$P(A^c|B) = 1 - P(A|B)$

# IES302 2011/1 Part I.3 Dr.Prapun

**Example 6.24.** Medical Diagnostic: Because a new medical procedure has been shown to be effective in the early detection of an illness, a medical screening of the population is proposed. The probability that the test correctly identifies someone with the illness as positive is 0.99, and the probability that the test correctly identifies someone without the illness as negative is 0.95. The incidence of the illness in the general population is 0.0001. You take the test, and the result is positive. What is the probability that you have the illness? [12, Ex. 2-37]

$P(T_p|D) = 0.99$ $P(T_p^c|D) = 0.01$ $P(T_p^c|D^c) = 0.95$ $P(D) = 10^{-4}$ $P(T_p|D^c) = 0.05$

Let $D$ denote the event that you have the illness, and let $T_P$ denote the event that the test signals positive.

$$P(D|T_p) = \frac{P(D \cap T_p)}{P(T_p)} = \frac{P(T_p|D)P(D)}{P(T_p)} = \frac{0.99 \times 10^{-4}}{0.0501} \approx 0.002$$

$$P(T_p) = P(T_p|D)P(D) + P(T_p|D^c)P(D^c)$$
$$= 0.99 \times 10^{-4} + (0.05)(1 - 10^{-4})$$
$$= 0.0501$$

**Example 6.25.** Bayesian networks are used on the Web sites of high-technology manufacturers to allow customers to quickly diagnose problems with products. An oversimplified example is presented here.

A printer manufacturer obtained the following probabilities from a database of test results. Printer failures are associated with three types of problems: hardware, software, and other (such as connectors), with probabilities 0.1, 0.6, and 0.3, respectively. The prob-

$P(H) = 0.1$
$P(S) = 0.6$
$P(O) = 0.3$

printer failure

$P(F|H) = 0.9$

$P(F|S) = 0.2$

$P(F|O) = 0.5$

ability of a printer failure given a hardware problem is 0.9, given a software problem is 0.2, and given any other type of problem is 0.5. If a customer enters the manufacturers Web site to diagnose a printer failure, what is the most likely cause of the problem?

Let the events $H$, $S$, and $O$ denote a hardware, software, or other problem, respectively, and let $F$ denote a printer failure.

which one is the largest

$$P(H|F) = \frac{P(H \cap F)}{P(F)} = \frac{P(F|H)P(H)}{P(F)} = \frac{0.9 \times 0.1}{0.36} = \frac{1}{4} = 0.25$$

$$P(S|F) = \frac{P(S \cap F)}{P(F)} = \frac{P(F|S)P(S)}{P(F)} = \frac{.2 \times .6}{.36} = \frac{1}{3} \approx 0.33$$

$$P(O|F) = \frac{P(O \cap F)}{P(F)} = \frac{P(F|O)P(O)}{P(F)} = \frac{.5 \times .3}{0.36} = \frac{5}{12} \approx 0.4167$$

The most likely cause is in the "other" category.

$$P(F) = P(F \cap H) + P(F \cap S) + P(F \cap O) = P(F|H)P(H) + P(F|S)P(S) + P(F|O)P(O)$$
$$= .9 \times .1 + .2 \times 0.6 + .5 \times .3 = 0.36$$

## 6.2 Event-based Independence

Plenty of random things happen in the world all the time, most of which have nothing to do with one another. If you toss a coin and I roll a dice, the probability that you get heads is 1/2 regardless of the outcome of my dice. Events that in this way are unrelated to each other are called *independent*.

**6.26.** Sometimes our definition for independence above does not agree with the everyday-language use of the word "independence". Hence, many authors use the term "statistically independence" to distinguish it from other definitions.

In general, $P(A \cap B) = P(B|A)P(A) = P(A|B)P(B)$

**Definition 6.27. Two events** $A$, $B$ are called (statistically) ***independent*** if

$$P(A \cap B) = P(A)P(B) \tag{8}$$

- Notation: $A \perp\!\!\!\perp B$ in (Recall that $A \perp B$ means $A$ and $B$ are disjoint)
- Read "$A$ and $B$ are independent" or "$A$ is independent of $B$"
- We call (8) the ***multiplication rule*** for probabilities.

printer failure

$P(F|H) = 0.9$

$P(F|S) = 0.2$

$P(F|O) = 0.5$

ability of a printer failure given a hardware problem is 0.9, given a software problem is 0.2, and given any other type of problem is 0.5. If a customer enters the manufacturers Web site to diagnose a printer failure, what is the most likely cause of the problem?

