Probabilities

Chance behavior is unpredictable in the short run but has a regular and predictable pattern in the long run.

We call a phenomenon **random** if individual outcomes are uncertain but there is a regular **distribution** of outcomes in a large number of repetitions as the figure shows.

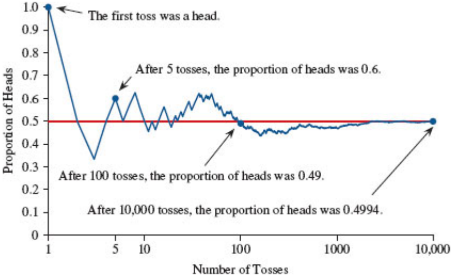


Figure 4.1 As the number of tosses increases, the proportion of heads fluctuates around the true probability of 0.5, and gets closer to 0.5. The horizontal axis is not drawn to scale.

Define probability

The **probability** of an outcome is the proportion of times the outcome would occur in a very long series of repetitions.

In the above example, the proportion of heads showing up in a very long series of repetitions is 1/2, then we say the probability of having a head in a single toss is 1/2.

probability models

The **sample space S** of a random phenomenon is the set of all possible outcomes.
e.g., toss a coin, S={head, tail}; throw a die, S={1,2,3,4,5,6}.

An **event** is an outcome or a set of outcomes of a random phenomenon. That is, an event is a subset of the sample space.
e.g., throw a die, event A= get an odd number={1,3,5}.

A **probability model** consists of a sample space S and a way of assigning probabilities to events.

Compute probability: Equally likely events

If a sample space has n equally likely outcomes and an event A has k outcomes, then $P(A) = \frac{k}{n}$.

Example: A fair die is rolled. Find the probability that an odd number would come up.
$P(\text{odd number}) = \frac{3}{6} = \frac{1}{2}$.

Compute probability: Empirical method

When the outcomes are not equally likely, just use empirical proportion to compute probability.

Example 4.8. In 2002, it is reported that 2,057,979 boys and 1,963,747 girls were born. Find the probability that a newborn baby is a boy.

$P(\text{boy}) = \frac{2057979}{2057979+1963747} = 0.5117$

Probability rules

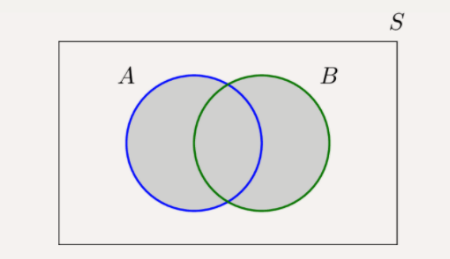
- The probability $P(A)$ of any event A satisfies $0 \leq P(A) \leq 1$.
- If S is the sample space, then $P(S) = 1$.
- Two events A and B are disjoint (or mutually exclusive) if they have no outcomes in common. If A and B are disjoint,
  $P(A or B) = P(A) + P(B)$
  A or B includes outcomes that are either in A or in B.
- For any event A,
  P(A does not occur) $= P(A^C) = 1 - P(A)$. (C means complement)

  In the previous example,
  P(girl) = P(not boy)=1-P(boy)=1-0.5117=0.4883.

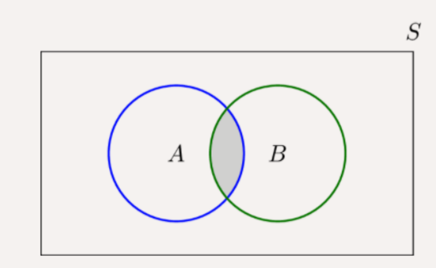
General addition rule

For any two events A and B,
$P(A \text{or} B) = P(A) + P(B) - P(A \text{and} B)$.
A and B includes outcomes that are both in A and B.
e.g., throw a die, A=$\{1, 2, 3\}$ B=$\{2, 4, 6\}$,
A or B=$\{1, 2, 3, 4, 6\}$, A and B=$\{2\}$.
Here P (A or B)= $\frac{5}{6}$ =P(A)+P(B)-P(A and B)=$\frac{3}{6} + \frac{3}{6} - \frac{1}{6}$.

