STM4PSD – Workshop 3 Solutions

1. Let C denote the event that a customer purchases coffee, and let S denote the event that a customer purchases a sandwich. We are given:

$$P(C) = 0.9,$$
$$P(S) = 0.4,$$
$$P(C \cap S) = 0.15.$$

Then,

$$P(C \mid S) = \frac{P(C \cap S)}{P(S)} = \frac{0.15}{0.4} = 0.375.$$

2. (a) $P(X = 2) = \binom{5}{2} \times 0.2^2 \times (1 - 0.2)^{5-2} = \frac{5!}{2!\,3!} \times 0.2^2 \times 0.8^3 = 10 \times 0.04 \times \times 0.512 = 0.2048$

   (b) $P(X \neq 0) = 1 - P(X = 0) = 1 - \binom{5}{0} \times 0.2^0 \times (1 - 0.2)^{5-0} = 1 - 1 \times 1 \times 0.8^5 = 1 - 0.8^5 = 0.67232$

   (c) $P(X = 2 \mid X \neq 0) = \frac{P(X=2 \cap X \neq 0)}{P(X \neq 0)} = \frac{P(X=2)}{P(X \neq 0)} = \frac{0.2048}{0.67232} \approx 0.305$
   Note that the event $X = 2 \cap X \neq 0$ simplifies to $X = 2$ because the intersection of the sets $\{2\}$ and $\{1, 2, 3, 4, \ldots\}$ is the set $\{2\}$.

   The event $X = 2 \mid X \neq 0$ is more likely. Intuitively speaking, this is because if we already know that $X \neq 0$, then this only increases the likelihood that $X = 2$.

3. (a) $P(X = 1 \cap Y = 2) = 0.5$. This represents the probability that there is low precipitation and critical flow has not been reached.

   (b) $P(Y = 1) = 0.0 + 0.06 + 0.12 = 0.18$
   $P(Y = 2) = 0.5 + 0.24 + 0.08 = 0.82$

   (c)

y	1	2
$P(Y = y)$	0.18	0.82

   (d) $P(X = 1) = 0.0 + 0.5 = 0.5$
   $P(X = 2) = 0.06 + 0.24 = 0.3$
   $P(X = 3) = 0.12 + 0.08 = 0.2$
   So the PMF for X is

x	1	2	3
$P(X = x)$	0.5	0.3	0.2

   (e) $P(Y = 1 \mid X = 3) = \frac{P(Y=1 \cap X=3)}{P(X=3)} = \frac{0.12}{0.2} = 0.6$

   (f) $P(X = 2 \mid Y = 2) = \frac{P(X=2 \cap Y=2)}{P(Y=2)} = \frac{0.24}{0.82} \approx 0.293$

4. (a)
$$P(M) = 0.002$$
$$P(S) = 0.007$$
$$P(F \mid M) = 0.975$$
$$P(F \mid S) = 0.4$$

   (b) $P(F) = P(F \mid M) \times P(M) + P(F \mid S) \times P(S) = 0.975 \times 0.002 + 0.4 \times 0.007 = 0.00475$

   (c) Here, we want $P(F \cap S)$ (as opposed to $P(F \mid S)$).
   From the definition of conditional probability, we have

$$P(F \mid S) = \frac{P(F \cap S)}{P(S)} \implies P(F \cap S) = P(F \mid S) \times P(S)$$

   Hence $P(F \cap S) = 0.4 \times 0.007 = 0.0028$

(d) This is asking for $P(S \mid F)$. We can use Bayes' theorem here:

$$P(S \mid F) = \frac{P(F \mid S) \times P(S)}{P(F)} = \frac{0.4 \times 0.007}{0.00475} \approx 0.589$$

One way this can be interpreted is: almost 60% of the machine's failures are due to a software error.

5. (a) $P(F^c \mid D^c)$ represents the probability that a non-defective unit passes the quality control test.

(b) $P(F \mid D^c) = 1 - P(F^c \mid D^c) = 1 - 0.95 = 0.05$.

(c) $$\begin{aligned} P(F) &= P(F \mid D)P(D) + P(F \mid D^c)P(D^c) \\ &= 0.97 \times 0.015 + 0.05 \times (1 - 0.015) \\ &= 0.0638 \end{aligned}$$

(d) $P(D \mid F) = \dfrac{P(F \mid D)P(D)}{P(F)} = \dfrac{0.97 \times 0.015}{0.0638} = 0.228$

(e) i. True negative rate: $P(F^c \mid D^c)$
False discovery rate: $P(D^c \mid F)$

ii. $$\begin{aligned} P(F) &= P(F \mid D)P(D) + P(F \mid D^c)P(D^c) \\ &= P(F \mid D)P(D) + P(F \mid D^c)[1 - P(D)] \end{aligned}$$

iii. $P(D \mid F) = \dfrac{P(F \mid D)P(D)}{P(F)} = \dfrac{P(F \mid D)P(D)}{P(F \mid D)P(D) + P(F \mid D^c)[1 - P(D)]}$

iv. $$\begin{aligned} & P(D \mid F) = \frac{P(F \mid D)P(D)}{P(F \mid D)P(D) + P(F \mid D^c)[1 - P(D)]} \\ \implies & P(D \mid F) \times \big(P(F \mid D)P(D) + P(F \mid D^c)[1 - P(D)]\big) = P(F \mid D)P(D) \\ \implies & P(F \mid D)P(D) + P(F \mid D^c)[1 - P(D)] = \frac{P(F \mid D)P(D)}{P(D \mid F)} \\ \implies & P(F \mid D^c)[1 - P(D)] = \frac{P(F \mid D)P(D)}{P(D \mid F)} - P(F \mid D)P(D) \\ \implies & P(F \mid D^c) = \frac{1}{1 - P(D)} \left(\frac{P(F \mid D)P(D)}{P(D \mid F)} - P(F \mid D)P(D) \right) \end{aligned}$$

v. First note that we have $P(D) = 0.015$ and $P(F \mid D) = 0.97$. We also have the false discovery rate $P(D^c \mid F) = 0.2$, giving $P(D \mid F) = 1 - 0.2 = 0.8$. Substituting these values gives:

$$P(F \mid D^c) = \frac{1}{1 - 0.015} \left(\frac{0.97 \times 0.015}{0.8} - 0.97 \times 0.015 \right) = 0.0037$$

vi. From part (v), the required true negative rate is $P(F^c \mid D^c) = 1 - P(F \mid D^c) = 1 - 0.0037 = 0.9963$.
Thus, we report to the manager that the required true negative rate is 0.996.