STAT 390 A
Statistical Methods in Engineering and Science
Week 1 Lectures – Part 2 – Spring 2023
Axiomatic Introduction to Probability Theory

Alexander Giessing

Department of Statistics

University of Washington

April 3, 2023

Outline

Fundamental Concepts

Experiment

An experiment is any activity or process whose outcome is subject to uncertainty.

- rolling dice, Bitcoin price next year, tomorrow's weather, genetic variation in a population, ...

Sample Space

The sample space of an experiment is the set of all possible outcomes of the experiment. We denote the sample space by Ω.

- throwing one die once: $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- tossing one coin twice: $\Omega = \{HH, HT, TH, TT\}$.
- Bitcoin price next year: $\Omega = [0, \infty)$.

Fundamental Concepts

Event

Events are subsets of the sample space Ω.

(examples continued from above)

- rolling an even number with a die: $A = \{2, 4, 6\}$.
- tossing heads at least once: $B = \{HH, HT, TH\}$.
- Bitcoin price next year: $C = [50k, 100k]$.

- Let $\omega \in \Omega$ be the outcome of an experiment. We say that the **event** A **occurs** if $\omega \in A$.
  - die lands on 2, i.e. $\omega = 2$
    $\implies$ event A occurs.
  - coin shows tail twice, i.e. $\omega = TT$
    $\implies$ event B does not occur.
  - Bitcoin price next year exceeds 300, i.e. $\omega = (300k, \infty)$
    $\implies$ event C does not occur.

Set Operations and Interpretation

Let $A, B \subset \Omega$ be events.

- **Intersection:** $A \cap B = \{\omega \in \Omega : \omega \in A \text{ and } \omega \in B\}$
  Event $A \cap B$ occurs. $\Leftrightarrow$ Both A and B occur.

- **Union:** $A \cup B = \{\omega \in \Omega : \omega \in A \text{ or } \omega \in B\}$
  Event $A \cup B$ occurs. $\Leftrightarrow$ At least one of the events A and B occurs.

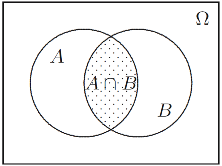
- **Complementation:** $A^c = \{\omega \in \Omega : \omega \notin A\}$
  Event A^c occurs. $\Leftrightarrow$ Event A does not occur.

- **Implication:** $\omega \in A \implies \omega \in B$ (equivalent, $A \subset B$)
  If event A occurs, then so does event B.

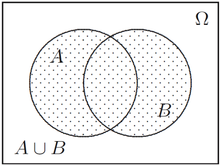
- **Mutually exclusiveness:** $A \cap B = \emptyset$ ($\emptyset =$ "empty set" with no element)
  At most one of the events A and B occurs.

Set Operations and Interpretation

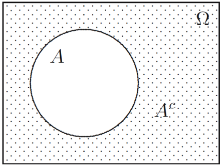
(Venn diagram illustration of the preceding set operations)



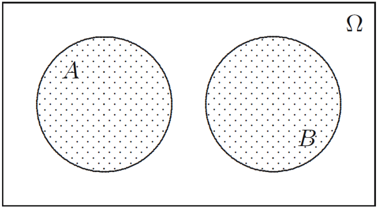
Intersection $A \cap B$



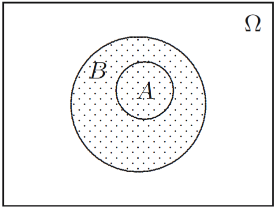
Union $A \cup B$



Complement A^c



Disjoint sets A and B



A subset of B

Set Operations and Interpretation

DeMorgan's Law

For any events A and B we have

$$(A \cup B)^c = A^c \cap B^c, \qquad \text{and} \qquad (A \cap B)^c = A^c \cup B^c.$$

- These two identities can be iterated: For events $A_1, \ldots, A_m$,

$$(A_1 \cup A_2 \cup \ldots \cup A_m)^c = A_1^c \cap A_2^c \cap \ldots \cap A_m^c,$$
$$(A_1 \cap A_2 \cap \ldots \cap A_m)^c = A_1^c \cup A_2^c \cup \ldots \cup A_m^c.$$

Example: Who is to blame?

