$\cup_{i=1}^n E_i = E_n$$

and we have the following “donuts” decomposition

$$E_n = \cup_{i=1}^n F_i$$

with

$$F_1 = E_1 \text{ and } F_n = E_n \cap E_{n-1}^c.$$

- It is easy to check that

$$F_i \cap F_j = \emptyset \text{ and } \cup_{i=1}^n E_i = \cup_{i=1}^n F_i \text{ for all } n \geq 1$$

- We have

$$\begin{aligned}P\left(\bigcup_{i=1}^{\infty} E_i\right) &= P\left(\bigcup_{i=1}^{\infty} F_i\right) = \sum_{i=1}^{\infty} P\left(F_i\right) \text{ (axiom 3)} \\&= \lim_{n \rightarrow \infty} \sum_{i=1}^n P\left(F_i\right) \\&= \lim_{n \rightarrow \infty} P\left(\underbrace{\bigcup_{i=1}^n F_i}_{= \bigcup_{i=1}^n E_i}\right) \\&= \lim_{n \rightarrow \infty} P\left(E_n\right)\end{aligned}$$

- For decreasing events $\{E_n\}_{n \geq 1}$, use the fact that $\{E_n^c\}_{n \geq 1}$ is increasing.

Example: A Paradox

- You possess an infinitely large urn and an infinite collection of balls labelled number 1,2,3 etc. At time 0, balls 1 to 10 are placed in the urn and ball 10 withdrawn. After $1/2$ minute, balls 11 to 20 are placed in the urn and ball 20 withdrawn. After one more $1/4$ minute, balls 21 to 30 are placed in the urn and ball 30 withdrawn etc. How many balls do we have in the urn after one minute?
- We clearly have an infinite number as, at time

$$\frac{1}{2} + \dots + \frac{1}{2^n}$$

we have $9(n+1)$ balls after the $n+1$ th withdrawal; i.e. balls 1 to 9, 11 to 19, ..., $10n+1$ to $10n+9$.

- We now change the experiment so that at time 0, balls 1 to 10 are placed in the urn and ball 1 is withdrawn. After $1/2$ minute, balls 11 to 20 are placed in the urn and ball 2 withdrawn. After $1/4$ more minute, balls 21 to 30 are placed in the urn and ball 3 withdrawn etc. How many balls do we have in the urn after one minute?

Example: A Paradox (cont.)

- The urn is empty as ball number n is removed at time 0 for $n = 1$ and time

$$\frac{1}{2} + \dots + \frac{1}{2^{n-1}} < 1 \text{ for } n > 1.$$

- Consider now the case where, whenever a ball is to be withdrawn, that ball is randomly selected from among those present; i.e. at time 0 balls 1 to 10 are placed in the urn and one is randomly selected and withdrawn, and so on. How many balls will you have in the urn after one minute?
- Consider ball number 1. Define E_n the event that this ball is still in the urn after n withdrawals. We have

$$P(E_n) = \frac{9 \cdot 18 \cdot 27 \cdot \dots \cdot 9n}{10 \cdot 19 \cdot 28 \cdot \dots \cdot (9n + 1)} = \prod_{i=1}^n \frac{9i}{9i + 1}$$

- Being in the urn at time 1 is the event $\cap_{i=1}^{\infty} E_i$ and $\{E_i\}_{i \geq 1}$ is decreasing so $P(\cap_{i=1}^{\infty} E_i) = \lim_{n \rightarrow \infty} P(E_n)$.

Example: A Paradox (cont.)

- We have

$$\frac{1}{P(E_n)} = \prod_{i=1}^n \left(1 + \frac{1}{9i}\right) \geq \frac{1}{9} + \frac{1}{18} + \cdots + \frac{1}{9n} \xrightarrow{n \rightarrow \infty} \infty$$

so

$$P(\cap_{i=1}^{\infty} E_i) = 0.$$

- This reasoning can be extended to any ball number k . Define $p(k)$ the positive integer such that $k = 10p(k) + d(k)$ where $0 \leq d(k) \leq 9$ and define E_n the event that ball k is still in the urn after n ($n \geq p(k) + 1$) withdrawals then similarly

$$P(E_n) = \prod_{i=p(k)+1}^n \frac{9i}{9i+1} \xrightarrow{n \rightarrow \infty} 0$$

- Hence, the urn will be empty at time 1.