ST102/ST109 Outline solutions to Exercise 7

1. Let X denote the number of components which will fail, hence $X \sim \text{Bin}(15, 0.4)$. Therefore:

$$\begin{aligned}
P(X \geq 3 \,|\, X \geq 2) = \frac{P(X \geq 3)}{P(X \geq 2)} &= \frac{1 - P(X = 0) - P(X = 1) - P(X = 2)}{1 - P(X = 0) - P(X = 1)} \\
&= \frac{1 - 0.0005 - 0.0047 - 0.0219}{1 - 0.0005 - 0.0047} \\
&= \frac{0.9729}{0.9948} \\
&= 0.9780.
\end{aligned}$$

2.* We require:

$$P\left(X_1 = 1 \,\middle|\, \sum_{i=1}^n X_i = k\right) = \frac{P\left(X_1 = 1 \text{ and } \sum_{i=1}^n X_i = k\right)}{P\left(\sum_{i=1}^n X_i = k\right)} = \frac{P\left(X_1 = 1 \text{ and } \sum_{i=2}^n X_i = k - 1\right)}{P\left(\sum_{i=1}^n X_i = k\right)}.$$

Since the random variables $X_1, X_2, \ldots, X_n$ are independent, it follows that X_1 and $\displaystyle\sum_{i=2}^n X_i$ are independent. Therefore, the final expression above can be written as:

$$\frac{P(X_1 = 1)\, P\left(\sum_{i=2}^n X_i = k - 1\right)}{P\left(\sum_{i=1}^n X_i = k\right)}.$$

We note $\displaystyle\sum_{i=2}^n X_i \sim \text{Bin}(n-1, \pi)$, and $\displaystyle\sum_{i=1}^n X_i \sim \text{Bin}(n, \pi)$. Hence:

$$P\left(\sum_{i=2}^n X_i = k - 1\right) = \binom{n-1}{k-1} \pi^{k-1} (1-\pi)^{(n-1)-(k-1)} = \binom{n-1}{k-1} \pi^{k-1} (1-\pi)^{n-k}$$

and:

$$P\left(\sum_{i=1}^n X_i = k\right) = \binom{n}{k} \pi^k (1-\pi)^{n-k}.$$

We also have $P(X_1 = 1) = \pi$. It now follows that:

$$P\left(X_1 = 1 \,\middle|\, \sum_{i=1}^n X_i = k\right) = \frac{\binom{n-1}{k-1} \pi^k (1-\pi)^{n-k}}{\binom{n}{k} \pi^k (1-\pi)^{n-k}} = \frac{k}{n}.$$

3.* Let Y denote the number of typos on a given page, such that $Y \sim \text{Pois}(\lambda)$. The probability, π, that a given page will contain more than k typos is:

$$\pi = P(Y > k) = \sum_{y=k+1}^{\infty} p_Y(y) = \sum_{y=k+1}^{\infty} \frac{\mathrm{e}^{-\lambda}\lambda^y}{y!}$$

and so:

$$1 - \pi = \sum_{y=0}^{k} \frac{\mathrm{e}^{-\lambda}\lambda^y}{y!}.$$

Now let X denote the number of pages, among the n pages in the book, on which there are more than k typos. Hence, for $0, 1, 2, \ldots, n$, we have:

$$P(X = x) = \binom{n}{x} \pi^x (1-\pi)^{n-x}$$

and so:

$$P(X \geq m) = \sum_{x=m}^{n} \binom{n}{x} \pi^x (1-\pi)^{n-x}$$

where π and $1 - \pi$ are as defined above.

4. (a) If $X \sim \text{Bin}(n, \pi)$, then the moment generating function (mgf) of X is:

$$M_X(t) = \mathrm{E}(\mathrm{e}^{tX}) = \sum_{x=0}^{n} \mathrm{e}^{tx} \binom{n}{x} \pi^x (1-\pi)^{n-x} = \sum_{x=0}^{n} \binom{n}{x} (\mathrm{e}^t\pi)^x (1-\pi)^{n-x}.$$

The binomial theorem states that, for any integer $n \geq 0$ and any real numbers y and z, we have:

$$\sum_{x=0}^{n} \binom{n}{x} y^x z^{n-x} = (y+z)^n.$$

The expression for the binomial mgf is of this form, if we set $y = \mathrm{e}^t\pi$ and $z = 1 - \pi$. Therefore:

$$M_X(t) = \left(\mathrm{e}^t\pi + (1-\pi)\right)^n.$$

(b) Using the binomial mgf:

$$M_X'(t) = n\pi\mathrm{e}^t(\mathrm{e}^t\pi + (1-\pi))^{n-1}$$

and:

$$M_X''(t) = n(n-1)(\pi\mathrm{e}^t)^2 \left(\mathrm{e}^t\pi + (1-\pi)\right)^{n-2} + n\pi\mathrm{e}^t \left(\mathrm{e}^t\pi + (1-\pi)\right)^{n-1}.$$

So, since $\mathrm{e}^0 = 1$, we have:

$$\mathrm{E}(X) = M_X'(0) = n\pi$$

$$\mathrm{E}(X^2) = M_X''(0) = n(n-1)\pi^2 + n\pi$$

and:

$$\text{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2 = n^2\pi^2 - n\pi^2 + n\pi - n^2\pi^2 = n\pi - n\pi^2 = n\pi(1-\pi).$$

5. (a) If we assume that the calving process is random (as the remark about seasonality hints) then we are counting events over periods of time (with, in particular, no obvious upper maximum), and hence the appropriate distribution is the Poisson distribution.

(b) The rate parameter for one week is 0.4, so for three weeks we use $\lambda = 1.2$, hence:

$$P(X=0) = \frac{\mathrm{e}^{-1.2} \times (1.2)^0}{0!} = \mathrm{e}^{-1.2} = 0.3012.$$

(c) If it is correct to use the Poisson distribution then events are independent, and hence:

$$P(\text{none in weeks 1, 2 \& 3}) = P(\text{none in weeks 4, 5 \& 6}) = 0.3012.$$

(d) The rate parameter for four weeks is $\lambda = 1.6$, hence:

$$P(X=5) = \frac{\mathrm{e}^{-1.6} \times (1.6)^5}{5!} = 0.0176.$$

(e) Bayes' theorem tells us that:

$$P(\text{5 in weeks 5 to 8} \,|\, \text{5 in weeks 1 to 4}) = \frac{P(\text{5 in weeks 5 to 8} \,\cap\, \text{5 in weeks 1 to 4})}{P(\text{5 in weeks 1 to 4})}.$$

If it is correct to use the Poisson distribution then events are independent. Therefore:

$$P(\text{5 in weeks 5 to 8} \cap \text{5 in weeks 1 to 4}) = P(\text{5 in weeks 5 to 8})\, P(\text{5 in weeks 1 to 4}).$$

So, cancelling, we get:

$$\begin{aligned} P(\text{5 in weeks 5 to 8} \,|\, \text{5 in weeks 1 to 4}) &= P(\text{5 in weeks 5 to 8}) \\ &= P(\text{5 in weeks 1 to 4}) \\ &= 0.0176. \end{aligned}$$

(f) The fact that the results are identical in the two cases is a consequence of the independence built into the assumption that the Poisson distribution is the appropriate one to use. A Poisson process does not 'remember' what happened before the start of a period under consideration.

Discrete uniform distributions in R (for reference only)

Not required, but below is an example of a histogram from:

```
> hist(sample(100,50000,replace=T))
```

Note the bar heights are not all exactly equal. This is due to random variation which we expect in sampling.