Lecture 1
Introduction to Probability

Chao Song

College of Ecology
Lanzhou University

September 11, 2023

Philosophy of learning statistics

What is the best approach to learning statistics for ecologists?

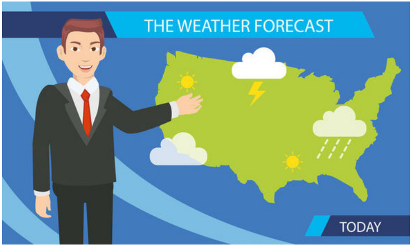
- **Application focus:** Which method do I use?
- **Implementation focus**: How do I implement an analysis in software?
- **Mathematical focus**: Why does this statistical method work?

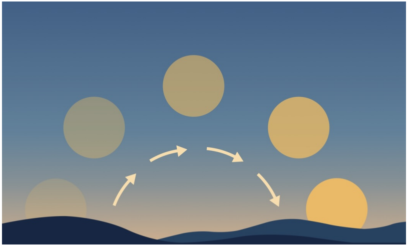
Which approach do we take in this class?

- A hybrid of mathematical and application focus.
- Know which method to use and understand, at least roughly, why it works.

What is probability?

Probability is the branch of mathematics concerning numerical descriptions of how likely an event is to occur; Probability focuses on events that are uncertain. Events that occurs with certainty is often not within the scope of probability.



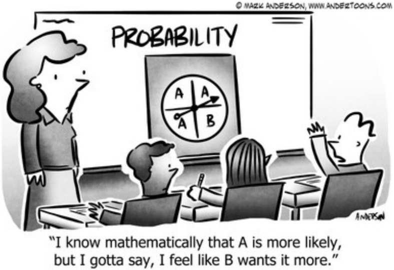


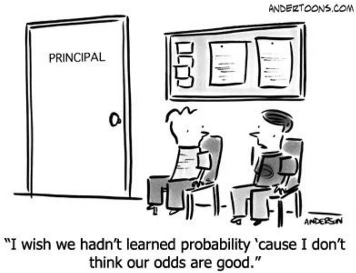
Events with uncertain outcomes, such as the weather tomorrow, is within the scope of probability while events with certain outcomes, such as the sunrise time tomorrow, is not.

What is probability?

We may interpret probability in various intuitive ways.

- **Frequentist view**: Probability measures how frequent an outcome occur if we perform the experiment repeatedly under the same condition;
- **Bayesian view**: Probability measures our belief on how likely a particular outcome is going to occur.





Cartoons illustrating the difference between frequentist and Bayesian view of probability.

Essential terminologies to define probability

Random experiment: experiment for which the outcome cannot be predicted with certainty;

Sample space (*S*): the collection of all possible outcomes;

Event (*A*): a collection of outcomes that is part of the sample space, i.e., $A \subset S$.

When the random experiment is performed and the outcome of the experiment is in A, we say that **event A has occurred**.

Essential terminologies to define probability

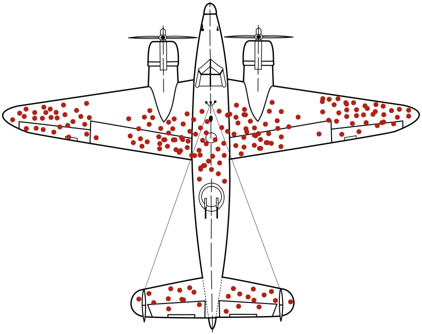
Example: Flip a coin 2 times and record the side facing up each time.

- Sample space $S = \{HH, HT, TH, TT\}$
- Event A_1: $\{$get all heads$\} = \{HH\}$
- Event A_2: $\{$get at least one head$\} = \{HH, HT, TH\}$
- Event A_3: $\{$get both head and tail$\} = \{HT, TH\}$

A note on sample space

Sample space is a seemingly intuitive concept. In practice, however, one needs to be careful in understanding what the sample space is.

Survival bias: the logical error of concentrating on entities that passed a selection process while overlooking those that did not.



Use patterns of damage on surviving planes to identify locations of reinforcement is an example of survival bias

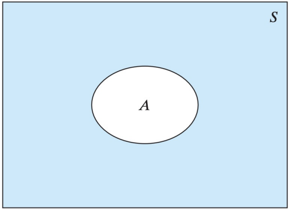
Probability and algebra of sets

In studying probability, the words **event** and **set** are often interchangable. Algebra of sets can be useful when calculating probability.