Let the events $H$, $S$, and $O$ denote a hardware, software, or other problem, respectively, and let $F$ denote a printer failure.

which one is the largest

$$P(H|F) = \frac{P(H\cap F)}{P(F)} = \frac{P(F|H)P(H)}{P(F)} = \frac{0.9\times 0.1}{0.36} = \frac{1}{4} = 0.25$$

$$P(S|F) = \frac{P(S\cap F)}{P(F)} = \frac{P(F|S)P(S)}{P(F)} = \frac{.2\times .6}{.36} = \frac{1}{3} \approx 0.33$$

$$P(O|F) = \frac{P(O\cap F)}{P(F)} = \frac{P(F|O)P(O)}{P(F)} = \frac{.5\times .3}{0.36} = \frac{5}{12} \approx 0.4167$$

The most likely cause is in the "other" category.

$$P(F) = P(F\cap H) + P(F\cap S) + P(F\cap O) = P(F|H)P(H) + P(F|S)P(S) + P(F|O)P(O)$$

$$= .9\times .1 + .2\times 0.6 + .5\times .3 = 0.36$$

## 6.2 Event-based Independence

Plenty of random things happen in the world all the time, most of which have nothing to do with one another. If you toss a coin and I roll a dice, the probability that you get heads is 1/2 regardless of the outcome of my dice. Events that in this way are unrelated to each other are called *independent*.

**6.26.** Sometimes our definition for independence above does not agree with the everyday-language use of the word "independence". Hence, many authors use the term "statistically independence" to distinguish it from other definitions.

In general, $P(A\cap B) = P(B|A)P(A) = P(A|B)P(B)$

**Definition 6.27. Two events** $A$, $B$ are called (statistically) ***independent*** if

$$P(A\cap B) = P(A)P(B) \tag{8}$$

independent

- Notation: $A \perp\!\!\!\perp B$

(Recall that $A \perp B$ means $A$ and $B$ are disjoint)

- Read "$A$ and $B$ are independent" or "$A$ is independent of $B$"
- We call (8) the ***multiplication rule*** for probabilities.

$$P(A\cap B) = P(A)P(B) \rightarrow P(A|B) = \frac{P(A\cap B)}{P(B)} = P(A)$$

$$\hookrightarrow P(B|A) = \frac{P(A\cap B)}{P(A)} = P(B)$$

- If two events are not independent, they are ***dependent***. If two events are dependent, the probability of one changes with the knowledge of whether the other has occurred.

- In classical probability, this is equivalent to

$$|A \cap B||\Omega| = |A||B|.$$

**6.28.** Intuition: Again, here is how you should think about independent events: "If one event has occurred, the probability of the other does not change."

$$P(A|B) = P(A) \quad \text{and} \quad P(B|A) = P(B). \tag{9}$$

In other words, "the unconditional and the conditional probabilities are the same". We can almost use (9) as the definitions for independence. However, we use (8) instead because it also works with events whose probabilities are zero. In fact, in 6.30, we show how (9) can be used to define independence with extra condition that deals with the case when zero probability is involved.

**Example 6.29.** (Slide) [24, Ex. 5.4] Which of the following pairs of events are independent?

(a) The card is $\overset{A}{\text{a club}}$, and the card $\overset{B}{\text{is black}}$.

(b) The card is $\overset{A}{\text{a king}}$, and the card $\overset{B}{\text{is black}}$.

(a) $P(A) = \frac{13}{52} = \frac{1}{4}$

$P(B) = \frac{26}{52} = \frac{1}{2}$

$P(A \cap B) = P(A) = \frac{1}{4} \neq P(A)P(B)$

Not independent

(b) $P(A) = \frac{4}{52} = \frac{1}{13}$

$P(B) = \frac{1}{2}$

$P(A \cap B) = \frac{2}{52} = \frac{1}{26} = P(A)P(B)$

$A \perp\!\!\!\perp B$

**6.30.** Two events $A$, $B$ *with positive probabilities* are independent if and only if $P(B|A) = P(B)$, which is equivalent to $P(A|B) = P(A)$.

