MATH3200: APPLIED LINEAR ALGEBRA SELF-STUDY AND PRACTICE MODULE 11: WHAT ARE PROBABILITIES, ANYWAY? A VERY BRIEF INTRODUCTION TO PROBABILITIES AND PROBABILITY DISTRIBUTIONS

WINFRIED JUST, OHIO UNIVERSITY

We will use the terminology and notation of Conversation 7.

We can think of an *event* or *outcome* as an observation that we may make when we do an *experiment.* For example,

- If we flip a coin, the relevant events might be H (that it comes up heads) and T (that it comes up tails).
- If we roll a die, the events might be $e_1, e_2, e_3, e_4, e_5, e_6$, where e_1 means the die comes up "1", e_2 means the die comes up "2", and so on.
- If we observe the weather on a given day at noon, the events might be S (sunshine), R (rain), O (dry, but overcast).

In all the above examples, the outcome of the experiment will not be entirely predictable. When we want to study the degree of unpredictability mathematically, we need to assign *probabilities* to events. *Probability theory* is the branch of mathematics that studies uncertainty. We need here only the following basic facts about probabilities:

- Probabilities are numbers p such that $0 \leq p \leq 1$.
- A probability of $p = 0$ for an event or outcome signifies that the event will *never* occur.
- A probability of $p = 1$ for an event signifies that the event *will occur with certainty.*
- A probability of $p = 0.5$ for an event signifies that there is a fifty-fifty chance that the event will occur. For example, if we flip a fair coin, then the event H that it will come up heads will be $p = 0.5$.
- Similarly, if we roll a fair die, then the event e_3 that it will come up "3" will be $p = 1/6$.

For a given *sample space* that consists of *elementary outcomes* that are listed in a given order, the *probability distribution* will be vector of probabilities of the elementary outcomes.

In the first example above, the sample space would be $\{H, T\}$ and the if the coin is fair, the probability distribution will be the vector $[0.5, 0.5]$.

In the second example above, the sample space would be $\{e_1, e_2, e_3, e_4, e_5, e_6\}$ and the if the die is fair, the probability distribution will be the vector $\left[\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}\right]$.

In the third example above, the sample space would be $\{S, R, O\}$. However, it is not immediately clear what the probability distribution would be. This will depend on the particular location. We can think of the probability of an event as the frequency with which the event would be observed if we repeat the experiment *under similar circumstances* many times. For example, suppose we go out on 56 consecutive days and record the following observations about the weather:

$SSROOSRRORSOSSSSOSRRRRSROOOOROOSSROOORSORRSSRRRRSOOORROS$

We observe the event S 17 times, the event R 20 times, and the event O 19 times. Thus we could *estimate* the probability of S as $\frac{17}{56}$, the probability of R as $\frac{20}{56}$ and the probability of O as $\frac{19}{56}$. Then the probability distribution will be the vector $\left[\frac{17}{56}, \frac{20}{56}, \frac{19}{56}\right]$.

Statistics is the branch of applied mathematics that studies, among other things, how to skillfully derive such estimates. You can see that even in this example matters are not entirely straightforward, because it could happen that at noon it rains while the sun shines. The observer would then see a beautiful rainbow, but would need to make a decision on whether to record the observation as S or R. Generally speaking, events that could occur simultaneously tend to complicate matters. Similarly, it is not immediately clear what our observer should record when it snows. For this reason mathematicians require that the elementary outcomes in a sample space must be *mutually exclusive,* which means that no two of them could occur simultaneously, and *exhaustive,* so that every possible observation could be classified as one of them.

Once we have such a sample space, we can lump together some elementary outcomes into larger events. To see how this works, let us assume for simplicity that in the above example event R signifies any precipitation and S is understood as implying "dry", similarly to O. Then the set $\{S, R, O\}$ would be our sample space. We could now optimistically define a new event "sunny day" that would comprise the observations S and O, call it SD, and then let RD be the event "rainy day" that occurs whenever R is observed. Of course, R and RD are the same events, but let's use different names here to indicate the level at which we have lumped together some elementary outcomes. Note that the events SD and RD are still mutually exclusive and that exactly one of them must occur.

Question 11.1: Find an estimate of the probability of the event SD based on the above sequence of observations.

Question 11.2: What is the probability distribution that you would estimate from the above sequence of observations for the sample space $\{SD, RD\}$?