ST2131 (Probability)

Jia Cheng

January 2021

1 Definitions and Formula

2 Observations

2.1 Ways of defining outcomes

When defining outcomes regarding sequences or selections, it may be wise to specify 2 important properties.

- Distinguishable/Indistinguishable
- Ordered/unordered selection

The most **basic outcome** is usually an ordered selection of distinguishable elements. For e.g. when we want to select r balls from n red balls and m blue balls, the most basic way to define an outcome is to consider the n + m balls to be labelled, and that our order of selection matters.

Here, we define "**basic outcome**" as above.
Our discussion now assumes that each basic outcome has equal probability. In a finite sample space, the consequence of this is that $P(E) = \frac{|E|}{|S|}$.
However, sometimes we want to consider events of equal probabilities (or equal number of outcomes). I like to think of these as equivalence classes with equal cardinalities.
It is then convenient to define these as outcomes instead.
For e.g. we define an outcome to be an unordered selection of r distinct elements, since each unordered selection can be considered as an equivalence class of $rPr = r!$ ordered selections of r distinct elements.

From now on, I will use the notion of an "equivalence class" to refer to an event or a collection of basic outcomes which may then be used to define new outcomes.

2.1.1 Ensuring outcomes are of equal probability

The above process may lead to mistakes if the equivalence classes are not actually of equal size.

1. Suppose we want to distribute 20 (all distinct) items between 2 people, such that each person gets 10 items. 10 items are of type A, 10 items are of type B.
   Here, there are $\binom{20}{10,10}$ basic outcomes. Suppose I now define the events $E_{(a_1,a_2,\ldots,a_{10})}$ as follows, $E_{(a_1,a_2,\ldots,a_{10})}$ is the event where person 1 gets i-th item of type A if $a_i = 1$, person 2 gets i-th item of type A if $a_i = 0$.
   We can easily check that $|E_{(1,1,\ldots,1)}|$ is much smaller than $|E_{(1,1,1,1,1,0,0,0,0,0)}|$. Hence, if we define such events as outcomes, then the resulting outcomes would not be of equal probability.

This is especially relevant when it comes to conditional probability, as we will discuss below.

2.2 Conditional Probability

Consider example 2e in the textbook. Suppose that an urn contains 8 red balls and 4 white balls. We draw 2 balls from the urn without replacement.
(a) If we assume that at each draw, each ball in the urn is equally likely to be chosen, what is the probability that both balls drawn are red?

The traditional way to do this would to define outcomes as the unordered selection of 2 balls. (this is valid since each unordered selection corresponds to 2! ordered selections.) Hence, $P(R_1R_2) = \frac{|R_1R_2|}{|S|} = \frac{\binom{8}{2}}{\binom{12}{2}} = \frac{14}{33}$

Another way to do this is via conditional probability. We claim that $P(R_2|R_1) = \frac{7}{11}$ and hence $P(R_1R_2) = P(R_1)P(R_2|R_1) = \frac{8}{12}\frac{7}{11} = \frac{14}{33}$.

But why this claim that $P(R_2|R_1) = \frac{7}{11}$ justified? Remember that conditional probability is merely a definition, that $P(A|B) = \frac{P(AB)}{P(B)}$, and definitions do not contribute to the underlying theory.

This is because we are implicitly defining new outcomes based on events. Precisely stated, we now relabel the remaining 11 balls from 1 to 11. Let E_i be the event where the first ball chosen is a red ball and the 2nd ball chosen is ball i. So how many basic outcomes lie in each event E_i? Since the first ball chosen could have been any red ball, $|E_i| = 8 \forall 1 \leq i \leq 11$. Since all the events are equally large, we can consider these as outcomes.
Out of these 11 outcomes, only 7 result in the 2nd ball chosen being red. Hence, $\frac{P(R_1R_2)}{P(R_1)} = \frac{|R_1R_2|}{|R_1|} = \frac{7}{11}$. And by definition of conditional probability, $P(R_2|R_1) = \frac{7}{11}$.

In conclusion, when we make claims about conditional probability in such problems, we are actually implicitly redefining the underlying outcomes and then making use of the traditional method.

2.2.1 Taking advantage of symmetries

Suppose event B is the disjoint union of sets, where $B = \cup_i B_i$ and that $P(B_i)$ is constant for all i. Suppose also that for event A, we have $P(A \cap B_i)$ constant for all i.

Then,

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{AB_i}{B_i} = P(A|B_i)$$

This can be the case when we see statements like these:

- An ordinary deck of 52 playing cards is randomly divided into 4 piles of 13 cardseach. Compute the probability that each pile has exactly 1 ace.(Example 2g of textbook)
  Let B be the event where the ace of spades and the ace of hearts are in different piles . Note that here, it is not specified exactly which pile the 2 aces are in. However, using the symmetrical nature of things, we can specify a special case $B_{1,2}$ and let ace of spades reside in pile 1, ace of hearts reside in pile 2.

