National University of Singapore
School of Computing
CS3243

A Quick Guide to Basic Probability Theory

We will be using some basic concepts from probability theory that you may not have seen before. The objective of this writeup is to offer a semi-formal introduction to probability theory.

1 Basic Concepts

Probability theory is the study of *random events*[1]. Consider a room containing 20 students; there are 13 CS and 7 CEG students. Of the CS students, 7 stay in the student halls and 6 stay off campus, whereas 2 of the CEG students stay in the student halls and 5 stay off campus. We call a student from the room; an event would be 'the student stays on campus' or 'the student is a CS student', or 'Michelle, a CEG student who stays off-campus came out'.

Definition 1.1. A *random variable* X takes values from a given *sample space* $\mathcal{S}_X$. A *probability measure* assigns a non-negative value $\Pr[X = x] \triangleq p_X(x)$ to every $x \in \mathcal{S}_X$. If $\mathcal{S}_X$ is finite, then we require that

$$\sum_{x \in \mathcal{S}_X} p_X(x) = 1$$

The case of infinite sized $\mathcal{S}_X$ is not covered in this primer (or in the class).

When it is clear from context, we omit the subtext X from $\mathcal{S}_X$ and $p_X(x)$ and simply write $\mathcal{S}$ and $p(x)$. Events are subsets of $\mathcal{S}$. For example, in the case of the students described above, the event "a CEG student came out" is the set of all CEG students.

Definition 1.2. An *event* A is a subset of $\mathcal{S}$; the probability of an event is written as

$$\Pr[A] \triangleq \sum_{x \in A} p(x).$$

In particular, we have the following straightforward proposition.

Proposition 1.3. *Given two events $A, B \subseteq \mathcal{S}$,*

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

[1]One can go into an entire discussion about whether true randomness really exists. This is interesting, but for our purposes, we assume that it is possible to create, at the very least, a truly random coin flip.

Income (SGD)	15-24	25-34	35-44	45-54	55-64	65+
$< S\$2500$	0.062	0.051	0.037	0.019	0.015	0.039
$S\$2500 - S\$5000$	0.078	0.068	0.061	0.057	0.031	0.053
$> S\$5000$	0.015	0.051	0.094	0.119	0.111	0.039

Table 1: Age-Income distribution example.

Given two random variables X, Y, their *joint probability space* is given by $\mathcal{S}_X \times \mathcal{S}_Y$. A *joint probability distribution* over $S_X \times S_Y$ is a function assigning a non-negative value

$$p_{X,Y}(x, y) \stackrel{\triangle}{=} \Pr_{X,Y}[X = x \wedge Y = y]$$

to every pair $(x, y) \in \mathcal{S}_X \times \mathcal{S}_Y$. We require that $p_{X,Y}$ defines a probability over $\mathcal{S}_X \times \mathcal{S}_Y$:

$$\sum_{x \in \mathcal{S}_X} \sum_{y \in \mathcal{S}_Y} p_{X,Y}(x, y) = 1.$$

In particular note that

$$\Pr[X = x] = \sum_{y \in \mathcal{S}_Y} p_{X,Y}(x, y) = \sum_{y \in \mathcal{S}_Y} \Pr[X = x \wedge Y = y].$$

Example 1.4. Table 1 describes the (approximately accurate) distribution of Singaporeans by age/income. We let X be the random variable for age and Y be the random variable for income; then

$$\Pr[X = \text{15-24}, Y = \text{S\$2500-S\$5000}] = 0.078$$

1. What is the probability that $X =$ 35–44?
2. What is the probability that $Y = S\$5000$?

Definition 1.5 (Independence)**.** We say that two events A and B are independent if $\Pr[A \wedge B] = \Pr[A] \times \Pr[B]$.

For example, if we roll a die two times, then it is often assumed that the rolls are independent; thus, if X_1 and X_2 are the random variables corresponding to the first and second rolls respectively, then $\Pr[X_1 = 1 \wedge X_2 = 4] = \Pr[X_1 = 1] \times \Pr[X_2 = 4] = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$

2 Conditional Probabilities

Conditional probabilities ask "what is the probability that event A occurs, given that we know that event B occurs?". For example, suppose that we roll two dice, one can ask "what is the probability that the first die rolls a '2' given that the sum of the rolls is 8"? We denote this as $\Pr[x \in A \mid x \in B]$ or $\Pr[A \mid B]$ in shorthand. This is defined to be

$$\Pr[A \mid B] \triangleq \frac{\Pr[A \wedge B]}{\Pr[B]}$$

2.1 Bayes' Rule

The following is one of the most useful tools in probabilistic analysis.

