

Probability Definitions

- The **probability** of an outcome refers to how often the outcome would occur in the long run if a random process were repeated over and over under identical conditions (relative frequency interpretation).
- An **experiment** is any activity or situation in which there is uncertainty about the outcome.
- The **sample space** is the list of all possible outcomes of a random trial. (Called S .)
- An **event** is any potential subset of the sample space.
- A **simple event** is an event consisting of exactly one outcome.
- Two events are **mutually exclusive** or **disjoint** if they cannot both occur simultaneously.
- Two events are **independent** if the occurrence of one does not change the probability that the second will occur.

Set Theory

Consider three events.

A : {woman has a tattoo }

B : {someone age 18-29 has a tattoo }

C : {someone has a tattoo }

$A \subset C$ $B \subset C$ because both A and B are subsets of C

EMPTY SET Is the set with no outcomes and is denoted: \emptyset .

UNION The union of A and B is defined to be the event containing all outcomes that belong to A alone, to B alone, or to both A and B : $A \cup B$.

$$A \cup B = B \cup A \quad A \cup A = A$$

$$A \cup \emptyset = A \quad A \cup S = S$$

$$\text{if } A \subset B \Rightarrow A \cup B = B$$

$$A \cup B \cup C = (A \cup B) \cup C = A \cup (B \cup C) \text{ (associative)}$$

If we have n separate events, A_1, A_2, \dots, A_n ,

$$A_1 \cup A_2 \cup \dots \cup A_n = \bigcup_{i=1}^n A_i$$

INTERSECTION The intersection of A and B is defined as the events in *both* A and B: $A \cap B = AB$.

$$\begin{aligned}A \cap B &= B \cap A & A \cap A &= A \\A \cap \emptyset &= \emptyset & A \cap S &= A \\ \text{if } A \subset B &\Rightarrow A \cap B = A \\ A \cap B \cap C &= (A \cap B) \cap C = A \cap (B \cap C) \text{ (associative)}\end{aligned}$$

If we have n separate events, A_1, A_2, \dots, A_n ,

$$A_1 \cap A_2 \cap \dots \cap A_n = \bigcap_{i=1}^n A_i$$

COMPLEMENT The complement of an event A is defined to be the event that contains all outcomes in the sample space that do *not* belong to A: A^c

$$\begin{aligned}(A^c)^c &= A & \emptyset^c &= S & S^c &= \emptyset \\ A \cup A^c &= S & A \cap A^c &= \emptyset\end{aligned}$$

DISJOINT A and B are disjoint (or mutually exclusive) if they have no outcomes in common, that is, if $A \cap B = \emptyset$.

Probability Rules

AXIOM 1 $P(A) \geq 0$ for all A .

AXIOM 2 $P(S) = 1$

AXIOM 3 for an infinite series of disjoint events, $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$

THM 1.5.1 $P(\emptyset) = 0$

THM 1.5.2 for a finite series of disjoint events, $P(\cup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$

THM 1.5.3 $P(A^c) = 1 - P(A)$

THM 1.5.4 if $A \subset B$, $P(A) \leq P(B)$

THM 1.5.5 for all A , $0 \leq P(A) \leq 1$

THM 1.5.6 for all A, B , $P(A \cup B) = P(A) + P(B) - P(AB)$

Examples

“Our brains are just not wired to do probability problems very well” (Persi Diaconis, 1989)

1. A cab was involved in a hit and run accident at night. To cab companies, the Green and the Blue operate in the city. Suppose you are told the following:
 - 85% of the cabs in the city are Green, and the remaining 15% are Blue.
 - A witness identified the cab as Blue (but it was dark outside!). The court tested the reliability of the witness under the same conditions that existed on the night of the accident. The court determined that the witness correctly identified the cab color 80% of the time and made a mistake 20% of the time.

What is the probability that the cab involved in the hit-and-run was actually Blue rather than Green?

table: $P(\text{blue} \mid \text{said blue}) = 120/290 = 0.4137$

2. A medical research team wished to evaluate a proposed screening test for Alzheimer’s disease. The test was given to a random sample of 450 patients with Alzheimer’s disease and an independent random sample of 500 patients without symptoms of the disease. The two samples were drawn from populations of subjects who were 65 years of age or older. It is believed that 11.3 percent of the population aged 65 and older have Alzheimer’s disease. The results are as follows:

	Alzheimer’s (Alz)	No Alzheimer’s (No Alz)	Total
Test Positive (Pos)	436	5	441
Test Negative (Neg)	14	495	509
Total	450	500	950

- (a) Are the variables disjoint? [no]
- (b) Are the variables independent? [no: $P(\text{Alz}) = 450/950 = 0.474$, $P(\text{Alz} \mid \text{Pos}) = 14/509 = 0.0275$]
- (c) (sensitivity) Find $P(\text{Pos} \mid \text{Alz})$ [= $436/450 = 0.969$]
- (d) (specificity) Find $P(\text{Neg} \mid \text{No Alz})$ [= $495/500 = 0.99$]
- (e) (predictive value positive) Find $P(\text{Alz} \mid \text{Pos})$ [= $14/509 = 0.0275$]
- (f) (predictive value negative) Find $P(\text{No Alz} \mid \text{Neg})$ [= $495 / 509 = 0.972$]

3. The Birthday Problem

Consider a situation where you and a friend have the same birthday. Because there are 365 days in a year (usually), you think it is remarkable that your friend shares your birthday. However, in spreading the news, you find that lots of people share birthdays with their friends! You wonder to yourself, in a room full of 50 people, how likely is it that at least two people will share a birthday? To answer the question, let’s assume that there are only 365 days in the year and that a birthday is equally likely to be on

any of the 365 days. Additionally, let's assume that there aren't any twins in the room (or other dependencies); that is, we assume that each person's birthday is independent of the other people in the room.

- (a) What is the probability that a randomly selected person in the room shares your birthday?
- (b) What is the probability that 2 randomly selected people are both born on December 22nd?
- (c) What is the probability that out of 3 people, exactly two of them share a birthday? In order to answer this question, think about people A, B, C. For exactly two of them to share a birthday, the other one must have a different birthday. That is, if A and B share a birthday, C must have a different birthday. What are all the possible combinations?
- (d) Continuing with the previous question, what is the probability that at least 2 of the randomly selected people share a birthday? Notice that here we've simply added the additional possibility that all three of the people can have the same birthday.
- (e) Thinking of the birthday problem another way, what is the probability that none of the 3 randomly selected people shares a birthday? How does this probability compare to the previous one? They should be related but not equal.
- (f) Notice that it was slightly easier to compute the probability that none of the 3 shared a birthday than that at least 2 shared a birthday (and both gave us the answer we were looking for!) If we have 4 people, it will get even harder to think about at least 2 sharing a birthday (2 of 4, 3 of 4, 4 of 4). What if we had 50 people? Calculate the probability that out of 5 people, none of them share a birthday.
- (g) Using the logic from above, find the probability that out of 5 people at least 2 of them share a birthday.
- (h) Derive a formula for computing the probability that out of n people at least 2 of them share a birthday.