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

Alexander Giessing

Department of Statistics

University of Washington

April 3, 2023

Outline

1. Conditional Probability

2. Independence of Events

3. Probability on Product Spaces

Conditional Probability

Knowing that an event has occurred sometimes forces us to reassess the probability of another event.

- The probability of a random person in the street being born in January is
$$P(\text{“born in January”}) = \frac{1}{12}.$$
- Now, suppose that we know that the person was born in Winter. We should then (intuitively?) calculate the probability as
$$P(\text{“born in January”} \mid \text{“born in Winter”}) = \frac{1}{3}.$$
This new probability is called *conditional probability*.

Conditional Probability

Conditional Probability

Let A, C be events. The conditional probability of A given C is defined as

$$P(A \mid C) := \frac{P(A \cap C)}{P(C)},$$

provided that $P(C) > 0$.

- Let $A = \{\text{January}\}$ and $C = \{\text{December, January, February}\}$.

  $\implies$ $P(A \cap C) = P(A) = 1/12$, $P(C) = 3/12$, and $P(A \mid C) = \frac{1}{3}$.

  $\implies$ Matches our intuition from the previous slide.

Multiplication Rule

- An immediate consequence of the definition of conditional probability is the following useful identity:

MULTIPLICATION RULE

For any events A and C,

$$P(A \cap C) = P(C)P(A \mid C) = P(A)P(C \mid A).$$

- Computing $P(A \cap C)$ can be decomposed into computing $P(C)$ and $P(A|C)$ separately; this often simplifies the calculations significantly!
- More generally, for any events $A_1, \dots A_m$,

$$\begin{aligned} & P(A_1 \cap A_2 \cap \dots A_m) \\ & = P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1 \cap A_2) \cdots P(A_m \mid A_1 \cap A_2 \cap \dots \cap A_{m-1}). \end{aligned}$$

Example: Birthday Problems

What is the probability two randomly chosen people, Ali and Bet, have different birthdays?

$$P(\text{“diff. birthdays”}) = P(\text{Ali} \neq \text{Bet}) = 1 - P(\text{Ali} = \text{Bet}) = 1 - \frac{365}{365^2} = \frac{364}{365}$$

[1] *Suppose Cal joins Ali and Bet. What is the probability that all three have different birthdays?*

$$\begin{aligned} P(\text{“diff. birthdays”}) &= P(\text{Cal} \neq \text{Ali and Cal } \neq \text{Bet} \mid \text{Ali} \neq \text{Bet}) P(\text{Ali} \neq \text{Bet}) \\ &= \frac{363}{365} \times \frac{364}{365}. \end{aligned}$$

What is the probability that Ali, Bet, Cal, and Dan have all four different birthdays?

$$\begin{aligned} P(\text{“diff. birthdays”}) &= P(D \neq A, D \neq B, D \neq C) \mid A \neq B, A \neq C, B \neq C) \\ &\quad \times P(A \neq B, A \neq C, B \neq C) \\ &= \frac{362}{365} \times \frac{363}{365} \times \frac{364}{365}. \end{aligned}$$

---

[1] $P(\text{Ali's birthday on day x}) = \frac{1}{365}$, $P(\text{Bet's birthday on day y}) = \frac{1}{365}$, and # (pairs (x,y) with x=y) = 365. So, $P(A = B) = \frac{365}{365^2} = \frac{1}{365}$.

Example: Drawing Lottery Tickets w/o Replacement

A box contains 10 tickets. Three of them are winning tickets. Pick three at random without replacement.

- What is the chance that the first three randomly picked tickets are the winning ones?

  Let W_i be the event that the ith draw is a winning ticket. Then,

$$
\begin{aligned}
P(W_1 \cap W_2 \cap W_3) &= P(W_1)P(W_2 \mid W_1)P(W_3 \mid W_1 \cap W_2) \\
&= \frac{3}{10} \times \frac{2}{9} \times \frac{1}{8} = \frac{1}{120}.
\end{aligned}
$$

- What is the probability that the second ticket is a winning one?

$$
\begin{aligned}
P(W_2) &= P(W_2 \cap W_1) + P(W_2 \cap W_1^c) \\
&= P(W_1)P(W_2 \mid W_1) + P(W_1^c)P(W_2 \mid W_1^c) \\
&= \frac{3}{10} \times \frac{2}{9} + \frac{7}{10} \times \frac{3}{9} = \frac{3}{10}.
\end{aligned}
$$

Law of Total Probability

- In the last example we have split the event W_2 into two disjoint events $W_2 \cap W_1$ and $W_2 \cap W_1^c$ and used conditioning to find the probability of the original event W_2.

  The following law generalizes this result.