Venn Diagram for events



Shaded area is A or B



Shaded area is A and B

Addition rule application

It is known that 70% of JMU students have a visa credit card, 50% have a master card, and 35% have both. What percent of JMU students have either a visa card or a master card or both? What percent have neither?

$P(VorM) = 0.7 + 0.5 - 0.35 = 0.85,$

$P(Neither) = 1 - 0.85 = 0.15.$

Exercises

Choose an American adult at random. The probability that you choose a women is 0.52, you choose a person who is never married is 0.25, you choose a woman who is never married is 0.11. What is the probability that the person you choose is either a women or never married?

Answer: 0.52+0.25-0.11=0.66.

100 senators classified by gender and party.

	Male	Female	Total
Democrat	36	16	52
Republican	42	4	46
Independent	2	0	2
Total	80	20	100

Compute the probabilities

a. The senator is a male Republican. $P=\frac{42}{100} = 0.42$

b. The senator is a Democrat or a female P=0.52+0.20-0.16=0.56

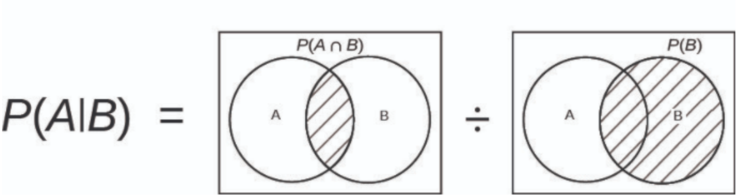
c. The senator is not a Republican. P=1-0.46=0.54

d. The senator is a Democrat or an Independent. P=0.52+0.02=0.54

Conditional probability

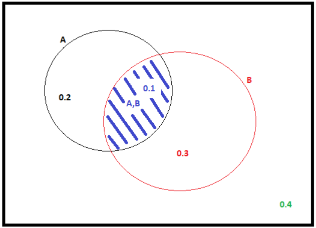
When $P(B) > 0$, the conditional probability of A given B is
$P(A|B) = \frac{P(A \text{and} B)}{P(B)}$.

$A \cap B$ in the plot below means A and B.



So roughly $P(A|B)$ is the ratio of the size of (A and B) over size of B.
Given B means we fix B as the total. Non conditional probabilities always takes S as the total.

Conditional probability



From this plot, $P(A|B) = \frac{\text{size of A and B}}{\text{size of B}} = \frac{0.1}{0.1+0.3} = \frac{1}{4}$
and $P(B|A) = \frac{0.1}{0.1+0.2} = \frac{1}{3}$. (now A is the total and is on the bottom)

But for non conditional probability:
$P(A) = \frac{\text{size of A}}{\text{size of S}} = \frac{0.3}{1} = 0.3$
$P(AandB) = \frac{\text{size of A and B}}{\text{size of S}} = \frac{0.1}{1} = 0.1.$

Compute conditional probability

	BigEater	NormalEater	total
Rich	0.2(20)	0.1(10)	0.3(30)
Modest	0.2(20)	0.5(50)	0.7(70)
total	0.4(40)	0.6(60)	1.0(100)

Table: 100 customers classified by income level or big eater

The numbers in the tables are probabilities, and the numbers in the parentheses are counts. The counts divided by the table total 100 give probabilities.

Find $P(BigEater|Rich), P(Rich|NormalEater)$.

$P(BigEater|Rich) = \frac{0.2}{0.3} = \frac{2}{3}$

$P(Rich|NormalEater) = \frac{0.1}{0.6} = \frac{1}{6}$.

Also read Example 4.18 in textbook for another example.

Exercise: practice conditional probabilities

The table below classifed victims who died of violent deaths by their gender and cause of death. Note the blue numbers are row or column totals, the red number is the table total.

Gender	Accidents	Homicide	Suicide	Total
female	1818	457	345	2620
male	6457	2870	2152	11479
Total	8275	3327	2497	14099

1. The probability that a randomly selected victim was male is
a) 0.81; b) 0.78; c) 0.19
Answer: $P(male) = \frac{11479}{14099} = 0.81$.
2. The conditional probability that the victim was male given that the death was accidental, is
a) 0.81; b) 0.78; c) 0.56.
Answer: $P(male|accidental) = \frac{6457}{8275} = 0.78$.

