# ECE 361 Probability for Engineers (Fall, 2016)
# Lecture 2a

## §1.3 Conditional probability

### Using conditional probability for modeling

Generalizing the arguments in the above example we identify the following multiplication rule:

**Multiplication rule.**

$$\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2)\cdots\mathbb{P}\left(A_n \middle| \bigcap_{i=1}^{n-1} A_i\right) \tag{1}$$

The proof is given by cancelling terms in the expression:

$$\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \mathbb{P}(A_1)\frac{\mathbb{P}(A_1 \cap A_2)}{\mathbb{P}(A_1)}\frac{\mathbb{P}(A_1 \cap A_2 \cap A_3)}{\mathbb{P}(A_1 \cap A_2)}\cdots\frac{\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right)}{\mathbb{P}\left(\bigcap_{i=1}^{n-1} A_i\right)} \tag{2}$$

We give two examples illustrating the power of this rule.

**Example.** Three cards are drawn without replacement from an ordinary 52 card deck. What is the probability none of the three cards is a heart ($A$)? Define $A_i$ as the event that card $i$ is not a heart, for $i \in [3]$. Then $A = A_1 \cap A_2 \cap A_3$, and

$$\mathbb{P}(A) = \mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2) = \frac{39}{52}\frac{38}{51}\frac{37}{50}. \tag{3}$$

**Example.** A class consisting of 4 graduate (G) and 12 undergraduate (UG) students is divided at random into 4 groups of 4. What is the probability that each group contains a graduate student? Label the four graduate students as $1, 2, 3, 4$. Define events:

$$\begin{aligned} A_1 &= \text{students 1 and 2 in different groups} \\ A_2 &= \text{students 1 and 2 and 3 in different groups} \\ A_3 &= \text{students 1 and 2 and 3 and 4 in different groups} \end{aligned} \tag{4}$$

Note $A_3$ is the event of interest, and that we can write $A_3 = A_1 \cap A_2 \cap A_3$ since $A_3 \subset A_2 \subset A_1$:

$$\begin{aligned} \mathbb{P}(A_3) &= \mathbb{P}(A_1 \cap A_2 \cap A_3) \\ &= \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2) \\ &= \frac{12}{15}\frac{8}{14}\frac{4}{13}. \end{aligned} \tag{5}$$

These fractions can be obtained by arguing from first principles. For example, $\mathbb{P}(A_2|A_1) = 8/14$ since there are eight open seats in the two groups without graduate students out of fourteen seats total.

## §1.4 Total probability theorem and Bayes' rule

**Total probability theorem.** Let $A_1, \ldots, A_n$ be events that partition the sample space $\Omega$. Then for any event $B$

$$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(A_i \cap B) = \sum_{i=1}^{n} \mathbb{P}(B|A_i)\mathbb{P}(A_i). \tag{6}$$

The importance of this theorem cannot be overstated – it provides a “divide and conquer” approach to analyzing the probabilities of “complex” events $B$, by breaking them down into multiple simpler events $\mathbb{P}(B|A_i)$ where the constituent events $A_i$ have known probabilities. See Fig. 1.13 for a visualization.

**Example: chess tournament.** Your win probability in a chess tournament is 0.3 against half the players (type 1), 0.4 against a quarter of the players (type 2), and 0.5 against the remaining quarter (type 3). You play against a random opponent – what is the probability of winning ($B$)?

Set $A_i$ to be the event that the opponent is of type $i \in [3]$, with

$$\begin{array}{ccc} \mathbb{P}(A_1) = 0.5 & \mathbb{P}(A_2) = 0.25 & \mathbb{P}(A_3) = 0.25 \\ \mathbb{P}(B|A_1) = 0.3 & \mathbb{P}(B|A_2) = 0.4 & \mathbb{P}(B|A_3) = 0.5 \end{array} \tag{7}$$

Then:

$$\mathbb{P}(B) = \mathbb{P}(B|A_1)\mathbb{P}(A_1) + \mathbb{P}(B|A_2)\mathbb{P}(A_2) + \mathbb{P}(B|A_3)\mathbb{P}(A_3) = 0.3 \times 0.5 + 0.4 \times 0.25 + 0.5 \times 0.25 = 0.375. \tag{8}$$

**Example: Monty Hall.** Quoting from Rosenhouse [2]:

> You are shown three identical doors. Behind one of them is a car. The other two conceal goats. You are asked to choose, but not open, one of the doors. After doing so, Monty, who knows where the car is, opens one of the two remaining doors. He always opens a door he knows to be incorrect, and randomly chooses which door to open when he has more than one option (which happens on those occasions where your initial choice conceals the car). After opening an incorrect door, Monty gives you the option of either switching to the other unopened door or sticking with your original choice. You then receive whatever is behind the door you choose. What should you do?