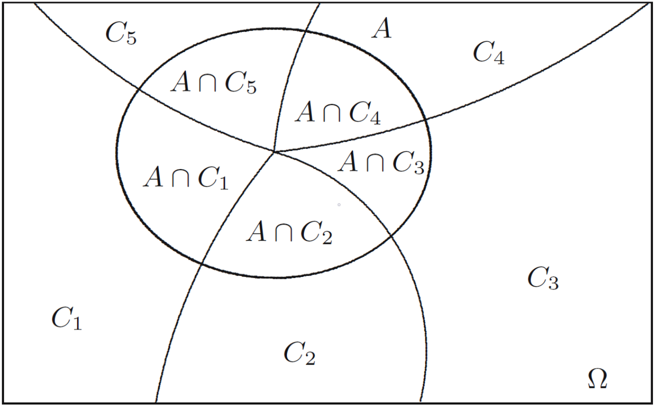
LAW OF TOTAL PROBABILITY

Suppose that $C_1, C_2, \ldots, C_m$ are pairwise disjoint events and $C_1 \cup C_2 \cup \ldots \ldots \cup C_m = \Omega$. The probability of an arbitrary event A can be expressed as

$$P(A) = P(A \mid C_1)P(C_1) + P(A \mid C_2)P(C_2) + \ldots + P(A \mid C_m)P(C_m).$$

Law of Total Probability

(Illustration of the law of total probability for $m = 5$.)



Bayes' Rule

- For events A, C with $P(A) > 0$ the definition of conditional probability yields,

$$P(C \mid A) = P(C) \times \frac{P(A \mid C)}{P(A)}.$$

  - $P(C)$ is called the **prior** probability of C.
  - $P(C|A)$ is called the **posterior** probability of C given A.
  - Tells us how to modify knowledge about C based on information about A.

Combined with the Law of Total Probability we have the following:

Bayes' Rule

Suppose that $C_1, C_2, \ldots, C_m$ are pairwise disjoint events and $C_1 \cup C_2 \cup \ldots \ldots \cup C_m = \Omega$. The conditional probability of C_i, given an arbitrary event A, can be expressed as

$$P(C_i \mid A) = \frac{P(A \mid C_i)P(C_i)}{P(A \mid C_1)P(C_1) + P(A \mid C_2)P(C_2) + \ldots + P(A \mid C_m)P(C_m)}.$$

Example: Testing for Rare Diseases

Lab tests produce positive and negative results. Assume that a lab test has 95% sensitivity and 98% specificity. Assume that the prevalence probability of the disease is 1%.

- sensitivity (= "true positive rate"): propensity of the test to correctly identify a sick person as being sick.
- specificity (= "true negative rate"): propensity of the test to correctly identify a healthy person as being healthy.
- What is the probability that lab test is positive for a random person?

$$\begin{aligned} P(+) &= P(\text{“sick”})P(+ \mid \text{“sick”}) + P(\text{“not sick”})P(+ \mid \text{“not sick”}) \\ &= 0.01 \times 0.95 + 0.99 \times 0.02 = 2.93\%. \end{aligned}$$

Example: Testing for Rare Diseases (Cont.)

- Suppose the lab test of a randomly selected person is positive. What is the probability that the person really has the disease?

$$
\begin{aligned}
&P(\text{“sick”} \mid +) \\
&\quad = \frac{P(\text{“sick”} \cap +)}{P(+)} \\
&\quad = \frac{P(\text{“sick”})P(+ \mid \text{“sick”})}{P(\text{“sick”})P(+ \mid \text{“sick”}) + P(\text{“not sick”})P(+ \mid \text{“not sick”})} \\
&\quad = \frac{0.01 \times 0.95}{0.01 \times 0.95 + 0.99 \times 0.02} < 33\%.
\end{aligned}
$$

Example: Testing for Rare Diseases (Cont.)

- What is the probability that, given that the lab test is negative, the person not does have the disease?

$$
\begin{aligned}
&P(\text{“not sick”} \mid -) \\
&\quad = \frac{P(\text{“not sick”} \cap -)}{P(-)} \\
&\quad = \frac{P(\text{“not sick”})P(- \mid \text{“not sick”})}{P(\text{“sick”})P(- \mid \text{“sick”}) + P(\text{“not sick”})P(- \mid \text{“not sick”})} \\
&\quad = \frac{0.99 \times 0.98}{0.01 \times 0.05 + 0.99 \times 0.98} > 99\%.
\end{aligned}
$$

- What is the probability that, given that the lab test is negative, the person does have the disease? Answer: $P(\text{“sick”} \mid -) = 0.052\%$.

