



COLUMBIA
BUSINESS
SCHOOL

Probability Basics

Fall 2001
B6014: Managerial Statistics

Professor Paul Glasserman
403 Uris Hall

Terminology and basic concepts

1. In discussing probability, the **sample space** is the set of possible outcomes. Examples:

- The sample space for a coin toss is {Head, Tail}.
- The sample space for the roll of a die is {1, 2, 3, 4, 5, 6}.
- What is the sample space for the roll of 2 dice? For the sum of 2 dice?
- What is the sample space for today's IBM closing price?

2. An **event** is a subset of the sample space. Examples:

- The die comes up 3.
- The die comes up even, which corresponds to the set {2, 4, 6}.

We denote generic events by the letters A, B , etc.

3. The **probability** of an event A is a number $P(A)$ between 0 and 1. The probability of the entire sample space is 1.

- The *frequentist* interpretation of probability interprets $P(A)$ to mean the long-run proportion of times A occurs; e.g., the proportion of die-rolls that come up 3 out of many tries.
- The *subjectivist* interpretation takes $P(A)$ to be a subjective assessment of the chance that A will occur; e.g., the probability that a particular candidate will win an election.

Both interpretations are useful in practice.

4. We can combine events to get new events. The event (A or B) occurs if either A or B occurs. The event (A and B) occurs if both A and B occur.

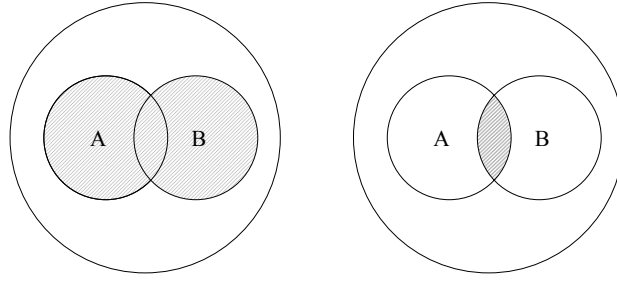


Figure 1: The shaded area on the left is the event $(A \text{ or } B)$. The shaded area on the right is the event $(A \text{ and } B)$. The large circle containing both events A, B represents the sample space.

5. We sometimes use the notation $(A \cup B)$ [the *union* of A and B] for $(A \text{ or } B)$. We sometimes use the notation $(A \cap B)$ [the *intersection* of A and B] for $(A \text{ and } B)$. See Figure 1.
6. The **complement** of an event A , denoted by $\bar{A}$ is the event that A does not occur. Clearly, $P(\bar{A}) = 1 - P(A)$.
7. We say that two events are **mutually exclusive** if they do not overlap.
 - (a) {die comes up 3} and {die comes up even} are mutually exclusive.
 - (b) {die comes up less than 3} and {die comes up even} are not mutually exclusive, because {die comes up 2} is contained in both.

Saying that A and B are mutually exclusive is equivalent to saying that the intersection $(A \text{ and } B)$ is empty.

8. The **addition rule** for mutually exclusive events is the following: If A and B are mutually exclusive, then

$$P(A \text{ or } B) = P(A) + P(B).$$

Let's try this on the die example above. Clearly, $P(\text{die comes up 3}) = 1/6$ and $P(\text{die comes up even}) = 1/2$, so

$$P(\text{die comes up 3}) + P(\text{die comes up even}) = \frac{1}{6} + \frac{1}{2} = \frac{2}{3}.$$

On the other hand,

$$P(\text{die comes up 3 or even}) = P(\text{die comes up 2,3,4 or 6}) = \frac{4}{6} = \frac{2}{3}.$$

9. If A and B are not mutually exclusive, we need to subtract the overlap:

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B).$$

Let's try the other die example. Define

$$\begin{aligned} A &= \{\text{die comes up less than 3}\} \\ B &= \{\text{die comes up even}\}; \end{aligned}$$

then

$$\begin{aligned}A \text{ or } B &= \{\text{die comes up } 1, 2, 4 \text{ or } 6\} \\A \text{ and } B &= \{\text{die comes up } 2\}.\end{aligned}$$

We can now verify that

$$P(A \text{ or } B) = \frac{2}{3}$$

and

$$P(A) + P(B) - P(A \text{ and } B) = \frac{2}{6} + \frac{3}{6} - \frac{1}{6} = \frac{2}{3},$$

as expected.

Conditional Probability

1. The **conditional probability** of A **given** B is the probability that A occurs given that B is known to occur. This is denoted by $P(A|B)$.
2. Before discussing conditional probability more formally, we look at an example. The following table gives hypothetical percentages of people that are male smokers, female smokers, male nonsmokers and female nonsmokers:

	Smoker	Nonsmoker
Male	.12	.37
Female	.06	.45

Given that a person is male, what is the probability that he smokes? Given that a person is a smoker, what is the probability she is female?

3. To address these questions, we define the conditional probability of A given B , as follows:

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

When we condition on B , we shrink the sample space down to B . Thus, $P(A|B)$ is the probability of the part of A that lies in B , relative to the probability of all of B .

We can now answer the questions posed above, using the information in the table:

$$P(\text{smoker}|\text{male}) = \frac{P(\text{smoker and male})}{P(\text{male})} = \frac{.12}{.49} = .24$$

Similarly,

$$P(\text{female}|\text{smoker}) = \frac{P(\text{female and smoker})}{P(\text{smoker})} = \frac{.06}{.18} = .33$$

Thus, 24% of males are smokers and 33% of smokers are female.

