

TABLE 1. Translations between events and sets. To interpret the Venn diagrams in terms of events, imagine that a point is picked at random from the square. Each point in the square then represents an outcome, and each region of the diagram represents the event that the point is picked from that region.

Event language	Set language	Set notation	Venn diagram
outcome space	universal set	Ω	
event	subset of Ω	$A, B, C, \text{ etc.}$	
impossible event	empty set	\emptyset	
not A , opposite of A	complement of A	A^c	
either A or B or both	union of A and B	$A \cup B$	
both A and B	intersection of A and B	$AB, A \cap B$	
A and B are mutually exclusive	A and B are disjoint	$AB = \emptyset$	
if A then B	A is a subset of B	$A \subseteq B$	

Introduction: Summary

Outcome space: A set of all possible outcomes of a situation or experiment, such that one and only one outcome must occur.

Events: Represented as subsets of an outcome space.

A and B , AB , $A \cap B$, *intersection*: event that both A and B occur.

A or B , $A \cup B$, *union*: event that either A or B (or both) occur.

$AB = \emptyset$, *disjoint; mutually exclusive*: no overlap, no intersection.

not A , A^c , *complement*: opposite of A : event that occurs if A does not.

$A \subset B$, *inclusion*: A is a part of B , A implies B , if A occurs then so does B .

Ω , *whole set, outcome space*: certain event, all possibilities, sure to happen.

\emptyset , *empty set, impossible event*: no way to happen.

partition of A : disjoint sets A_1, \dots, A_n with union A .

Rules of Probability and Proportion

- Non-negative: $P(A) \geq 0$
- Addition: $P(A) = \sum_{i=1}^n P(A_i)$ if A_1, \dots, A_n is a partition of A
- Total of 1: $P(\Omega) = 1$.
- Between 0 and 1: $0 \leq P(A) \leq 1$
- Empty set: $P(\emptyset) = 0$
- Complements: $P(A^c) = 1 - P(A)$
- Difference: $P(BA^c) = P(B) - P(A)$ if $A \subset B$
- Inclusion–Exclusion: $P(A \text{ or } B) = P(A) + P(B) - P(AB)$.

Relative frequency: Proportion of times something happens: $\frac{\text{\#of times it happens}}{\text{\#of trials}}$

Interpretations of Probability

- long-run relative frequency (statistical average): $P_n(A) \approx P(A)$ for large n .
- degree of belief (probabilistic opinion)

Probability distribution over Ω : Assignment of probabilities to events represented as subsets of Ω , satisfying rules of probability. A distribution over a finite set Ω can be specified with a *distribution table*:

outcome ω	a	b	c	...
probability $P(\omega)$	$P(a)$	$P(b)$	$P(c)$...

The probabilities must sum to 1 over all outcomes.

Odds

Chance odds: ratio of probabilities, e.g., the following are equivalent: $P(A) = 3/10$; the odds of A are 3 in 10; the odds *in favor of* A are 3 to 7; the odds *against* A are 7 to 3.

Payoff odds: ratio of stakes: $\frac{\text{what you get}}{\text{what you bet}}$ (what you get does not include what you bet).

Fair odds rule: in a fair bet, payoff odds equal chance odds.

Conditional Probability

$P(A|B)$ = probability of A given B : probability of A with outcome space reduced to B . Compare with $P(A)$ = overall or unconditional probability of A .

Interpretations of conditional probability:

- *Intuitive/subjective:* chance of A if B is known to have occurred:
- *Long-run frequency:* long-run relative frequency of A 's among trials that produce B .

As a function of A , for fixed B , conditional probabilities satisfy the rules of probability, e.g., $P(A^c|B) = 1 - P(A|B)$

Rules of Conditional Probability

Division: $P(A|B) = \frac{P(AB)}{P(B)}$ (note: $AB = BA$)

For probabilities defined by counting, $P(A|B) = \#(AB)/\#(B)$. Similarly for length, area, or volume instead of #.

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

The following rules refer to a *partition* B_1, \dots, B_n of Ω , so $P(B_1) + \dots + P(B_n) = 1$; for example, $B_1 = B$, $B_2 = B^c$ for any B .

Average rule: $P(A) = P(A|B_1)P(B_1) + \dots + P(A|B_n)P(B_n)$

Bayes' rule: $P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A)}$ where $P(A)$ is given by the weighted average formula.

Independence

Two trials are *independent* if learning the result of one does not affect chances for the other, e.g., two draws at random with replacement from a box of known composition.

The trials are *dependent* if learning the result of one does affect chances for the other, e.g., two draws at random without replacement from a box of known composition, or two draws at random with replacement from a box of random composition.

Independent events: A and B are such that
 $P(AB) = P(A)P(B) \iff P(A|B) = P(A)$ (learning B occurs does not affect chances of A)
 $\iff P(B|A) = P(B)$ (learning A occurs does not affect chances of B)

Independence of n events A_1, \dots, A_n :

$$P(A_1 A_2 \cdots A_n) = P(A_1) \cdots P(A_n),$$

and the same with any number of complements A_i^c substituted for A_i (2^n identities).