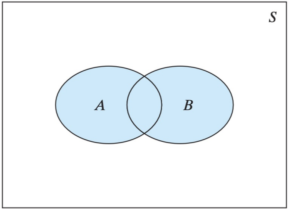
- **Commutative laws**:
  $A \cup B = B \cup A$
  $A \cap B = B \cap A$
- **Associative laws** :
  $(A \cup B) \cup C = A \cup (B \cup C)$
  $(A \cap B) \cap C = A \cap (B \cap C)$
- **Distributive laws**:
  $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
  $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
- **De Morgan's laws**:
  $(A \cup B)' = A' \cap B'$
  $(A \cap B)' = A' \cup B'$

Probability and algebra of sets

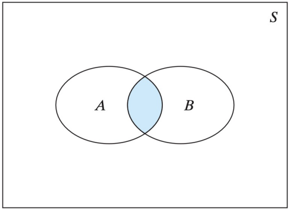
Venn diagram is a useful tool to visualize relationship among events and help calculate probability.



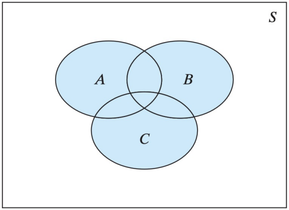
(a) A'



(b) $A \cup B$



(c) $A \cap B$



(d) $A \cup B \cup C$

Calculate probability

Example: Flip a fair coin 2 times and record the side facing up each time.

- Sample space $S = \{HH, HT, TH, TT\}$
- Event A_1: {get all heads}$= \{HH\}$
- Event A_2: {get at least one head}$= \{HH, HT, TH\}$
- Event A_3: {get both head and tail}$= \{HT, TH\}$

Each sampling point in the sample space occurs equally likely. We thus can calculate probability by counting the number of elements in the event and the sample space.

- $P(A_1) = \frac{1}{4}$;
- $P(A_2) = \frac{3}{4}$;
- $P(A_3) = \frac{2}{4}$

Classical model of probability

Let the sample space of a random experiment S be a finite set and $\#$ denotes the number of sampling points in the set. If the chance of occurrence for every sampling point is equal, the probability of event A is

$$P(A) = \frac{\# A}{\# S}.$$

This model of probability is referred to as the **classical model of probability**. Its development is attributed to the work by Jacob Bernoulli and Pierre-Simon Laplace.

Jacob Bernoulli (1655–1705) and Pierre-Simon Laplace (1749–1827)

Classical definition of probability

Based on the definition of the classical model of probability, probability can be calculated by **method of enumeration**, i.e. counting number of sampling points in event and sample space.

Example: There are 6 female and 16 male athletes in the team. Randomly select 11 athletes to participate in the Olympics Game. What is the probability that exactly 3 females athletes are selected?

Let A denote the event that 3 female athletes are selected.

$$
\begin{aligned}
{}^{\#}S &= \mathbf{C}_{22}^{11} = \frac{22!}{11!(22-11)!} = 705432 \\
{}^{\#}A &= \mathbf{C}_{6}^{3}\mathbf{C}_{16}^{8} = 257400 \\
P(A) &= \frac{{}^{\#}A}{{}^{\#}S} = 0.365
\end{aligned}
$$

Limitations of the classical model of probability

The classical model of probability only applies to scenarios when sampling points are equally likely. When sampling points do not occur with equal probability, the classical model of probability and the method of enumeration does not work any more.

Frequentist interpretation of probability: Let A be an event and f_N be the frequency of event A occurring in N repeated experiments. Probability of event A can be defined as $P(A) = \lim_{N \to \infty} f_N$.

Frequentists interpretation of probability

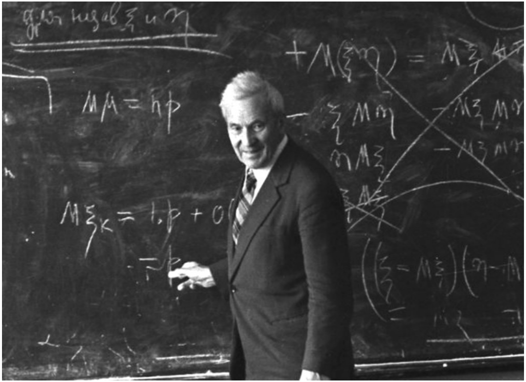
The frequentists definition of probability has been extensively tested by statisticians using the coin flipping experiment. These experiments show that as N increases, frequency of getting heads approaches 0.5.