"It is certainly not true that neither John nor Mary is to blame."

- Define events

$$J = \text{"John is to blame."} \qquad \text{and} \qquad M = \text{"Mary is to blame."}$$

- We can now write above statement succinctly as

$$(J^c \cap M^c)^c \,.$$

- But by DeMorgan's Law,

$$(J^c \cap M^c)^c = (J^c)^c \cup (M^c)^c = J \cup M.$$

- We conclude that

"Either John or Mary is to blame, or both!"

Example with R Code: Titanic Survivors

Let Ω be the sample space of all passengers aboard the Titanic. Based on the data set "titanic.csv" (a subset of all passengers), we extract some interesting events/ subsets.

- The data set contains the following variables:
  - **pclass** – passenger class (1 = 1st; 2 = 2nd; 3 = 3rd)
  - **survived** – survival (0 = No; 1 = Yes)
  - **sex** – male, female
  - **age** – age in years
  - **sibsp** – number of siblings/spouses aboard
  - **parch** – number of parents/children aboard
  - **fare** – passenger fare in USD
  - **embarked** – port of embarkation (C = Cherbourg; Q = Queenstown; S = Southampton)

```
> titanic <- read.csv("titanic.csv", header=T)
> titanic[1:3,]
pclass survived    sex     age sibsp parch     fare embarked
1       1        1 female 29.0000     0     0 211.3375        S
2       1        1   male  0.9167     1     2 151.5500        S
3       1        0 female  2.0000     1     2 151.5500        S
```

Example with R Code: Titanic Survivors

- The age of male and female passengers can be extracted as follows:

```
> mAge <- titanic$age[titanic$sex=="male"]
> fAge <- titanic$age[titanic$sex=="female"]
> summary(fAge)
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.1667 19.0000 27.0000 28.5766 37.7500 76.0000
```

- The age of all female passengers who embarked in Cherbourg:

```
> fCAge<-titanic$age[(titanic$embarked=="C")&(titanic$age=="female")]
> summary(CfAge)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.75   19.00   30.00   31.22   44.75   64.00
```

- Observation: Female passengers embarking in Cherbourg were on average about 3 years older than the typical female passenger.

- Fore more logical operators to extract subsets see STAT390-R-Intro-2.pdf on Canvas!

Outline

Probability

- Probability is a measure of the likelihood with which an event occurs.
- Two (philosophical) viewpoints on probability:
  - **Frequentist:** probability = "long-run relative freq." (appealing for repeatable experiments, e.g. dice rolls, coin tosses, ...).
  - **Bayesian:** probability = "measure of prior belief" (appealing for non-repeatable experiments, e.g. tomorrow's weather, ...).
- Mathematical study of probability:
  - originated from analyses of games of chance (dice, cards, roulette, ...) by Blaise Pascal (1623–1662) and Pierre de Fermat (1607–1665).
  - formalized as a distinct branch of mathematics by Andrey Kolmogorov (1903–1987).

Probability

PROBABILITY MEASURE (ON A FINITE SAMPLE SPACE)

Let Ω be a finite sample space and $\mathcal{A}$ be the collection of all subsets of Ω. A probability measure P on $(\Omega, \mathcal{A})$ is a function P with domain $\mathcal{A}$ that satisfies

(i) $P(A) \geq 0$ for all $A \in \mathcal{A}$;

(ii) $P(\Omega) = 1$;

(iii) $P(A \cup B) = P(A) + P(B)$ for all pairwise disjoint $A, B \in \mathcal{A}$.

The number $P(A)$ is called the probability that event A occurs.

- formalizes our intuitive understanding of a “measure”:
  - **non-negativity:** negative distances, areas, weights, volumes, etc. don’t make sense.
  - **normalization:** we measure in meters, miles, square feet, gallons, etc.
  - **additivity:** we obtain the same volume whether we measure the volume of a ball directly or cut it into two halves and measure each half separately.

A Few Properties of Probability Measures

1. $P(A^c) = 1 - P(A)$.

2. $P(\emptyset) = 0$.