Let $A$ be the event where you win by sticking with your original choice, and let $B$ be the event where you win by switching your original choice. What are $\mathbb{P}(A)$ and $\mathbb{P}(B)$? Clearly $\mathbb{P}(A) = 1/3$. It remains to compute $\mathbb{P}(B)$. Define $C$ as the event that your initial choice is correct. Note that if your initial choice is correct then switching loses while if your initial guess is incorrect then switching wins. Then

$$\mathbb{P}(B) = \mathbb{P}(B|C)\mathbb{P}(C) + \mathbb{P}(B|C^c)\mathbb{P}(C^c) = 0 \times \frac{1}{3} + 1 \times \frac{2}{3} = \frac{2}{3}. \tag{9}$$

Thus switching is superior to sticking and doubles your chance of winning. There is a long history to this problem. [1]

**Example.** Monty Hall with $n$ doors and 1 car. In this case $\mathbb{P}(A) = 1/n$, $\mathbb{P}(C) = 1/n$, and thus

$$\mathbb{P}(B) = \mathbb{P}(B|C)\mathbb{P}(C) + \mathbb{P}(B|C^c)\mathbb{P}(C^c) = 0 \times \frac{1}{n} + \frac{1}{n-2} \times \frac{n-1}{n} = \frac{n-1}{(n-2)n}. \tag{10}$$

**Example.** You roll a 4-sided die – if you roll 1 or 2 you roll once more, otherwise you stop. What is the probability that the sum of your rolls is at least 4 ($B$)? Let $A_i$ be the event that the first roll is $i \in [4]$ and note $\mathbb{P}(A_i) = 1/4$ for $i \in [4]$. Note:

$$\mathbb{P}(B|A_1) = 1/2,\ \mathbb{P}(B|A_2) = 3/4,\ \mathbb{P}(B|A_3) = 0,\ \mathbb{P}(B|A_4) = 1, \tag{11}$$

and so

$$\mathbb{P}(B) = (1/2)(1/4) + (3/4)(1/4) + (0)(1/4) + (1)(1/4) = 9/16. \tag{12}$$

[1] Please see: http://en.wikipedia.org/wiki/Monty_Hall_problem. This quote from the Wikipedia article gives some context:

> Many readers refused to believe that switching is beneficial. After the Monty Hall problem appeared in Parade, approximately 10,000 readers, including nearly 1,000 with PhDs, wrote to the magazine claiming that vos Savant was wrong. (Tierney 1991) Even when given explanations, simulations, and formal mathematical proofs, many people still do not accept that switching is the best strategy.

This controversy is discussed in this 1991 NYT article: http://www.nytimes.com/1991/07/21/us/behind-monty-hall-s-doors-puzzle-debate-and-answer.html.

### Inference and Bayes' rule

In the following context we think of an event $B$ as the effect (which is observable, i.e., we will know whether the random outcome is in $B$ or $B^c$), and we are interested in ascertaining the cause (which is not observable), represented by the partition $A_1, \ldots, A_n$ of $\Omega$. We are interested in the probabilities of each possible cause given the effect $\mathbb{P}(A_i|B)$, but our model tells us only the probability of the effect given the cause $\mathbb{P}(B|A_i)$. Bayes' rule allows us to obtain each $\mathbb{P}(A_i|B)$ from all the $\mathbb{P}(B|A_i)$ and $\mathbb{P}(A_i)$ values.

**Bayes' rule.** Let $A_1, \ldots, A_n$ be a partition of $\Omega$ and let $B$ be an event. Then

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_{j=1}^n \mathbb{P}(B|A_j)\mathbb{P}(A_j)}, \ i \in [n]. \tag{13}$$

Note the first equality is simply the definition of conditional probability, and the second equality above is simply the TPT for $\mathbb{P}(B)$. In this light the rule is rather trivial, but its power in solving problems is profound. See Fig. 1.14 for an illustration.

**Example: radar detection.** If an aircraft is present ($A$) it may generate an alarm ($B$). The probabilities are specified conditionally (this is the model):

$$P(B|A) = 0.99, \ P(B|A^c) = 0.10, \ P(A) = 0.05 \tag{14}$$

Find $\mathbb{P}(A|B)$, the probability an aircraft is present given an alarm:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|A^c)\mathbb{P}(A^c)} = \frac{0.05 \cdot 0.99}{0.05 \cdot 0.99 + 0.95 \cdot 0.1} = 0.3426. \tag{15}$$

**Example: chess tournament.** Your win probability in a chess tournament is 0.3 against half the players (type 1), 0.4 against a quarter of the players (type 2), and 0.5 against the remaining quarter (type 3). You play against a random opponent and win – what is the probability you had an opponent of type 1 ($\mathbb{P}(A_1|B)$)?

