Notes on Probability Theory

There are 3 kinds of probabilities:

1. Marginal probability – the probability of an event.

   Example 1. Given a uniformly distributed random variable X with range $R_X = \{1, 2, 3, 4, 5, 6\}$, we can ask what is the probability $\Pr[X = 6]$. Considering the uniform distribution of X, we can say $\Pr[X = 6] = \frac{1}{6}$.

2. Joint probability – the probability of intersection of two or more events. I.e., $\Pr[A \cap B]$ is the probability of an event that events A and B happen at the same time.

3. Conditional probability – the probability that some event A happens given that some other event B has happened. Written as $\Pr[A|B]$ – the probability of event A given B.

The **multiplication rule** binds all the 3 kinds of probabilities together into one relation:

$$\Pr[A \cap B] = \Pr[A|B] \cdot \Pr[B] = \Pr[B|A] \cdot \Pr[A] \ .$$

If events A and B are **independent**, then $\Pr[A|B] = \Pr[A]$ and $\Pr[B|A] = \Pr[B]$, and the multiplication rule takes the form of

$$\Pr[A \cap B] = \Pr[A] \cdot \Pr[B] \ .$$

Example 2. Getting "A" in this course does not depend on the color of your hair, and we can say that $\Pr[\text{"A"}|\text{red hair}] = \Pr[\text{"A"}|\text{green hair}] = \Pr[\text{"A"}]$.

Example 3. Consider two uniformly distributed random variables X with range $R_X = \{1, 2, 3, 4, 5, 6\}$, and Y with range $R_Y = \{\text{heads}, \text{tails}\}$. Calculate $\Pr[X = 6, Y = \text{heads}]$.

Solution. The probability that $X = 6$ does not depend on whether $Y = \text{heads}$ or $Y = \text{tails}$. And so

$$\Pr[X = 6, Y = \text{heads}] = \Pr[X = 6|Y = \text{heads}] \cdot \Pr[Y = \text{heads}] = \Pr[X = 6] \cdot \Pr[Y = \text{heads}] = \frac{1}{6} \cdot \frac{1}{2} = \frac{1}{12} \ .$$

The probability of a union of two events is calculated as

$$\Pr[A \cup B] = \Pr[A] + \Pr[B] - \Pr[A \cap B] \ .$$

In class we've drawn a picture on the whiteboard demonstrating the intuition behind this formula using counting. If A and B are **mutually exslusive** (meaning that they can never happen at the same time), then $A \cap B = \emptyset$, and $\Pr[A \cap B] = 0$, and the formula takes its simplified form

$$\Pr[A \cup B] = \Pr[A] + \Pr[B] \ .$$

Example 4. Continuing the story of Example 3, we can ask what is $\Pr[X = 6 \text{ or } Y = \text{heads}]$. Since these events are not mutually exclusive (they can happen at the same time), then

$$\Pr[X = 6 \text{ or } Y = \text{heads}] = \Pr[X = 6] + \Pr[Y = \text{heads}] - \Pr[X = 6 \cap Y = \text{heads}] = \frac{1}{6} + \frac{1}{2} - \frac{1}{12} = \frac{7}{12} \ .$$

As an example of mutually exclusive events, consider an events $X = 5$ and $X = 6$. The probability of getting a 5 or a 6 is

$$\Pr[X = 5 \text{ or } X = 6] = \Pr[X = 5] + \Pr[X = 6] = \frac{1}{6} + \frac{1}{6} = \frac{1}{3} \ .$$

Example 5. Consider a class of 30 students, 17 of whom are foreigners, and the rest 13 are local students. The rest results show that 4 foreigners and 5 local students made an "A". What is the probability that a randomly selected student will be a local student, or the one who got an "A"?

Solution. Since these events are not mutually exclusive (a local student can get an "A"),

$$\Pr[\text{local or "A"}] = \Pr[\text{local}] + \Pr[\text{"A"}] - \Pr[\text{local and "A"}] = \frac{13}{30} + \frac{9}{30} - \frac{5}{30} = \frac{17}{30} \ .$$

Example 6. Imagine that we wish to learn if the salary rate influences the color of one's car. For this reason, we have conducted interviews with employees of an enterprise, and collected their responses. The aggregated view can be seen in Table 1. It can be seen that

Table 1: Salary rate vs car color

	red car	other color
low salary	28	252
high salary	7	63

$$\begin{aligned}
&\Pr[\text{high salary}|\text{red car}] = \frac{1}{5} = \Pr[\text{high salary}] \ , \\
&\Pr[\text{low salary}|\text{red car}] = \frac{4}{5} = \Pr[\text{low salary}] \ , \\
&\Pr[\text{high salary}|\text{not red car}] = \frac{1}{5} = \Pr[\text{high salary}] \ , \\
&\Pr[\text{low salary}|\text{not red car}] = \frac{4}{5} = \Pr[\text{low salary}] \ .
\end{aligned}$$

So we can conclude that the salary rate and the color of a car are independent and do not influence each other.