3. If $A \subset B$, then $P(A) \leq P(B)$.

4. If A and B are mutually exclusive, then $P(A \cap B) = 0$.

5. If $A_1, \ldots, A_m$ are mutually exclusive[1], i.e. $A_j \cap A_k = \emptyset$ for all $1 \leq j, k \leq m$,

$$P(A_1 \cup A_2 \cup \ldots \cup A_m) = P(A_1) + P(A_2) + \ldots + P(A_m).$$

6. For any events A and B,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

[1] equivalent "pairwise disjoint"

A Few Properties of Probability Measures[2]

1. *Proof:* $1 \overset{(a)}{=} P(\Omega) = P(A \cup A^c) \overset{(b)}{=} P(A) + P(A^c)$, where (a) holds by (ii) and (b) by (iii).
2. *Proof:* Note that $\emptyset = \Omega^c$. Thus, from above, $P(\emptyset) = 1 - P(\Omega) \overset{(a)}{=} 1 - 1 = 0$, where (a) holds by (ii).
3. *Proof:* Note that $B = A \cup (B \cap A^c)$ and A and $B \cap A^c$ are mutually exclusive. Hence, by (iii) and (i), $P(B) = P(A) + P(B \cap A^c) \geq P(A)$.
4. *Proof:* $P(A \cap B) = P(\emptyset) = 0$.
5. *Proof:* Note that $A_1 \cap (A_2 \ldots \cup A_m) = \emptyset$. Thus, by (iii), $P(A_1 \cup A_2 \cup \ldots \cup A_m)$ $= P(A_2 \cup \ldots \cup A_m) + P(A_1)$. Since $A_2 \cap (A_3 \cap \ldots A_m) = \emptyset$, we have, by (iii) again, $P(A_2 \cup \ldots \cup A_m) = P(A_3 \cup \ldots \cup A_m) + P(A_2)$. Iterate, to conclude.
6. *Proof:* Note that $A = (A \cap B) \cup (A \cap B^c)$, where $A \cap B$ and $A \cap B^c$ are mutually exclusive. Hence, by (iii), $P(A) = P(A \cap B) + P(A \cap B^c)$ (*). Similarly, $A \cup B = \big((A \cup B) \cap B\big) \cup \big((A \cup B) \cap B^c\big) = B \cup (A \cap B^c)$, where B and $A \cap B^c$ are mutually exclusive. Thus, by (iii), $P(A \cup B) = P(A \cap B^c) + P(B)$ (**). Combine (*) and (**) to conclude.

---

[2] I expect you to learn and apply these properties, the proofs are optional.

Example: A Fair Coin Toss

Suppose we toss a fair coin one time. Find the probability triple $(\Omega, \mathcal{A}, P)$ *which describes this experiment!*

- all possible outcomes: $\Omega = \{H, T\}$.
- all possible subsets of Ω: $\mathcal{A} = \{\emptyset, \{H\}, \{T\}, \{H, T\}\}$.
- Since the coin is fair, heads and tail are equally likely. Thus, we assign

$$P(\{H\}) = P(\{T\}) = \frac{1}{2}, \quad P(\emptyset) = 0, \quad P(\{H, T\}) = 1.$$

  By definition, $P(\{H, T\}) = P(\Omega) = 1$ and $P(\emptyset) = 0$ (see above). Thus, we do not need to explicitly assign the last two probabilities.

- Formally, we have to write $\{H\}$ for the set consisting of the single element H, because probability measures are defined on events, not on outcomes. From now on we shall drop the brackets.

Example: A Simple Lottery

Suppose we buy a single ticket in a lottery with 10,000 tickets and only one prize. We think of this experiment as having two possible outcomes, "success" (we win the prize) and "failure" (we don't win the prize). Find the probability triple $(\Omega, \mathcal{A}, P)$ *which describes this experiment!*

- all possible outcomes: $\Omega = \{S, F\}$.
- all possible subsets of Ω: $\mathcal{A} = \{\emptyset, \{S\}, \{F\}, \Omega\}$.
- the probability measure P on $(\Omega, \mathcal{A})$ is completely defined by

$$P(S) = \frac{1}{10,000} \quad \text{and} \quad P(F) = \frac{9,999}{10,000}.$$

Example: Tossing a Fair Coin Twice

Suppose we toss a fair coin twice. Find the probability triple $(\Omega, \mathcal{A}, P)$ *which describes this experiment!*