$$\mathbb{P}(A_1|B) = \frac{\mathbb{P}(B|A_1)\mathbb{P}(A_1)}{\mathbb{P}(B|A_1)\mathbb{P}(A_1) + \mathbb{P}(B|A_2)\mathbb{P}(A_2) + \mathbb{P}(B|A_3)\mathbb{P}(A_3)} = \frac{0.5 \cdot 0.3}{0.5 \cdot 0.3 + 0.25 \cdot 0.4 + 0.25 \cdot 0.5} = 0.4. \tag{16}$$

**Example: false positive puzzle.** A test for a rare disease is correct 95% of the time: if a person has the disease the test is positive 95% of the time, if the person does not have the disease the test is negative 95% of the time. A random person drawn from the population has the disease with probability 0.001. Given the test is positive, what is the probability of having the disease? Let $A$ be the event the person has the disease (the cause) and let $B$ be the event the test is positive (the effect). We want to find $\mathbb{P}(A|B)$:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|A^c)\mathbb{P}(A^c)} = \frac{0.001 \cdot 0.95}{0.001 \cdot 0.95 + 0.999 \cdot 0.05} = 0.0187. \tag{17}$$

Note the asymmetry: a positive test on a 95% correct test for a disease affecting one in a thousand in a population is in fact correct only 2% of the time.

## §1.5 Independence

Events $A$ and $B$ are independent when partial information $B$ has no impact on the probability of $A$, i.e.:

$$\mathbb{P}(A|B) = \mathbb{P}(A). \tag{18}$$

Using the definition of conditional probability we see independence is equivalent to:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) \tag{19}$$

This last equation explains the phrase "independence means multiply" often used to explain independence. Note disjoint events are not independent. If $A \cap B = \emptyset$ then knowledge of $B$ being true means we instantly know $A$ to be false, so $A, B$ are clearly dependent events.

**Example.** Consider two successive rolls of a four sided die in which all 16 outcomes are equally likely.

1. Are events that the first roll is $i$ ($A_i$) and the second roll is $j$ ($B_j$) independent? Yes:
$$\mathbb{P}(A_i \cap B_j) = 1/16,\ \mathbb{P}(A_i) = 1/4,\ \mathbb{P}(B_j) = 1/4. \tag{20}$$
2. Are events $A$ (first roll is a 1) and $B$ (sum of the two rolls is 5) independent? Yes:
$$\mathbb{P}(A \cap B) = \mathbb{P}((1,4)) = 1/16,\ \mathbb{P}(A) = 1/4,\ \mathbb{P}(B) = 1/4. \tag{21}$$
3. Are events $A$ (max is 2) and $B$ (min is 2) independent? No:
$$\mathbb{P}(A \cap B) = \mathbb{P}((2,2)) = 1/16,\ \mathbb{P}(A) = 3/16,\ \mathbb{P}(B) = 5/16. \tag{22}$$

## Conditional independence

Conditional independence is the natural generalization of independence. Given events $A, B, C$ we say $A$ is conditionally independent of $B$ given $C$ if:
$$\mathbb{P}(A \cap B|C) = \mathbb{P}(A|C)\mathbb{P}(B|C), \tag{23}$$
which illustrates "independence means multiply". More insightful, however, is to observe that in general:
$$\mathbb{P}(A \cap B|C) = \frac{\mathbb{P}(A \cap B \cap C)}{\mathbb{P}(C)} = \frac{\mathbb{P}(A|B \cap C)\mathbb{P}(B|C)\mathbb{P}(C)}{\mathbb{P}(C)} = \mathbb{P}(A|B \cap C)\mathbb{P}(B|C). \tag{24}$$
So, using the definition and the above expression, an alternate definition of conditional independence is:
$$\mathbb{P}(A|B \cap C) = \mathbb{P}(A|C). \tag{25}$$
This states: given knowledge that $C$ is true, the estimates of the probability of $A$ are unchanged by further knowing $B$.

It is important to note that conditional independence does not imply independence, nor does independence imply conditional independence, as illustrated in the following examples.