When $A$ and/or $B$ has zero probability, $A$ and $B$ are automatically independent.

suppose $P(A) = 0 \Rightarrow P(A \cap B) = 0 \Rightarrow \underbrace{P(A \cap B)}_{0} = \underbrace{P(A)}_{0} P(B) \Rightarrow A \perp\!\!\!\perp B$ for any B

**6.31.** When $A$ and $B$ have nonzero probabilities, the following statements are equivalent:

1) $A \perp\!\!\!\perp B$
2) $P(A \cap B) = P(A)P(B)$
3) $P(A|B) = P(A)$
4) $P(B|A) = P(B)$
5) $A \perp\!\!\!\perp B^c \rightarrow P(A \cap B^c) = P(A)P(B^c)$
6) $A^c \perp\!\!\!\perp B$
7) $A^c \perp\!\!\!\perp B^c$

**6.32.** If $A$ and $B$ are independent events, then so are $A$ and $B^c$, $A^c$ and $B$, and $A^c$ and $B^c$. By interchanging the roles of $A$ and $A^c$ and/or $B$ and $B^c$, it follows that if any one of the four pairs is independent, then so are the other three. [7, p.31]

In fact, the following four statements are equivalent:

$$A \perp\!\!\!\perp B, \quad A \perp\!\!\!\perp B^c, \quad A^c \perp\!\!\!\perp B, \quad A^c \perp\!\!\!\perp B^c.$$

**Example 6.33.** If $P(A|B) = 0.4$, $P(B) = 0.8$, and $P(A) = 0.5$, are the events $A$ and $B$ independent? [12]

$P(A|B) \neq P(A) \Rightarrow$ not independent

**6.34.** Keep in mind that **independent and disjoint** are ***not synonyms***. In some contexts these words can have similar meanings, but this is not the case in probability.

- If two events cannot occur at the same time (they are disjoint), are they independent? At first you might think so. After all, they have nothing to do with each other, right? Wrong! They have a lot to do with each other. If one has occurred, we know for certain that the other cannot occur. [16, p 12]

- To check whether $A$ and $B$ are disjoint, you only need to look at the sets themselves and see whether they have shared

$A \perp\!\!\!\perp B \Rightarrow P(A \cap B) = P(A)P(B)$



$A \perp B \Rightarrow P(A \cup B) = P(A) + P(B)$

element(s). This can be answered without knowing probabilities.

To check whether $A$ and $B$ are independent, you need to look at the probabilities $P(A)$, $P(B)$, and $P(A \cap B)$.

- Reminder: If events $A$ and $B$ are disjoint, you calculate the probability of the union $A \cup B$ by adding the probabilities of $A$ and $B$. For independent events $A$ and $B$ you calculate the probability of the intersection $A \cap B$ by multiplying the probabilities of $A$ and $B$.

**Example 6.35.** Experiment of flipping a fair coin twice. $\Omega = \{HH, HT, TH, TT\}$. Define event $A$ to be the event that the first flip gives a H; that is $A = \{HH, HT\}$. Event $B$ is the event that the second flip gives a H; that is $B = \{HH, TH\}$. Note that even though the events $A$ and $B$ are not disjoint, they are independent.

$A \cap B = \{HH\} \neq \emptyset$ $\Rightarrow A \not\perp B$

$\underbrace{P(A)}_{\frac{1}{2}} \underbrace{P(B)}_{\frac{1}{2}} = \underbrace{P(A \cap B)}_{\frac{1}{4}}$ $\Rightarrow A \perp\!\!\!\perp B$

**Example 6.36. *Prosecutor's fallacy*:** In 1999, a British jury convicted Sally Clark of murdering two of her children who had died suddenly at the ages of 11 and 8 weeks, respectively. A pediatrician called in as an expert witness claimed that the chance of having two cases of infant sudden death syndrome, or "cot deaths," in the same family was 1 in 73 million. There was no physical or other evidence of murder, nor was there a motive. Most likely, the jury was so impressed with the seemingly astronomical odds against the incidents that they convicted. But where did the number come from? Data suggested that a baby born into a family similar to the Clarks faced a 1 in 8,500 chance of dying a cot death. Two cot deaths in the same family, it was argued, therefore had a probability of $(1/8,500)^2$ which is roughly equal to 1/73,000.000.