- all possible outcomes: $\Omega = \{HH, HT, TH, TT\}$.
- all possible subsets of Ω:

$$\begin{aligned}\mathcal{A} = \big\{&\emptyset,\ \Omega,\ \{HH\},\ \{HT\},\ \{TH\},\ \{TT\},\\ &\{HH, HT\},\ \{HH, TH\},\ \{HH, TT\},\\ &\{HT, TH\}\ \{HT, TT\},\ \{TH, TT\},\\ &\{HH, HT, TH\},\ \{HH, HT, TT\},\\ &\{HT, TH, TT\},\ \{HH, TH, TT\}\big\}.\end{aligned}$$

  How can we assign probabilities to so many events? Not easy!

- Throwing one die once, results in $2^6 = 64$ possible events; throwing one die twice results in $2^{36} = 68,719,476,736$ possible events...
- It is impractical to explicitly assign probabilities to all possible events.

Outline

Equally Likely Outcomes

Simple Event

Simple events are subsets of the sample space that contain only one outcome.

- all simple events associated with throwing one die once: $\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}$.

Experiment with Equally Likely Outcomes

Suppose that all outcomes of an experiment with finite sample space Ω are equally likely. Then, for all events $A \subset \Omega$,

$$P(A) = \frac{\#\,\text{simple events in } A}{\#\,\text{simple events in } \Omega}. \qquad (\star)$$

- If all outcomes are equally likely, assigning probabilities reduces to counting simple events.

Equally Likely Outcomes[3]

Proof of formula in $(\star)$. To simplify notation, define

$$N := \#\,\text{simple events in } \Omega,$$
$$N_A := \#\,\text{simple events in } A.$$

Denote the simple events in Ω by $E_1, \ldots, E_N$. Since all outcomes are equally likely, so are the simple events, and, hence,

$$P(E_1) = P(E_2) = \ldots = P(E_N) = \frac{1}{N}.$$

Therefore,

$$P(A) = P\left(\bigcup_{E_i \in A}\right) \overset{(a)}{=} \sum_{E_i \in A} P(E_i) = \sum_{E_i \in A} \frac{1}{N} = \frac{N_A}{N},$$

where (a) follows from the additivity property of probability measures for disjoint events.

[3] I expect you to learn and apply this result, the proof is optional.

Example: Tossing a Fair Coin Twice (Cont.)

Suppose we toss a fair coin twice. Find the probability of tossing heads at least once!

- all possible outcomes $\Omega = \{HH, HT, TH, TT\}$.

  $\Longrightarrow$ Ω consists of 4 simple events.

- $A = \{\text{"at least one head"}\} = \{HH, HT, TH\}$.

  $\Longrightarrow$ A consists of 3 simple events.

- Hence, since all outcomes are equally likely,

$$P\left(\text{"at least one head"}\right) = \frac{3}{4}.$$

Example: Tossing Five Fair Coins

A machine tosses 5 coins at random. What is the probability of at least 4 heads?

- all possible outcomes:

$$\Omega = \{HHHHH, THHHH, HTHHH, HHTHH, HHHTH, \\ HHHHT, TTHHH \ldots\}.$$

$\implies 2^5 = 32$ simple events.

- $\{\text{“at least 4 heads”}\} = \underbrace{\{\text{“exactly 4 heads”}\} \cup \{\text{“exactly 5 heads”}\}}_{\text{disjoint events}}.$

- $\{\text{“exactly 4 heads”}\} \implies 5$ simple events.

  $\{\text{“exactly 5 heads”}\} \implies 1$ simple event.

- Thus, by the additivity property of probabilities for disjoint events,

$$\begin{aligned} P(\text{“at least 4 heads”}) &= P(\text{“exactly 4 heads”}) + P(\text{“exactly 5 heads”}) \\ &= \frac{5}{32} + \frac{1}{32} = \frac{3}{16}. \end{aligned}$$

Example: Rolling Two Dice

You roll two dice. How many simple events (or, possible outcomes) are there? What is the probability that the sum of the two dice is greater than 7?

