Tutorial 3: Probability Theory, Random Variables and Distributions (Revision 1)

Rob Nicholls – nicholls@mrc-lmb.cam.ac.uk

MRC LMB Statistics Course 2014

A very basic understanding of the R software environment is assumed.
See Tutorial 1 for an introduction to R.

This tutorial begins by stating basic facts required in order to complete the following series of tasks. Some tasks are paper-based, whilst some utilise R.

A set is a collection of objects. In probability theory, sets are used to represent events. If A and B are sets, then:

- A' – the complement of A – all outcomes except those in A.
- $A \cup B$ – the union of A and B – all outcomes in A or B.
- $A \cap B$ – the intersection of A and B – all outcomes common to both A and B.
- $A \backslash B$ – A not B – all outcomes in A that are not in B.
- $A \subseteq B$ – A is a subset of B – all outcomes in A are also in B.

All probabilities satisfy the identity: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

If A and B are *mutually exclusive* (disjoint) then $P(A \cap B) = 0$.

If A and B are *independent* then $P(A \cap B) = P(A)P(B)$.

Note that:

$$\begin{aligned}
P(A \backslash B) &= P(A \cap B') \\
&= P(A \cap (\Omega \backslash B)) \\
&= P((A \cap \Omega) \backslash (A \cap B)) \\
&= P(A \backslash (A \cap B)) \\
&= P(A) - P(A \cap B) \qquad\qquad \text{since } A \cap B \subseteq A
\end{aligned}$$

Task 1:

Consider a standard pack of 52 playing cards. Let A be the set of hearts, and B be the set of even cards (counting 2, 4, 6, 8, 10 and Queen as even cards).

1. What is $P(A)$? (calculate manually by thinking about it)
2. What is $P(B)$? (calculate manually by thinking about it)
3. What is $P(A \cap B)$? (calculate manually by thinking about it)
4. Are A and B *mutually exclusive*?
5. Are A and B *independent*?
6. Use your answers to the above questions to calculate $P(A \cup B)$.
7. Use your answers to the above questions to calculate $P(A \cap B')$. Which cards does the set $A \cap B'$ correspond to?

Task 2:

Suppose A and B are events, where $P(A \cup B) = 0.8$, $P(A) = 0.4$ and $P(B) = x$.

1. What is the value of x if A and B are *mutually exclusive*?
2. What is the value of x if $P(A \backslash B) = 0.2$?
3. What is the value of x if $P((A \cap B')') = 0.8$?
4. What is the value of x if $P(A' \cup B) = 0.8$?
5. What is the value of x if A and B are *independent*?
6. If A and B are *independent*, what is the probability of exactly one of A or B occurring? (hint – consider the set $(A \cap B') \cup (B \cap A')$)

Task 3:

Suppose A and B are events, where $P(B) = x$, $P(A \cup B) = 0.7$ and event A is three times as likely as B.

Calculate x, under the assumption that A and B are independent. Note that you will need to solve a quadratic equation of the form: $ax^2 + bx + c = 0$ using the formula: $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$. You can use R as a calculator for this calculation.

How many solutions are there to this problem, i.e. how many values can x adopt?

Task 4:

Note the following:

- If event A is partitioned by a series of n subsets B_i then $P(A) = \sum_i P(A \cap B_i)$.
- Conditional Probability is denoted $P(A|B)$ – this is the probability that event A occurs given that event B has occurred. The conditional probability of A given B may be calculated: $P(A|B) = \frac{P(A \cap B)}{P(B)}$.
- According to Bayes' Theorem $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$.

Suppose there are two bags, B_1 and B_2, each of which contain a number of coloured balls. Bag B_1 has 2 red and 3 yellow balls, bag B_2 has 4 red and 2 yellow balls. Someone removes a ball from one the bags. They select bag B_1 with probability 1/3, and bag B_2 with probability 2/3. This information can be summarised:
$P(red|B_1) = 2/5$
$P(yellow|B_1) = 3/5$
$P(red|B_2) = 2/3$
$P(yellow|B_2) = 1/3$
$P(B_1) = 1/3$
$P(B_2) = 2/3$

1. Using the facts provided regarding partitioned sets and conditional probabilities to calculate the probability that the selected ball is red.
2. Use Bayes' Theorem to evaluate the probability that Bag B_1 was selected, given knowledge that a red ball was selected.
3. Calculate the probability that Bag B_2 was selected, given knowledge that a red ball was selected.

Task 5:

Suppose a fair coin is tossed three times.

1. Write out the full set Ω, i.e. the sample space, using H to denote heads, and T to denote tails.
2. Let A be the event that the outcome is exactly two heads. Write out set A.
3. What is $P(A)$?
4. Suppose the coin is tossed 20 times. Use R's `choose` function to calculate the number of ways in which 10 heads could result from 20 trials.
5. Calculate the probability of achieving exactly 10 heads from 20 trials, using the mass (density) function corresponding to an appropriate probability distribution (type `?distributions` in R to get a list of distributions; hint – see the online lecture slides to help identify the appropriate distribution).

Task 6:

In a wine cellar, on average 20% of the bottles are not good.

1. Out of 10 bottles, what is the probability that at least 8 bottles are still good? (hint – look at the R help for `dbinom`, or alternatively `pbinom`)

2. Out of 20 bottles, what is the probability that at least 16 bottles are still good?

Task 7:

The local Health Department found an unexpected boom in male births, relative to female births, during May. There were 60 boys and 35 girls born during the month. Assume that the probability of a male birth (rather than a female birth) is $p = 0.514$. Is the above the above observation really extraordinary? (hint – look at the R help for the cumulative distribution function of the Binomial distribution: `pbinom`)

Task 8:

In Lecture 3, probability mass functions for Binomially-distributed random variables were shown (on pages 62–64 of the presentation).

1. Recreate the plot of the probability mass function for a random variable:

   $$X \sim \text{Bin}(10, 0.5)$$

   utilising the R functions:

   (a) `dbinom`

   (b) `plot`

   (c) `lines`

   Do this using a 'for loop' to iterate over the commands to draw the vertical lines corresponding to the mass function heights. Note – 'for loop' syntax:

   $$\texttt{for(i in 1:n)\{***INSERT\_COMMAND(S)***\}} \tag{1}$$

2. Create a custom R function that takes two parameters n and p, which automatically plots the probability mass function of a Binomially-distributed random variable with parameters n and p, using the syntax:

   $$\texttt{plot\_binom=function(n,p)\{***INSERT\_COMMAND(S)***\}} \tag{2}$$