MATH255 - Autumn 2023 Computer Lab - Week 9

Note. Question 3(a)(i), (ii) and (iii) (not 3(b), just 3(a)) is your Lab Preparation exercise for this week. Your answers must be submitted as a pdf document before the start of your lab.

Key Results from Lectures

- $P(A^c) = 1 - P(A)$ — Probability that A does **not** occur
- $P(A \cap B) = 0$ if $A \cap B = \emptyset$ (empty set) — **Disjoint** or **mutually exclusive** events
- $P(A \cup B) = P(A) + P(B)$ if $A \cap B = \emptyset$ — Probability of union (**or**) for **disjoint** events
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ — Probability of union (**or**) for **general** events
- $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$ — **Conditional** probabiity of A **given** B
- $P(A \cap B) = P(B)P(A|B) = P(A)P(B|A)$ — Prob. of intersection (**and**) for **general** events
- $P(A \cap B) = P(A)P(B)$ — Probability of intersection (**and**) for **independent** events

1. The `stackloss` dataset is available in R without having to import it. You can just enter `stackloss` to see the raw data, or `help(stackloss)` to obtain background information. First compute the correlations between each pair of variables.

   ```
   cor(stackloss)
   ```

   A correlation coefficient close to zero means little or no correlation between those two variables. A coefficient close to 1 (-1) indicates a strong positive (negative) relation between the variables. The stackloss correlations reveal a weak positive association between `stack.loss` and `Acid.Conc.`, and quite a strong positive association between `stack.loss` and `Air.Flow`. Check this out by looking at scatterplots with superimposed least squares lines. Are there any outliers?

   ```
   plot(stack.loss~Acid.Conc.,data=stackloss)
   abline(lm(stack.loss~Acid.Conc.,data=stackloss))
   ```

   Repeat the previous two commands for air flow and for water temperature.

2. To draw a Venn diagram in R, you first need to install the package.

   ```
   install.packages("VennDiagram")
   library("VennDiagram")
   ```

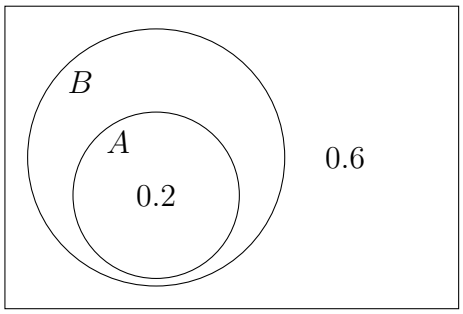
   To draw two (three) areas, the command is `draw.pairwise.venn` (`draw.triple.venn`). The basic inputs needed for a two-area plot are Area1, Area2 and cross_area. Optional additional inputs allow you to put titles, fill with colours, etc. For example, try (you may need to clear the graphing space with `grid.newpage()` first)

   ```
   draw.pairwise.venn(25,45,10,fill=c("Green","Red"))
   ```

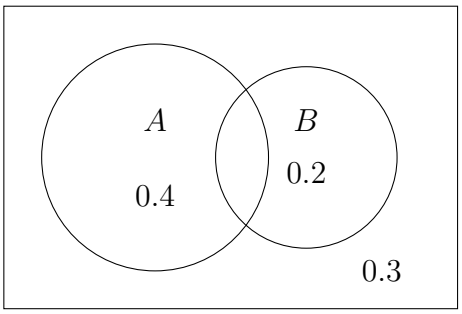
   The inputs for a three-area plot are Area1, Area2, Area3, Area12, Area23, Area13, Area123. Try making one.

3. For each of the following Venn diagrams:

(a)



(b)



(i) Evaluate $P(B)$, $P(A \cap B)$, $P(A \cup B)$, $P(A|B)$ and $P(B|A)$.

(ii) Are A and B^c disjoint (*i.e.* mutually exclusive)?

(iii) Are A and B^c independent?

(iv) Write down the corresponding two-way table of probabilities.

(v) Construct a tree diagram, branching first with respect to B, labelling all branches with the numerical values of the corresponding probabilities. Also calculate the probabilities of each path through the tree.

4. (a) An advertising agency notes that approximately one in 50 potential buyers of a product sees a given magazine advertisement and one in five sees the corresponding advertisement on television. One in 100 sees both. Construct a Venn diagram or a two-way table, and use it to find the probabilities that a potential customer does or does not see at least one form of advertisement.

(b) One in three of potential customers who have seen the advertisement in (a) purchase the product, and one in 10 of those who haven't seen the advertisement purchase the product. Construct a tree diagram and use it to find the probability that a randomly selected potential customer will purchase the product.