- number of simple events: $6^2 = 36$.
- split the event {"sum greater than 7"} into 5 disjoint events, i.e.

$$\begin{aligned}\{\text{"sum greater than 7"}\} = &\{\text{"sum equals 8"}\} \cup \{\text{"sum equals 9"}\} \\ &\cup \{\text{"sum equals 10"}\} \cup \{\text{"sum equals 11"}\} \\ &\cup \{\text{"sum equals 12"}\}.\end{aligned}$$

- {"sum equals 8"} $\Longrightarrow$ 5 simple events.

  {"sum equals 9"} $\Longrightarrow$ 4 simple events.

  {"sum equals 10"} $\Longrightarrow$ 3 simple events.

  {"sum equals 11"} $\Longrightarrow$ 2 simple events.

  {"sum equals 12"} $\Longrightarrow$ 1 simple events.

$$\Longrightarrow \; P\left(\text{"sum greater than 7"}\right) = \frac{5+4+3+2+1}{32} = \frac{5}{12}.$$

Counting Techniques

Permutations

An ordered sequence of k objects taken from a set of n distinct objects is called a permutation (or arrangement). The total number of permutations is

$$P_{n,k} := n(n-1)\cdots(n-k+1) = \frac{n!}{(n-k)!},$$

where $m! := m(m-1)\cdots 1$ is called "m factorial".

Combinations

An unordered sequence of k objects taken from a set of n distinct objects is called a combination. The total number of combinations is

$$C_{n,k} := \binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Example: Permutation or Combination?

Consider a group of 120 hobby marathon runners. In how many ways can the first, second, and third prize be awarded to the runners?

- The ordering matters: Jane, Richard and Caroline are different from Richard, Caroline and Jane for the first, second, and third prize.
- There are 120 runners to choose from for first prize.
- After choosing a runner for first place, there are 119 runners left to choose from for second prize.
- After choosing two runners for first and second place, there are 118 runners left to choose from for third prize.
- Hence, there exist

$$120 \times 119 \times 118 = \frac{120!}{117!} = 1,685,040$$

  permutations for awarding first, second, and third prize to 120 runners.

Example: Permutation or Combination? (Cont.)

Consider the same group of 120 hobby marathon runners. Now, they want to form a committee to organize the next race. In how many ways can three runners be selected to form a committee?

- The ordering does not matter: Jane, Richard, and Caroline constitute the same committee as Richard, Caroline, and Jane.
- For every committee there exist $3! = 6$ equivalent permutations (or arrangements).
- Thus, there exist

$$\frac{120 \times 119 \times 118}{6} = \frac{120!}{3!117!} = 280,840$$

  combinations for different committees of 3 out of 120 runners.

Example: Probabilities of Card Hands

Three cards are selected at random from a 52-card deck.

- What is the probability of a hand of King ♠, Queen ♣, and Jack ♣?

$$\frac{1}{\binom{52}{3}} = 0.00004524886.$$

- What is the probability of getting a hand of King, Queen, and Jack?

$$\frac{\binom{4}{1}\binom{4}{1}\binom{4}{1}}{\binom{52}{3}} = 0.0028959276.$$

- What is the probability of getting a hand of King, Queen, and Jack with three distinct suits?

$$\frac{\binom{4}{1}\binom{3}{1}\binom{2}{1}}{\binom{52}{3}} = 0.00108597285.$$

Example: Probabilities of Card Hands (Cont.)

Three cards are drawn from a 52-card deck and their order is noted.

- How many possible outcomes are there?

$$52 \times 51 \times 50 = 132,600.$$

- What is the probability of first drawing a King, then a Queen, and then another King?

$$\frac{\binom{4}{1}\binom{4}{1}\binom{3}{1}}{132,600} = 0.00036199095.$$

Example: Probabilities of Card Hands (Cont.)

Thirteen cards are selected at random from a 52-card deck.

- How many possible hands are there?

$$\binom{52}{13} = \frac{52!}{13!39!} = 635,013,559,600.$$

- What is the probability of getting a hand consisting entirely of spades and clubs with at least one card of each suit?

$$\frac{\binom{26}{13} - 2}{\binom{52}{13}} = 0.000016379.$$

- What is the probability of getting a hand consisting of exactly two suits?

$$\frac{\binom{4}{2}\left[\binom{26}{13} - 2\right]}{\binom{52}{13}} = 0.000098271.$$