Experimenter	N	N_{head}	N_{heads}/N
De Morgan	2048	1061	0.5181
Buffon	4040	2048	0.5069
Kerrich	7000	3516	0.5023
Kerrich	8000	4034	0.5043
Kerrich	9000	4538	0.5042
Kerrich	10000	5067	0.5067
Feller	10000	4979	0.4979
Pearson	12000	6019	0.5016
Pearson	24000	12012	0.5005
Romanovsky	80640	40173	0.4982

Table 1.5.2 in Shuyuan He, Probability and Statistics, 2021, Higher Education Press.

Axiomatic definition of probability

The axiomatic definition of probability was introduced by Soviet mathematician **Andrey Kolmogorov** in his 1933 book Foundations of the Theory of Probability.



Andrew N. Kolmogorov (1903–1987)

Axiomatic definition of probability

Probability is a real-valued set function P that assigns, to each event A in the sample space S, a number $P(A)$, called the probability of the event A, such that the following properties are satisfied:

- $P(A) \geqslant 0$;
- $P(S) = 1$;
- if $A_1, A_2, A_3 \ldots$ are events and $A_i \cap A_j = \varnothing$, $i \neq j$, then

$$P(A_1 \cup A_2 \cup A_3 \cup \ldots \cup A_k) = P(A_1) + P(A_2) + P(A_3) + \ldots P(A_k)$$

  for each positive integer k and

$$P(A_1 \cup A_2 \cup A_3 \cup \ldots) = P(A_1) + P(A_2) + P(A_3) + \ldots$$

  for an infinite, but countable, number of events.

Properties of probability

Based on the axiomatic definition of probability, we have:

- For each event A, $P(A) = 1 - P(A')$;
- $P(\varnothing) = 0$;
- If events A and B are such that $A \subset B$, $P(A) < P(B)$;
- For each event A, $P(A) \leqslant 1$;
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$;
- If events A and B are disjoint, $P(A \cup B) = P(A) + P(B)$.

Conditional probability

Given two events A and B. We denote the probability of event A happens **given** that event B is known to happen as $P(A|B)$.

Example: A deck of cards (52 cards without jokers) is well shuffled and one card is drawn randomly. What is the probability that the card is a K?

$$P(K) = \frac{4}{52} = \frac{1}{13}$$

What is the probability of drawing a K if the card is known to be a face card (J, Q. K)?

$$P(K|J, Q, K) = \frac{4}{12} = \frac{1}{3}$$

Why does conditional probability differ from unconditional probability? Current knowledge (face card) has changed or restricted the sample space.

Conditional probability

The conditional probability of event A given event B can be calculated as

$$P(A|B) = \frac{P(AB)}{P(B)}$$

Intuitively, we can view events A and B both occur as a two step process: event B occurs and then event A occur given that event B has already occurred.

The formula for conditional probability can be used the other way around. Multiplying both side by $P(B)$, we get the **general multiplication rule**:

$$P(AB) = P(B)P(A|B)$$

General multiplication rule for several events

The general multiplication rule can be extended to more than two events:

$$P(ABC) = P(A) \times P(B|A) \times P(C|A, B);$$

$$P(ABCD) = P(A) \times P(B|A) \times P(C|A, B) \times P(D|A, B, C);$$

$$P(ABCDE) = P(A) \times P(B|A) \times P(C|A, B) \times P(D|A, B, C) \times P(E \mid A, B, C, D).$$

Independent events

Two events are **independent** if knowing the outcome of one provides no useful information about the outcome of the other.

- **Independent**: Knowing that the coin landed on a head on the first toss does not provide any information for determining what the coin will land in the second toss.
- **Dependent**: Knowing that the first card drawn from a deck is an ace provide information on determining the probability of drawing an ace in the second draw.

Independent events

Definition: Events A and B are independent if and only if $P(AB) = P(A) \times P(B)$. In other words, $P(AB) = P(A) \times P(B)$ is a necessary and sufficient condition for A and B being independent.