2.3 Defining the sample space/outcomes

This uses example 3o from chapter 3 as an example. A crime has been committed by a solitary individual, who left some DNA at the scene of the crime. Forensic scientists who studied the recovered DNA noted that only five strands could be identified and that each innocent person, independently, would have a probability of of having his or her DNA match on all five strands. The district attorney supposes that the perpetrator of the crime could be any of the 1 million residents of the town. Ten thousand of these residents have been released from prison within the past 10 years; consequently, a sample of their DNA is on file. Before any checking of the DNA file, the district attorney thinks that each of the 10,000 ex-criminals has probability of being guilty of the new crime, whereas each of the remaining 990,000 residents has probability where (That is, the district attorney supposes that each recently released convict is times as likely to be the crime's perpetrator as is each town member who is not a recently released convict.) When the DNA that is analyzed is compared against the database of the 10,000 ex-convicts, it turns out that A. J. Jones is the only one whose DNA matches the profile. Assuming that the district attorney's estimate of the relationship between and is accurate, what is the probability that A. J. is guilty?

Consider the 2 sample spaces that can be defined. S1: For an ex-con, the sample space has 2 outcomes guilty, not guilty. Based on the qn, $P(\{guilty\}) = \alpha$ and $P(\{notguilty\} = 1 - \alpha$.

S2: For the population of the whole town, we can label each of the one million residents from 1 to 1 million. The sample space then has 1 million outcomes $\{p1, p2, \dots, p_{1000000}\}$, where outcome p_i indicates the i-th person is guilty.

Now for a qn: What is the probability that none of the ex-convicts are guilty? Is it $(1-\alpha)^{10000}$ or $1 - 10000\alpha$?
Ans: $1 - 10000\alpha$. Since the event E that the guilty is a member of the ex-convicts is the subset of S2 containing all the ex-convicts. Hence $P(E) = 10000\alpha$.
It would be wrong to multiply together $(1-\alpha)^{10000}$ since the event that person i is not guilty is not independent from the event that person j is not guilty.

2.4 Principle of Inclusion Exclusion

The upper and lower bounds of $P(\cup_{i=1}^n E_i)$ are an excellent exercise in manipulating summation indices.

Notation: Given 2 sets E, F, define $EF = E \cap F$.

$$P(\cup_{i=1}^n E_i) = \sum_{1 \le j \le n} (-1)^{j+1} \sum_{1 \le i_1 < \cdots < i_j \le n} P(E_{i_1} \dots E_{i_j})$$

Claim: For odd k,

$$P(\cup_{i=1}^n E_i) \le \sum_{1 \le j \le k} (-1)^{j+1} \sum_{1 \le i_1 < \cdots < i_j \le n} P(E_{i_1} \dots E_{i_j})$$

For even k,

$$P(\cup_{i=1}^n E_i) \ge \sum_{1 \le j \le k} (-1)^{j+1} \sum_{1 \le i_1 < \cdots < i_j \le n} P(E_{i_1} \dots E_{i_j})$$

Here, we shall go from the odd k to even $k+1$. We first note that

$$P(\cup_{i=1}^n E_i) = \sum_{1 \le i \le n} P(E_i) - \sum_{1 \le i \le n} P(\cup_{1 \le j < i} E_j E_i)$$

Fix some $1 \le i \le n$.Applying the inequality for odd k, we have

$$\begin{aligned}
P(\cup_{1\le j<i} E_j E_i) &\le \sum_{1\le j\le k} (-1)^{j+1} \sum_{1\le i_1<\cdots<i_j\le n} P(E_{i_1}E_iE_{i_2}E_i \dots E_{i_j}E_i) \\
&= \sum_{1\le j\le k} (-1)^{j+1} \sum_{1\le i_1<\cdots<i_j\le n} P(E_{i_1}E_{i_2} \dots E_{i_j}E_i)
\end{aligned}$$

Hence,

$$\begin{aligned}
P(\cup_{i=1}^n E_i) &\ge \sum_{1\le i\le n} P(E_i) - \sum_{1\le i\le n}\sum_{1\le j\le k} (-1)^{j+1} \sum_{1\le i_1<\cdots<i_j\le n} P(E_{i_1}E_{i_2} \dots E_{i_j}E_i) \\
&= \sum_{1\le i\le n} P(E_i) - \sum_{1\le j\le k} (-1)^{j+1} \sum_{1\le i_1<\cdots<i_j<i_{j+1}\le n} P(E_{i_1}E_{i_2} \dots E_{i_j}E_{i_{j+1}}) \\
&= \sum_{1\le j\le k+1} (-1)^{j+1} \sum_{1\le i_1<\cdots<i_{j+1}\le n} P(E_{i_1}E_{i_2} \dots E_{i_{j+1}})
\end{aligned}$$

Going from even k to odd $k+1$ is similar.