4. Examples for discussion:

- (a) A gun-control group mails questionnaires to 1000 randomly chosen people in a community. 350 people reply; of those 250 say they support gun control. Can we conclude that the proportion of people in the community supporting gun control is roughly 5/7? [Not really. This is the proportion among those who responded. Equivalently, 5/7 is the *conditional* probability that a randomly selected individual supports gun control, *given* that the individual is a respondent. This need not be the same as the probability that a randomly selected individual from the entire population supports gun control.]
 - (b) Given that a woman's first child is a boy, what is the probability that her next child will be a boy? [Presumably, 1/2. The conditional probability that the second child is a boy should be the same as the unconditional probability.]
 - (c) In the late 1980s, the state of Massachusetts considered requiring an HIV test of all couples applying for a marriage license. There are two types of tests: an ELISA test and a Western Blot. With the testing technology available at the time, the ELISA would miss 10% of samples known to be infected, and incorrectly give a positive result in 5% of infected samples. The Western Blot would miss 5% of infected samples and incorrectly give a positive result in 5% of infected cases. Under the proposed plan, the ELISA test would have been applied first; any positives would then be subjected to a Western Blot. The prevalence of the AIDS virus among the "low-risk" population of Massachusetts is considered 0.2%. What are the expected results of administering these tests to 100,000 low-risk individuals? In particular, what proportion of people that test positive on the ELISA test would actually have the virus? What proportion of those that test positive on both the ELISA and the Western Blot would have the virus?
 - (d) A woman tells you she has two children. Then she tells you at least one is a boy. What is the probability that the other is also a boy? [It's not 1/2!]
5. Two events A and B are **independent** if $P(A|B) = P(A)$; i.e., if event B carries no information about event A . This definition is symmetric in the sense that if $P(A|B) = P(A)$ then $P(B|A) = P(B)$. (Why? Substitute the definition of conditional probability.)
6. The **multiplication rule for independent events** is this:

$$P(A \text{ and } B) = P(A)P(B).$$

This does not hold if A and B are not independent. The general multiplication rule is

$$P(A \text{ and } B) = P(A)P(B|A) = P(A|B)P(B).$$

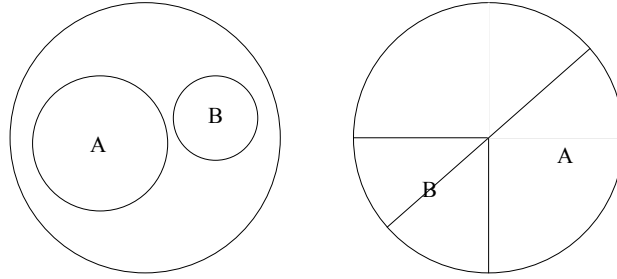


Figure 2: In the figure on the left, A and B are mutually exclusive. In the figure on the right, they are independent: A is a semicircle with $P(A) = 1/2$, and B is a quarter-circle with $P(B) = 1/4$. The intersection (A and B) is a wedge of area $1/8$ because it is half of B . So, we have $P(A|B) = P(A \text{ and } B)/P(B) = (1/8)/(1/4) = 1/2$, which is the same as the unconditional probability $P(A)$. Knowing that we're in B doesn't make it more likely or less likely that we're also in A .

7. Question: Suppose A and B are events. Which of the following statements is correct:

- (a) If A, B are independent then they are mutually exclusive.
- (b) If A, B are independent then they are not mutually exclusive.
- (c) If A, B are independent then they may or may not be mutually exclusive.

[The answer is (b). If A and B were mutually exclusive they would not overlap, so $P(A|B) = 0$ whereas $P(A) \neq 0$. If $P(A|B) \neq P(A)$, then A and B could not be independent.]

8. Figure 2 gives a graphical illustration of the distinction between mutually exclusive and independent events. The figure is explained in the caption. However, the graphical representation of independence is not very helpful. It's better to think of one event as independent of another if neither contains information about the other.
9. A tabular interpretation of independence is sometimes helpful. Replace the previous table with the following:

	Smoker	Nonsmoker
Married	.12	.48
Unmarried	.08	.32

In this table, whether or not an individual smokes is independent of marital status. For example,

$$P(\text{smoker}|\text{married}) = \frac{P(\text{smoker and married})}{P(\text{married})} = \frac{.12}{.12 + .48} = .20$$

and the unconditional probability is $P(\text{smoker}) = 0.12 + 0.08 = 0.20$. A married person has the same 20% chance of being a smoker as any other person. Similarly,

$$P(\text{unmarried}|\text{smoker}) = \frac{P(\text{unmarried and smoker})}{P(\text{smoker})} = \frac{.08}{.20} = .40,$$

and $P(\text{unmarried}) = 0.08 + 0.32 = 0.40$.

10. A current area of some controversy is DNA fingerprinting. Sample DNA from a crime scene is compared to DNA taken from a suspect. If there is a good match, prosecutors claim astronomical odds in favor of guilt. The controversy lies in the right way to compute those odds. Should, say, the probability of a blue-eyes gene be multiplied by the probability of a blonde-hair gene in computing the odds of a match? How do possible patterns within ethnic groups affect this calculation? [Multiplying probabilities is appropriate only if the attributes occur independently of each other. Patterns within ethnic groups suggest that attributes are not independent. For example, the conditional probability $P(\text{blue eyes}|\text{blonde hair})$ is probably much greater than the unconditional probability $P(\text{blue eyes})$.

Summary of Key Ideas

1. *Addition rule:*

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B).$$

If A and B are mutually exclusive then $P(A \text{ and } B) = 0$ and

$$P(A \text{ or } B) = P(A) + P(B).$$

2. *Multiplication rule:*

$$P(A \text{ and } B) = P(A)P(B|A) = P(B)P(A|B).$$

3. *Conditional probability:*

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}.$$

4. *Independence.* A and B are independent if

$$P(A \text{ and } B) = P(A)P(B);$$

equivalently, $P(A|B) = P(A)$.