Did you spot the error? I hope you did. The computation assumes that successive cot deaths in the same family are *independent* events. This assumption is clearly questionable, and even a person without any medical expertise might suspect that genetic factors play a role. Indeed, it has been estimated that if there is one cot death, the next child faces a much larger risk, perhaps around 1/100. To find the probability of having two cot deaths in the same family, we should thus use conditional probabilities and arrive at the computation $1/8,500 \times 1/100$, which equals 1/850,000. Now, this is still a small number and might not have made the jurors judge differently. But what does the probability 1/850,000 have to do with Sallys guilt? Nothing! When her first child died, it was certified to have been from natural causes and there was no suspicion of foul play. The probability that it would happen again without foul play was 1/100, and if that number had been presented to the jury, Sally would not have had to spend three years in jail before the verdict was finally overturned and the expert witness (certainly no expert in probability) found guilty of "serious professional misconduct."

You may still ask the question what the probability 1/100 has to do with Sallys guilt. Is this the probability that she is innocent? Not at all. That would mean that 99% of all mothers who experience two cot deaths are murderers! The number 1/100 is simply the probability of a second cot death, which only means that among all families who experience one cot death, about 1% will suffer through another. If probability arguments are used in court cases, it is very important that all involved parties understand some basic probability. In Sallys case, nobody did.

References: [11, 118–119] and [16, 22–23].

**Definition 6.37. Three events** $A_1, A_2, A_3$ are independent if and only if

$$\begin{aligned} P(A_1 \cap A_2) &= P(A_1)\, P(A_2) && A_1 \perp\!\!\!\perp A_2 \\ P(A_1 \cap A_3) &= P(A_1)\, P(A_3) && A_1 \perp\!\!\!\perp A_3 \\ P(A_2 \cap A_3) &= P(A_2)\, P(A_3) && A_2 \perp\!\!\!\perp A_3 \\ P(A_1 \cap A_2 \cap A_3) &= P(A_1)\, P(A_2)\, P(A_3) \end{aligned}$$

*Remarks*:

(a) When the first three equations hold, we say that the three events are ***pairwise independent***.

(b) We may use the term "mutually independence" to further emphasize that we have "independence" instead of "pairwise independence".

**Definition 6.38.** The events $A_1, A_2, \ldots, A_n$ are ***independent*** if and only if for any subcollection $A_{i_1}, A_{i_2}, \ldots, A_{i_k}$,

$n=4$

$$P\left(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}\right) = P\left(A_{i_1}\right) \times P\left(A_{i_2}\right) \times \cdots \times P\left(A_{i_n}\right).$$

$A_1, A_2, A_3, A_4$

$A_1 \perp\!\!\!\perp A_2$

$A_1 \perp\!\!\!\perp A_3$

$A_1 \perp\!\!\!\perp A_4$

$A_2 \perp\!\!\!\perp A_3$

$A_2 \perp\!\!\!\perp A_4$

$A_3 \perp\!\!\!\perp A_4$

- Note that part of the requirement is that

$$P\left(A_1 \cap A_2 \cap \cdots \cap A_n\right) = P\left(A_1\right) \times P\left(A_2\right) \times \cdots \times P\left(A_n\right).$$

Therefore, if someone tells us that the events $A_1, A_2, \ldots, A_n$ are independent, then one of the properties that we can conclude is that

$$P\left(A_1 \cap A_2 \cap \cdots \cap A_n\right) = P\left(A_1\right) \times P\left(A_2\right) \times \cdots \times P\left(A_n\right).$$

$P(A_1 \cap A_2 \cap A_3)$
$= P(A_1)P(A_2)P(A_3)$

- Equivalently, this is the same as the requirement that

$$P\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} P\left(A_j\right) \quad \forall J \subset [n] \text{ and } |J| \geq 2$$

$P(A_2 \cap A_3 \cap A_4)$
$= P(A_2)P(A_3)P(A_4)$

$P(A_1 \cap A_3 \cap A_4)$
$= P(A_1)P(A_3)P(A_4)$

- Note that the case when $j = 1$ automatically holds. The case when $j = 0$ can be regard as the $\emptyset$ event case, which is also trivially true.

$P(A_1 \cap A_2 \cap A_4)$
$= P(A_1)P(A_2)P(A_4)$

## 6.3 Bernoulli Trials

**Example 6.39.** Consider the following random experiments

(a) Flip a coin 10 times. We are interested in the number of heads obtained.

$P(A_1 \cap A_2 \cap A_3 \cap A_4)$
$= P(A_1)P(A_2)P(A_3)P(A_4).$

(b) Of all bits transmitted through a digital transmission channel, 10% are received in error. We are interested in the number of bits in error in the next five bits transmitted.