Outline

Independence of Events

- Consider the three probabilities from the previous example:

$$P(\text{“sick”}) = 1\%$$
$$P(\text{“sick”} \mid +) = 33\%$$
$$P(\text{“sick”} \mid -) = 0.052\%$$

- If we know nothing about the person, we would say that there is a 1% chance that he is infected. However, if we know that he tested negative, there is only a 0.052% chance. While if we know that he tested positive, the chance jumps to 33%.
- The probability of the event “sick” depends on the outcome of the test!
- Imagine the opposite: the lab test is useless. Whether the person is sick is unrelated to the outcome of the test. We would then say that the disease status {“sick”, “not sick”} and the events {‘+”, “−”} are *independent*.

Independence of Events

Independent Events

Let A and B be events. A is called independent of B if

$$P(A \mid B) = P(A).$$

Equivalent Formulations

Let A and B be events. The following statements are equivalent:

(i) A is independent of B.

(ii) B is independent of A.

(iii) $P(A \cap B) = P(A)P(B)$.

(iv) $P(A^c \cap B) = P(A^c)P(B)$.

(v) $P(A \cap B^c) = P(A)P(B^c)$.

(vi) $P(A^c \cap B^c) = P(A^c)P(B^c)$.

- Because of (i) and (ii) we also just say events A and B are independent.

Example: Independence of Two Events

Suppose a couple has two children. Given that the first child is a girl, what is the probability that the second is a boy? Are these two events independent?

- Sample space $\Omega = \{GG, GB, BG, BB\}$.
- Probability is fully specified (why?) by

$$P(GG) = P(GB) = (BG) = P(BB) = \frac{1}{4}.$$

- Define events

$$A = \{\text{“second child is a boy”}\} = \{GB, BB\},$$
$$C = \{\text{“first child is a girl”}\} = \{GG, GB\}.$$

- Then, $P(A) = P(C) = \frac{1}{2}$ (why?) and, by Bayes' rule,

$$P(A \mid C) = \frac{P(A \cap C)}{P(C)} = \frac{P(\{GB\})}{P(\{GG, GB\})} = \frac{1}{2}.$$

- We conclude that A and C are independent.

Independence of two or more Events

INDEPENDENCE OF TWO OR MORE EVENTS

Events $A_1, A_2, \ldots, A_m$ are called mutually independent if for every subset of indices $\{i_1, \ldots, i_k\} \subseteq \{1, \ldots m\}$,

$$P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_k}).$$

- Note: in total 2^m equalities have to hold for m events to be independent. (In fact, $m+1$ of these equations are redundant, but this version of the definition is easier to state.)
- Independence is an idealistic assumption and often violated in practice.

Example: Independence of Three Events

A coin is tossed twice. Let $\widetilde{H}_1 = \{H_1H_2, H_1T_2\}$ *and* $\widetilde{T}_1 = \{T_1H_2, T_1T_2\}$ *be the events that the first toss is head and tail,respectively. Similar notation applies to the second toss. Define the event* $E = \{H_1H_2, T_1T_2\}$ *(or, in plain English,* E *= "the two tosses are equal").*

- The events $E, \widetilde{H}_1, \widetilde{H}_2$ are pairwise independent because

$$P(E \cap \widetilde{H}_1) = \frac{1}{4} = P(E)P(\widetilde{H}_1)$$
$$P(E \cap \widetilde{H}_2) = \frac{1}{4} = P(E)P(\widetilde{H}_2)$$
$$P(\widetilde{H}_1 \cap \widetilde{H}_2) = \frac{1}{4} = P(\widetilde{H}_1)P(\widetilde{H}_2).$$

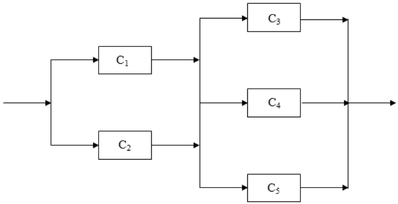
- But the events $E, \widetilde{H}_1, \widetilde{H}_2$ are not mutually independent because

$$P(E \mid \widetilde{H}_1 \cap \widetilde{H}_2) = 1 \neq \frac{1}{2} = P(E).$$

- This shouldn't surprise: whether E occurs follows deterministically from the outcomes of tosses 1 and 2.
- Conclusion: pairwise independence does not imply mutual independence for more than 2 events.

Example: Use of Independence

Suppose that a system consists of 5 independent components $C_1, \dots, C_5$ connected as below. If the probability that each component works properly is 90%, what is the probability that the system works properly?



- Let D_1 and D_2 be the the events that the first and second parallel device work properly, and C_k the event that $k$the device works properly. Then,

$$P(D_1) = P(C_1 \cup C_2) = 1 - P(C_1^c \cap C_2^c) = 1 - 0.1 \times 0.1 = 0.99,$$
$$P(D_2) = \dots = 1 - 0.1 \times 0.1 \times 0.1 = 0.999.$$

- Thus, the probability that the system as functions properly is

$$P(D_1 \cap D_2) = 0.99 \times 0.999 = 0.989.$$

- How often have we used characterization (iii) of independence? Answer: 5.

