# STM4PSD – Workshop 3

1. Suppose that of all customers visiting a cafe, 90% purchase coffee, 40% purchase a sandwich, and 15% purchase both coffee and a sandwich. Given that a customer has purchased a sandwich, what is the probability that they also purchase a coffee?

2. Let $X \sim \text{Bin}(5, 0.2)$. Determine each of the following:

   (a) $P(X = 2)$ (b) $P(X \neq 0)$ (c) $P(X = 2 \mid X \neq 0)$

   Which of (a) and (c) is more likely? Can you give an intuitive reason why?

3. This question is based on a problem given on page 55 of [1]. Suppose you were investigating the effects of rainfall on peak flow rates in a city sewer system. The random variable $X$ has $\Omega_X = \{1, 2, 3\}$, and denotes the precipitation level for a given day, with 1, 2 and 3 representing low, medium and high levels, respectively. Similarly, the random variable $Y$ has $\Omega_Y = \{1, 2\}$, and denotes the peak flow rates on that same day, with 1 indicating that critical peak flow has been reached and 2 indicating that it has not. The joint probability mass function is tabulated below.

| | $X = 1$ | $X = 2$ | $X = 3$ |
|---|---|---|---|
| $Y = 1$ | 0.0 | 0.06 | 0.12 |
| $Y = 2$ | 0.5 | 0.24 | 0.08 |

   Using this joint probability mass function, answer the following.

   (a) What is $P(X = 1 \cap Y = 2)$? In words, what does this probability represent?

   (b) Determine $P(Y = 1)$ and $P(Y = 2)$.

   (c) Using your answer to (b), write down the probability mass function for $Y$ in table form.

   (d) Similarly, write down the probability mass function for the random variable $X$.

   (e) Given that a high level of precipitation has occurred, what is the probability that peak flow will be achieved? That is, calculate $P(Y = 1 \mid X = 3)$.

   (f) Given that peak flow was not achieved, what is the probability that there was medium level of precipitation?

4. A machine runs continuously in a factory. The machine is managed by computer software. Two kinds of errors can occur during the running of the machine: a mechanical error, or a software error. Either of these two errors may lead to a critical failure of the machine. On any given day, the probability of a mechanical error is $0.002$. and the probability of a software error is $0.007$. When a mechanical error occurs, the probability of a critical failure is $0.975$. When a software error occurs, the probability of a critical failure is $0.40$.

   Let $M$ denote the event that a mechanical error occurs, let $S$ denote the event that a software error occurs, and let $F$ denote the event that a critical failure occurs.

   (a) Use symbols to represent the four probabilities given in the statement above.

   (b) Use the Law of Total Probability to determine the probability that the machine has a critical failure.

   (c) Determine the probability that the machine has a critical failure and a software error.

   **Hint**: rearrange the formula from the definition of conditional probability.

   (d) The machine had a critical failure. What is the probability that it was due to a software error?

5. Recall Example 7.2.1 from the reading materials, regarding quality control of GPUs. We have that $F$ denotes that a GPU fails the quality control test, and that $D$ denotes the event that a GPU is truly defective. We were given that that $P(D) = 0.015$, $P(F \mid D) = 0.97$ and $P(F^c \mid D^c) = 0.95$.

   (a) In non-technical terms, explain what $P(F^c \mid D^c)$ represents.

   (b) Explain why $P(F \mid D^c) = 0.05$.

   (c) Use the Law of Total Probability to show that $P(F) = 0.0638$.

   (d) Using (b) and Bayes' Theorem, verify that $P(D \mid F)$ is, to three decimal places, $0.228$.

(e) In Example 7.2.2, the manager desires an improved test which lowers the false discovery rate to 20%. This goal will be achieved by increasing the true negative rate of the test (i.e., by increasing the chance that a functional GPU passes the test). In this question, we will determine the required true negative rate. Assume that $P(D)$ and $P(F \mid D)$ remain unchanged throughout this question.

i. Express the true negative rate and false discovery rate using appropriate mathematical symbols for this context. (Do not assign numeric values to them at this stage.)

ii. Because changing the TNR will affect the probability of failing the test, after implementing the change, $P(F)$ will have a different value than in part (c). Using the Law of Total Probability, express $P(F)$ in terms of $P(D)$, $P(F \mid D)$ and $P(F \mid D^c)$.

iii. Use Bayes' Theorem and part (ii) to express $P(D \mid F)$ without explicit reference to $P(F)$.

iv. Rearrange your answer to part (iii) to express $P(F \mid D^c)$ in terms of $P(D)$, $P(F \mid D)$ and $P(D \mid F)$.

v. Substitute known quantities and the desired false discovery rate into your answer to (iv) to find the required $P(F \mid D^c)$ (to 4 decimal places).

vi. Based on your answer to (v), write a sentence summarising the results for the manager.

## References

[1] T. T. Soong. *Fundamentals of Probability and Statistics for Engineers*. Wiley, Hoboken, NJ, 2004.