(c) A multiple-choice test contains 10 questions, each with four choices, and you guess at each question. We are interested in the number of questions answered correctly.

These examples illustrate that a general probability model that includes these experiments as particular cases would be very useful.

**Example 6.40.** Each of the random experiments in Example 6.39 can be thought of as consisting of a series of repeated, random trials. In all cases, we are interested in the number of trials that meet a specified criterion. The outcome from each trial either meets the criterion or it does not; consequently, each trial can be summarized as resulting in either a success or a failure.

**Definition 6.41.** A ***Bernoulli trial*** involves performing an experiment once and noting whether a particular event $A$ occurs.

The outcome of the Bernoulli trial is said to be

(a) a "success" if $A$ occurs and    1    H    die

(b) a "failure" otherwise.    0    T    not die

We may view the outcome of a single Bernoulli trial as the outcome of a toss of an unfair coin for which the probability of heads (success) is $p = P(A)$ and the probability of tails (failure) is $1 - p$.

- The labeling ("success" and "failure") is not meant to be literal and sometimes has nothing to do with the everyday meaning of the words. We can just as well use A and B or 0 and 1.

**Example 6.42.** Examples of Bernoulli trials: Flipping a coin, deciding to vote for candidate A or candidate B, giving birth to a boy or girl, buying or not buying a product, being cured or not being cured, even dying or living are examples of Bernoulli trials.

- Actions that have multiple outcomes can also be modeled as Bernoulli trials if the question you are asking can be phrased in a way that has a yes or no answer, such as "Did the dice land on the number 4?" or "Is there any ice left on the North Pole?"

**Definition 6.43.** (Independent) ***Bernoulli Trials*** = a Bernoulli trial is repeated many times.

(a) It is usually assumed that the trials are independent. This implies that the outcome from one trial has no effect on the outcome to be obtained from any other trial.

(b) Furthermore, it is often reasonable to assume that the probability of a success in each trial is constant.

An outcome of the complete experiment is a sequence of successes and failures which can be denoted by a ***sequence of ones and zeroes***.

$\{\overbrace{HHTH \cdots T}^{n}\}$

$\vdots$

**Example 6.44.** If we toss ***unfair coin*** $n$ times, we obtain the space $\Omega = \{H, T\}^n$ consisting of $2^n$ elements of the form $(\omega_1, \omega_2, \ldots, \omega_n)$ where $\omega_i$ = H or T.

There are 10 ways to do this.

S F F S S

$\underbrace{\underline{S}\ \underline{S}\ \underline{F}\ \underline{F}\ \underline{S}}$

probability of this particular case is $p^3(1-p)^2$

**Example 6.45.** What is the probability of two failures and three successes in five Bernoulli trials with success probability $p$.

We observe that the outcomes with three successes in five trials are 11100, 11010, 11001, 10110, 10101, 10011, 01110, 01101, 01011, and 00111. We note that the probability of each outcome is a product of five probabilities, each related to one Bernoulli trial. In outcomes with three successes, three of the probabilities are $p$ and the other two are $1 - p$. Therefore, each outcome with three successes has probability $(1-p)^2 p^3$. There are 10 of them. Hence, the total probability is $10(1-p)^2 p^3$.

$P[R_1 = S \cap R_2 = S \cap R_3 = F \cap R_4 = F \cap R_5 = S] = P[R_1 = S]\, P[R_2 = S]\, P[R_3 = F]\, P[R_4 = F]\, P[R_5 = S]$

**Exercise 6.46** (F2011)**.** Kakashi and Gai are eternal rivals. Kakashi is a little stronger than Gai and hence for each time that they fight, the probability that Kakashi wins is 0.55. In a competition, they fight $n$ times (where $n$ is odd). We will assume that the results of the fights are independent. The one who wins more will win the competition.

Suppose $n = 3$, what is the probability that Kakashi wins the competition.

Conclusion:

The probability of exactly $\overset{3}{\overset{\shortparallel}{n_1}}$ successes in $\overset{5}{\overset{\shortparallel}{n}}$ $(= n_0 + n_1)$ Bernoulli trials is given by

$$\binom{n}{n_1} p^{n_1} (1-p)^{n-n_1}$$

$$\binom{n}{k} x^k y^{n-k}$$

$$\binom{5}{3} p^3 (1-p)^{5-3}$$