Example 7. If we study the relationship of the salary rate and the presence of B.Sc degree, we can see a different statistics. Among all the respondents, 30% have B.Sc degree, and 40% evaluated their salary rate as high. Among those who has a B.Sc, 80% have high salaries. 20% of people without B.Sc also get high salaries. We are interested to determine if the salary rate depends on the presence or absence of a B.Sc degree.

We can see that

$$\Pr[\text{high salary}|\text{B.Sc}] = \frac{4}{5} \neq \frac{2}{5} \neq \Pr[\text{high salary}] \ ,$$

and hence these parameters correlate.

The multiplication rule

$$\Pr[A \cap B] = \Pr[A|B] \cdot \Pr[B] = \Pr[B|A] \cdot \Pr[A]$$

provides us with the relationship between $\Pr[A|B]$ and $\Pr[B|A]$. We can see that

$$\underbrace{\Pr[A|B]}_{\text{posterior}} = \frac{\overbrace{\Pr[B|A]}^{\text{likelihood}} \cdot \overbrace{\Pr[A]}^{\text{prior}}}{\underbrace{\Pr[B]}_{\text{evidence}}} \ .$$

This is known as the **Bayes' Theorem**. It allows us to make guesses about observations based on our prior knowledge. Consider the following example.

Example 8. Assume you are standing in a line to a football match and see someone with long hair. You have no idea is it a man or a woman. The Bayes' theorem allows us to compute a distribution of the likelihood of this person being a man or a woman, considering that we observe a long hair.

Our prior knowledge about the world is the following. Since it is a line to a football match, we expect to meet men more likely then women. We believe that, on average, out of 100 people, there are 98 men and 2 women. 94 men have short hair, 4 men have long hair. Among women, we believe that the distribution is even, meaning that 1 woman has short hair, and 1 woman has long hair. We observe a long hair in front of us and ask is this person more likely a man or a woman?

Solution. Applying the Bayes' theorem, we get

$$\Pr[\text{man}|\text{long hair}] = \frac{\Pr[\text{long hair}|\text{man}] \cdot \Pr[\text{man}]}{\Pr[\text{long hair}]} \ .$$

Let us now calculate individual parts of this equation. Pr[man] is the probability that a randomly selected person is a man. This probebility is equal to

$$\Pr[\text{man}] = \frac{98}{100} \ .$$

Pr[long hair|man] is the probability that a randomly selected man will have long hair.

$$\Pr[\text{long hair}|\text{man}] = \frac{4}{98} \ .$$

$\Pr[longhair]$ is the probability that a randomly selected person will have a long hair. Since men and women both can have long hair, this event can be decomposed into a union of two joint probabilities

$$\Pr[\text{long hair}] = \Pr[\text{man}, \text{long hair} \cup \text{woman}, \text{long hair}] \ .$$

$\Pr[\text{man}, \text{long hair}]$ is the joint probability of an event that a randomly selected person will be a man with long hair. $Pr[\text{woman}, \text{long hair}]$ is the joint probability that a randomly selected person will be a woman with long hair.

$$\Pr[\text{man}, \text{long hair}] = \Pr[\text{man}] \cdot \Pr[\text{long hair}|\text{man}] = \frac{98}{100} \cdot \frac{4}{98} = \frac{4}{100} \ ,$$
$$\Pr[\text{woman}, \text{long hair}] = \Pr[\text{woman}] \cdot \Pr[\text{long hair}|\text{woman}] = \frac{2}{100} \cdot 12 = \frac{1}{100} \ ,$$

Since events "man with long hair" and "woman with long hair" are mutually exclusive (you can't be a man and a woman at the same time), then

$$\Pr[\text{long hair}] = \Pr[\text{man}, \text{long hair}] + \Pr[\text{woman}, \text{long hair}] = \frac{4}{100} + \frac{1}{100} = \frac{5}{100} \ .$$

Now we have all pieces of the puzzle ready to be plugged into the Bayes' formula.

$$\Pr[\text{man}|\text{long hair}] = \frac{\Pr[\text{long hair}|\text{man}] \cdot \Pr[\text{man}]}{\Pr[\text{long hair}]} = \frac{4 \cdot 98 \cdot 100}{98 \cdot 100 \cdot 5} = \frac{4}{5} = 0.8 \ .$$

The Bayes' theorem is extensively used in statistics (Bayesian inference) and machine learning (naive Bayes classifier).