- This comes from the general multiplication rule. $P(AB) = P(A) \times P(B|A)$ in which $P(B|A)$ reduces to $P(B)$ when events A and B are independent.
- More generally, if events $A_1, A_2, \ldots A_k$ are independent,

$$P(A_1 A_2 \ldots A_k) = P(A_1) \times P(A_2) \times \ldots P(A_k)$$

Abuse of the multiplication rule

As estimated in 2012, of the US population,

- 13.4% were 65 or older, and
- 52% of the population were male.

True or False: $0.134 \times 0.52 \approx 7\%$ of the US population were males aged 65 or older.

The answer is **false**

- Age and gender are not independent. On average women live longer than men. There are more old women than old men;
- According to survey data, among those 65 or older in the US, 44% are male, not 52%. Thus, $0.134 \times 0.44 \approx 5.9\%$ were males aged 65 or older in the US in 2012.

Law of total probability

Law of total probability: If events A_1, A_2, ..., A_n are pairwise disjoint events, $B \subset \bigcup_{k=1}^{n} A_k$, then

$$P(B) = \sum_{k=1}^{n} P(A_k)P(B|A_k)$$

Corollary 1: If events A_1, A_2, ..., A_n is a partition of the sample space, i.e., a set of pairwise disjoint events whose union is the entire sample space, then

$$P(B) = \sum_{k=1}^{n} P(A_k)P(B|A_k)$$

Corollary 2: For any events A and B, the probability of B can be calculated as $P(B) = P(A)P(B|A) + P(A')P(B|A')$.

Law of total probability

Seroprevalence adjustment: China Center for Disease Control and Prevention conducted a serological survey in April, 2020 in Wuhan and found 4.43% positive test. The test has a sensitivity of 90% and specificity of 98%. Does that mean that 4.43% of the Wuhan population were once infected?

Solution: Let A denote a positive test and B denote a true infection. Using the law of total probability, we have

$$P(A) = P(B)P(A|B) + P(B')P(A|B').$$

- $P(A) = 4.43\%$;
- Sensitivity $P(A|B) = 0.9$;
- Specificity $P(A'|B') = 0.98$;

Plug these values to the equation above, we get

$$0.0443 = P(B) \times 0.9 + (1 - P(B)) \times (1 - 0.98)$$

Solving the equation above for P(B), we arrive at $P(B) = 2.76\%$.

Bayes' theorem

Bayes' theorem, named after Thomas Bayes, describes the probability of an event, based on prior knowledge of conditions that might be related to the event:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)},$$

where A and B are events and $P(B) \neq 0$.

Thomas Bayes (1702–1761)

Bayes' theorem

Disease survey problem: Wu et al. (2019) estimated that people living with human immunodeficiency virus (HIV) has risen to more than 1.25 million in China, roughly 0.09% of the total population. A blood test for HIV typically has 95% accuracy, i.e., the test correctly detects positive cases or negative case 95% times. If a person is tested positive, what is the probability that this person is infected with HIV?

Solution: Let A denote that a person has HIV and B denote a positive test. We know $P(A) = 0.0009$, $P(B|A) = 0.95$, and $P(B'|A') = 0.95$. We want to know $P(A|B)$. Using Bayes' theorem, we have

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} = \frac{0.0009 \times 0.95}{P(B)},$$

Using law of total probability, we have

$$\begin{aligned} P(B) &= P(A)P(B|A) + P(A')P(B|A') \\ &= 0.0009 \times 0.95 + (1 - 0.0009) \times (1 - 0.95) \\ &= 0.05081 \end{aligned}$$

Finally, we have $P(A|B) = 1.68\%$

Bayes' theorem

Disease survey problem: Will a more accurate test help? If we improve the test sensitivity and specificity to 99.9%. What is the probability of having HIV if a person tested positive?

Solution: Using Bayes' theorem and the law of total probability:

$$
\begin{aligned}
P(A|B) &= \frac{P(A)P(B|A)}{P(A)P(B|A) + P(A')P(B|A')} \\
&= \frac{0.0009 \times 0.999}{0.0009 \times 0.999 + (1 - 0.0009) \times (1 - 0.999)} \\
&= 47.36\%.
\end{aligned}
$$

Implications: We should be cautious when confirming a rare disease, even if the diagnostic test is highly accurate.