Proposition 2.1 (Bayes' Rule)**.**

$$\Pr[A \mid B] = \Pr[B \mid A] \times \frac{\Pr[A]}{\Pr[B]}$$

Proof. We simply follow the definitions to obtain the result

$$\begin{aligned}\Pr[B \mid A] \times \frac{\Pr[A]}{\Pr[B]} &= \frac{\Pr[B \wedge A]}{\Pr[A]} \times \frac{\Pr[A]}{\Pr[B]} \\ &= \frac{\Pr[B \wedge A]}{\Pr[B]} = \Pr[A \mid B]\end{aligned}$$

□

Intuitively, it is much easier to think of Bayes' rule in a statistical way. If we think of $\Pr[X = x]$ to be the proportion of elements in the underlying sample space that have the property x, then $\Pr[A \mid B]$ translates to "how many elements have the property A, given that we already know that they have the property B?". This would be simply the proportion of elements having both properties A and B — i.e. $\Pr[A \wedge B]$ — out of the number of elements that have the property B — $\Pr[B]$.

Another intuition about conditional probability is that of *information*: the expression $\Pr[A \mid B]$ expresses how knowledge of the event B changes our belief about the likelihood that the event A will occur. For example, if A is the event "I will be late for class" and B is the event "the internal shuttle bus broke down", then one may think that $\Pr[A \mid B] > \Pr[A]$; on the other hand, if B is the event "the shuttle bus arrived right on time" then it is reasonable to assume that $\Pr[A \mid B] < \Pr[A]$.

Example 2.2. I roll two dice independently; formally, this probability space is the product of two independent probability spaces $\mathcal{S}_X, \mathcal{S}_Y$ — one for each die roll. Here $\Pr[X = i] = \Pr[Y = i] = \frac{1}{6}$ for all $i \in \{1, \ldots, 6\}$, and X and Y are independent. What is $\Pr[X = i \mid X + Y = 7]$?

Example 2.3. An urn has 7 black balls, 8 red balls and 9 blue balls. I pull out one ball uniformly at random; what's the probability that it is red? What is the probability that it is red, given that I tell you that it is not blue? Work out this example using the definitions of conditional probability.

Example 2.4. A lab is screening for the HIV virus. A person that carries the virus is screened positive in 95% of the cases; a person who does not carry the virus screens positive in 1% of the cases. Alice gets tested in the lab and receives a positive result. What is the probability that Alice is actually carrying the HIV virus? Note that the sample space in this case is a product of two variables

$$\{\text{carrier}, \text{not carrier}\} \times \{\text{tested positive}, \text{did not test positive}\}.$$

We can think of independent events in terms of conditional probability.

Proposition 2.5. *A and B are independent if and only if* $\Pr[A \mid B] = \Pr[A]$.

Verify why Proposition 2.5 is true! Intuitively, Proposition 2.5 simply says that knowing B adds no additional information about A.

2.2 Conditional Independence

Suppose that the Ministry for Environmental and Water Resources (MEWR) wants to test the water in Jurong Lake for pollution levels. They send out two teams who take two samples in different areas of the lake. We let X_1 and X_2 be the random variables corresponding to the pollution levels observed in the tests run by teams 1 and 2, respectively (say, they take values from 0 to 10, where 0 is no pollution and 10 is high pollution). X_1 and X_2 were measured independently; were they really independent? Not really; think of the 'additional information' interpretation of independence: knowing that $X_1 > 7$ tells us that it's very likely that X_2 will be high as well (assuming that the teams did their job properly). However, they 'feel' independent: they are independent random samples after all!

Both X_1 and X_2 depend strongly on an additional variable: the *actual pollution level in Jurong lake*, which we denote Y. In this case, *knowledge of Y makes X_1 and X_2 truly independent!*. More formally, we say that X_1 and X_2 are *conditionally independent given* Y if for all $x_1 \in \mathcal{S}_{X_1}, x_2 \in \mathcal{S}_{X_2}, y \in \mathcal{S}_Y$ we have

$$\Pr[X_1 = x_1 \wedge X_2 = x_2 \mid Y = y] = \Pr[X_1 = x_1 \mid Y = y] \times \Pr[X_2 = x_2 \mid Y = y].$$

As before, this is often written in shorthand as

$$\Pr[X_1 \wedge X_2 \mid Y] = \Pr[X_1 \mid Y] \times \Pr[X_2 \mid Y].$$

The simplest way in which conditional independence occurs is by having a *past* event affecting current outcomes. In the previous case, *past* pollution levels are affecting our current water samples.