Outline

Modeling More Complex Experiments

- The techniques developed so far allow us to model
  - single experiments with finite sample spaces, e.g. coin tosses, dice rolls, card hands.
  - repeated experiments with finite sample space and equally likely outcomes, e.g. tossing a fair coin twice or five times.
- How can we model experiments that are possibly repeated infinitely many times and do not have equally likely outcomes?
  - unfair games of chance, e.g. tossing a biased coin, roulette, ...
  - transmission of infectious diseases, spread of computer malware.
  - random mutations in DNA/ RNA.
  - Bitcoin price next year.
  - ...
- Repeated experiments are fundamental to statistics; repetitions are necessary in order to learn (aka *infer*) about real world phenomena.

Example: Tossing a Biased Coin Twice

We toss a coin twice. Let H_1 *and* T_1 *be the events that the first toss is head and tail, respectively. Similar notation applies to the second toss. The sample space is* $\Omega = \{H_1H_2, T_1H_2, H_1T_2, T_1T_2\}$.

- Suppose the coin is fair, i.e. $P(H_i) = P(T_i) = 0.5$ for $i = 1, 2$. Since all outcomes in Ω are equally likely, for any event $A \subset \Omega$,

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

- Now, suppose the coin is biased, say, $P(H_i) = 1 - P(T_i) = 0.1$ for $i = 1, 2$. Intuitively, outcome H_1H_2 should be less likely than T_1H_2 than T_1T_2.
  $\Longrightarrow$ outcomes in Ω are no longer equally likely.

- How can we assign probabilities to events associated with such experiments?

Example: Tossing a Biased Coin Twice (Cont.)

- Idea: assume that the two coin tosses are independent.
- Here, independence means that the outcome of the first coin toss does not affect the outcome of the second coin toss. Is that a reasonable assumption?
- Under independence, we can easily compute the probabilities of the simple events in Ω as

$$P(H_1H_2) = P(H_1) \times P(H_2) = 0.01, \quad P(T_1H_2) = P(T_1) \times P(H_2) = 0.09,$$
$$P(H_1T_2) = P(H_1) \times P(T_2) = 0.09, \quad P(T_1T_2) = P(T_1) \times P(T_2) = 0.81.$$

- Can use above probabilities to calculate the chances of more complicated events, e.g.

$$P(\text{“at least one head”}) = 0.01 + 2 \times 0.09 = 0.19.$$

Probability on Product Spaces

- The preceding example lends itself to the following generalization:

PRODUCT MEASURE AND PRODUCT SPACE

Suppose we perform n independent experiments with probability triples $(\Omega_i, \mathcal{A}_i, P_i)$, $i = 1, \ldots, n$. Then, the joint experiment has probability triple $(P, \Omega, \mathcal{A})$, with product measure $P := P_1 \times \ldots \times P_n$, product space $\Omega := \Omega_1 \times \ldots \times \Omega_n$, and $\mathcal{A} := \mathcal{A}_1 \times \ldots \times \mathcal{A}_n$.

- The Cartesian product between two sets $\Omega_1, \ldots, \Omega_n$ is defined as

$$\Omega_1 \times \ldots \times \Omega_n := \{(\omega_1, \ldots, \omega_n) : \omega_1 \in \Omega_1, \ \ldots, \ \omega_n \in \Omega_n\}.$$

- Convince yourself that for $\Omega_1 = \Omega_2 = \{H, T\}$

$$\Omega_1 \times \Omega_2 = \{(H, H), (H, T), (T, H), (T, T)\}.$$

Example: Tossing Five and More Biased Coins

A machine tosses 5 coins at random. Suppose the coins are mutually independent and each coin has $P(H) = 1 - P(T) = 0.1$. What is the probability that exactly one coin shows heads?

- There are 5 possible outcomes that comprise exactly one head.
- By independence each one of these events has probability 0.1×0.9^4.
- We conclude that

$$P(\text{“exactly one coin shows heads”}) = 5 \times 0.1 \times 0.9^4 = 32.81\%.$$

Suppose a machine tosses n coins, each biased with $P(H) = 1 - P(T) = 0.1$. You win, if exactly one coin shows heads. Which n maximizes your probability for winning this game?

- $P(\text{“exactly one coin shows heads”}) = n \times 0.1 \times 0.9^{n-1}$.
- The number n maximizing this probability is $n = 1/\log(1/0.9)$. Rounding to an integer yields $n \in \{9, 10\}$ with probability of winning of 38.74%.