**Example.** Toss two distinguishable fair coins, with all four outcomes equally likely. Define events $H_1$ (first toss is heads), $H_2$ (second toss is heads), and $D$ (tosses are different). Then $H_1, H_2$ are clearly unconditionally independent. They are not conditionally dependent given $D$:
$$\mathbb{P}(H_1|D) = 1/2,\ \mathbb{P}(H_2|D) = 1/2,\ \mathbb{P}(H_1 \cap H_2|D) = 0. \tag{26}$$
**Example.** Consider a (biased) red coin with probability of heads 0.01 and a (biased) blue coin with probability of heads 0.99. Select a coin at random (each one chosen with probability one half) and make two independent tosses of that coin. Define events $B$ (blue coin selected), $H_i$ (toss $i$ is heads) for $i \in [2]$. Then $H_1$ and $H_2$ are conditionally independent given $B$
$$\mathbb{P}(H_1 \cap H_2|B) = \mathbb{P}(H_1|B)\mathbb{P}(H_2|B) = 0.99 \cdot 0.99. \tag{27}$$
But $H_1, H_2$ are not unconditionally independent:
$$\begin{aligned}
\mathbb{P}(H_1) &= \mathbb{P}(H_1|B)\mathbb{P}(B) + \mathbb{P}(H_1|B^c)\mathbb{P}(B^c) = (1/2)(0.99) + (1/2)(0.01) = 1/2 \\
\mathbb{P}(H_2) &= \mathbb{P}(H_2|B)\mathbb{P}(B) + \mathbb{P}(H_2|B^c)\mathbb{P}(B^c) = (1/2)(0.99) + (1/2)(0.01) = 1/2 \\
\mathbb{P}(H1 \cap H_2) &= \mathbb{P}(H_1 \cap H_2|B)\mathbb{P}(B) + \mathbb{P}(H_1 \cap H_2|B^c)\mathbb{P}(B^c) = (1/2)(0.99 \cdot 0.99) + (1/2)(0.01 \cdot 0.01) \approx 1/2
\end{aligned} \tag{28}$$

## Independence of a collection of events

Thus far we have discussed pairwise independence, i.e., independence of pairs of events. We now generalize this definition.

**Definition of independence of several events.** Events $A_1, \ldots, A_n$ are independent if
$$\mathbb{P}\left(\bigcap_{i \in \S} A_i\right) = \prod_{i \in S} \mathbb{P}(A_i),\ \forall S \subseteq [n]. \tag{29}$$

As an example, for $n = 3$ the above definition asserts events $A_1, A_2, A_3$ are independent if

$$\begin{aligned}
\mathbb{P}(A_1 \cap A_2) &= \mathbb{P}(A_1)\mathbb{P}(A_2) \\
\mathbb{P}(A_1 \cap A_3) &= \mathbb{P}(A_1)\mathbb{P}(A_3) \\
\mathbb{P}(A_2 \cap A_3) &= \mathbb{P}(A_2)\mathbb{P}(A_3) \\
\mathbb{P}(A_1 \cap A_2 \cap A_3) &= \mathbb{P}(A_1)\mathbb{P}(A_2)\mathbb{P}(A_3) \qquad (30)
\end{aligned}$$

The following examples illustrate two important lessons about independence of multiple events.

**Example: pairwise independence does not imply independence.** Consider two independent tosses of a fair coin. Consider events $H_1$ (first toss is heads), $H_2$ (second toss is heads), and $D$ (tosses are different). Then:

$$\begin{aligned}
\mathbb{P}(H_1 \cap H_2) &= \mathbb{P}(H_1)\mathbb{P}(H_2) \\
\mathbb{P}(H_1 \cap D) &= \mathbb{P}(H_1)\mathbb{P}(D) \\
\mathbb{P}(H_2 \cap D) &= \mathbb{P}(H_2)\mathbb{P}(D) \\
\mathbb{P}(H_1 \cap H_2 \cap D) = 0 &\neq \mathbb{P}(H_1)\mathbb{P}(H_2)\mathbb{P}(D) = 1/8 \qquad (31)
\end{aligned}$$

**Example.** Consider two independent rolls of a fair six-sided die. Define events $A$ (first roll is $1, 2, 3$), $B$ (first roll is $3, 4, 5$), and $C$ (sum of the two rolls is 9):

$$\begin{aligned}
\mathbb{P}(A \cap B) = 1/6 &\neq \mathbb{P}(A)\mathbb{P}(B) = 1/4 \\
\mathbb{P}(A \cap C) = 1/36 &\neq \mathbb{P}(A)\mathbb{P}(C) = 1/18 \\
\mathbb{P}(B \cap C) = 1/12 &\neq \mathbb{P}(B)\mathbb{P}(C) = 1/18 \\
\mathbb{P}(A \cap B \cap C) = 1/36 = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) = 1/36 & \qquad (32)
\end{aligned}$$

The point being: the equality $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$ does not imply pairwise independence.

# References

[1] *Introduction to Probability, 2nd Edition* by Dimitri P. Bertsekas and John N. Tsitsiklis, Athina Scientific Press, 2008.

[2] *The Monty Hall Problem* by J. Rosenhouse, Oxford University Press, 2009.