Continued

3. The conditional probability that the death was accidental, given that the victim was male, is
a) 0.81; b) 0.78; c) 0.56.
Answer: $P(accident|male) = \frac{6457}{11479} = 0.56$

4. What is the probability that a randomly selected victim was male and had an accidental death?
Answer: P(male and accidental)=$\frac{6457}{14099} = 0.46$.

General multiplication rule

Recall that $P(A|B) = \frac{P(AandB)}{P(B)}$. (1)
Multiply both sides of (1) by P(B), we can get
the general multiplication rule: $P(AandB) = P(A|B)P(B)$. (2)
Switch A and B, we can also get $P(AandB) = P(B|A)P(A)$.

This rule can be used to compute P(A and B) if we know P(A|B) and P(B).

Multiplication rule for independent events

Two events A and B are independent if
$P(B|A) = P(B)$ or $P(A|B) = P(A)$, (3)
Two events A and B are independent means the probability of one event does not depend on the occurence of the other event.

If A and B are independent, we get
multiplication rule for independent events:
$P(A\text{and } B) = P(A)P(B)$. (derived from (2) and (3)).

More general, if $A_1, A_2, ...A_k$ are independent,
$P(A_1\text{and}A_2....\text{and}A_k) = P(A_1A_2...A_k) = P(A_1)P(A_2)...P(A_k)$.

Example: multiplication rule for independent events

Tom will throw darts at a target. Suppose the probability is 0.10 that he will hit the target in each throw.
Let A_1 = hit in 1st throw,
A_2 = hit in 2nd throw.
Find
$P(\text{ both are hits}) = P(A_1 and A_2) = P(A_1)P(A_2) = 0.1 * 0.1 = 0.01$.
$P(\text{both are misses}) = P(\text{1st is miss}) * P(\text{2nd is miss}) = 0.9 * 0.9 = 0.81$.

Suppose now Tom will throw 5 times.
Then P(all 5 are hits)=$0.1^5 = 0.00001$.
Here it is reasonable to assume the results are independent.

Exercise: Multiplication rule for independent events

An unfair coin has a probability of 0.4 of landing heads. Flip this coin 4 times. Find the probability it will land

1). all four heads.
$P(HHHH) = 0.4^4 = 0.0256$
2) two heads followed by two tails.
$P(HHTT) = 0.4 * 0.4 * 0.6 * 0.6 = 0.0576.$

It is reasonable to assume the results from each flip are independent.

Check independence

Note if we know P(A), P(B) and also know A and B are independent, we can use
P(A and B)= P(A)P(B) to compute P(A and B).

If we know P(A), P(B) and P(A and B), we can check if P(A an B)= P(A) P(B) holds to answer if A and B are independent.

We can also check if $P(A|B) = P(A)$ (or $P(B|A) = P(B)$) holds to answer if A and B are independent.

1. The following table contains 600 wafers classified by lot and whether they conform to specifications.

Lot	Conforming	Nonconforming
A	88	12
B	165	35
C	260	40

Randomly choose a wafer.

1). Find the probability the wafer is from lot A.

$P = \frac{88+12}{600} = 0.167$

2). Find the probability the wafer is conforming.

$P = \frac{88+165+260}{600} = 0.855$

3). Find the probability the wafer is from lot A and is conforming.

$P = \frac{88}{600} = 0.147$

Exercise continued: check if two events are independent

4). Let E_1 be the event the wafer is from lot A and E_2 be the event the wafer is conforming. Are E_1 and E_2 independent? Why or why not?

Note $P(E_1 and E_2) = 0.147, P(E_1) = 0.167, P(E_2) = 0.855$,
(these are computed in part 1), 2) and 3)) and
$0.147 \neq 0.167 * 0.855 = 0.143$, so E_1 and E_2 are not independent.

Here we check if $P(E_1 and E_2) = P(E_1)P(E_2)$ to answer if E_1 and E_2 are independent.