Example 2.6. An urn has 7 black balls, 8 red balls and 9 blue balls; first, I pull out a ball, do not show you its color, and *do not return it to the urn*. Next, I randomly pull out another ball, return it to the urn, and randomly pull out a ball again. Formally, let X be the random variable corresponding to the outcome of the first draw without replacement (e.g. $X =$ 'red' means that the first ball I took out was red), and let Y_1, Y_2 be the results of the two subsequent draws. Are Y_1 and Y_2 independent? Are they conditionally independent?

3 Expectation and Variance

The expected value of a random variable X is given by

$$\mathbb{E}[X] \stackrel{\triangle}{=} \sum_{x \in \mathcal{S}_X} x \times \Pr[X = x]$$

The conditional expectation of X given a random variable Y is written as

$$\mathbb{E}[X \mid Y = y] = \sum_{x \in \mathcal{S}_X} x \times \Pr[X = x \mid Y = y] = \frac{1}{\Pr[Y = y]} \sum_{x \in X} x \times \Pr[X = x \wedge Y = y]$$

Proposition 3.1 (Linearity of Expectation)**.** *Given two random variables X and Y, $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$. This holds even if X and Y are dependent.*

Proof.

$$\begin{aligned}
\mathbb{E}[X + Y] &= \sum_{x \in \mathcal{S}_X} \sum_{y \in \mathcal{S}_Y} (x + y) \times \Pr[X = x \wedge Y = y] \\
&= \sum_{x \in \mathcal{S}_X} \sum_{y \in \mathcal{S}_Y} x \times \Pr[X = x \wedge Y = y] + \sum_{x \in \mathcal{S}_X} \sum_{y \in \mathcal{S}_Y} y \times \Pr[X = x \wedge Y = y] \\
&= \sum_{x \in \mathcal{S}_X} x \times \left(\sum_{y \in \mathcal{S}_Y} \Pr[X = x \wedge Y = y] \right) + \sum_{y \in \mathcal{S}_Y} y \times \left(\sum_{x \in \mathcal{S}_X} \Pr[X = x \wedge Y = y] \right) \\
&= \sum_{x \in \mathcal{S}_X} x \times \Pr[X = x] + \sum_{y \in \mathcal{S}_Y} y \times \Pr[Y = y] = \mathbb{E}[X] + \mathbb{E}[Y]
\end{aligned}$$

□

The *variance* of a random variable is its expected squared distance from its expectation.

$$\mathrm{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

4 Probabilistic Inequalities

We conclude this primer with a list of useful inequalities.

Definition 4.1 (Markov's Inequality)**.** Let X be a non-negative random variable; then for all $t \geq 0$,

$$\Pr[X \geq t] \leq \frac{\mathbb{E}[X]}{t}$$

Definition 4.2 (Chebychev's Inequality)**.** Let X be a random variable with $\mathrm{Var}[X] < \infty$, then

$$\Pr[|X - \mathbb{E}[X]| \geq t] \leq \frac{\mathrm{Var}[X]}{t^2}$$

Definition 4.3 (Multiplicative (simplified) Chernoff Bound)**.** Let Y be the sum of n independent random variables $X_1, \dots, X_n$ where X_i takes the value 1 with probability p_i and 0 otherwise. Then for all $\delta \in (0, 1)$:

$$\Pr[Y < (1-\delta)\mathbb{E}[Y]] \leq \mathrm{e}^{-\frac{\delta^2 \mathbb{E}[Y]}{2}}$$
$$\Pr[Y > (1+\delta)\mathbb{E}[Y]] \leq \mathrm{e}^{-\frac{\delta^2 \mathbb{E}[Y]}{4}}$$

Acknowledgments There are many good primers on probability theory online. I have referenced some of them myself when writing this tutorial. Of particular help was Prof. Raz Kupferman's course notes on probability theory (which I have taken myself as an undergraduate student in the Hebrew University!)[2].

[2] http://www.ma.huji.ac.il/~razk/Teaching/LectureNotes/LectureNotesProbability.pdf