# ST102/ST109 Outline solutions to Exercise 9

1. (a) The Poisson distribution with parameter $\lambda$ has its expectation and variance both equal to $\lambda$, so we should take $\mu = \lambda$ and $\sigma^2 = \lambda$ in a normal approximation, i.e. use a $N(\lambda, \lambda)$ distribution as the approximating distribution.

   (b) $P(X > 10) \approx P(Y > 10.5)$ using a continuity correction, where $Y \sim N(14, 14)$. This is:

$$P(Y > 10.5) = P\left(\frac{Y - 14}{\sqrt{14}} > \frac{10.5 - 14}{\sqrt{14}}\right) = P(Z > -0.94) = 0.8264.$$

2.* (a) With a population of just 20, and 40% of them, i.e. 8, being ST102 FC supporters, the probability of success (a 'yes' response) changes each time a respondent is asked, and the change depends on the respondent's answer. For instance:

- the probability is $\pi_1 = 0.4$ that the first respondent supports ST102 FC
- if the first respondent supports ST102 FC, the probability is $\pi_{2|1=\text{'yes'}} = 7/19 = 0.3684$ that the second respondent supports ST102 FC too
- if the first respondent does not support ST102 FC, the probability is $\pi_{2|1=\text{'no'}} = 8/19 = 0.4211$ that the second respondent supports ST102 FC.

This means that we cannot assume that the probability of a 'yes' is (virtually) the same from one respondent to the next. Therefore, we cannot regard the successive questions to different people as being successive identical Bernoulli trials. Equivalently, we must treat the problem as if we were using sampling *without* replacement.

So, if '$S$' means 'supports ST102 FC' and '$S^c$' means 'does not support ST102 FC', there are four possible sequences of responses which give exactly three $S$s, and we want the sum of their separate probabilities. So we have:

$$P(SSSS^c) = \frac{8}{20} \times \frac{7}{19} \times \frac{6}{18} \times \frac{12}{17} = \frac{12 \times 8 \times 7 \times 6}{20 \times 19 \times 18 \times 17}$$

$$P(SSS^cS) = \frac{8}{20} \times \frac{7}{19} \times \frac{12}{18} \times \frac{6}{17} = \frac{12 \times 8 \times 7 \times 6}{20 \times 19 \times 18 \times 17}$$

$$P(SS^cSS) = \frac{8}{20} \times \frac{12}{19} \times \frac{7}{18} \times \frac{6}{17} = \frac{12 \times 8 \times 7 \times 6}{20 \times 19 \times 18 \times 17}$$

$$P(S^cSSS) = \frac{12}{20} \times \frac{8}{19} \times \frac{7}{18} \times \frac{6}{17} = \frac{12 \times 8 \times 7 \times 6}{20 \times 19 \times 18 \times 17}.$$

Therefore, $P(\text{exactly 3 } S\text{s}) = 4 \times (12 \times 8 \times 7 \times 6)/(20 \times 19 \times 18 \times 17) = 0.1387$.

   (b) In this case we can assume that the probability of an '$S$' is (for practical purposes) constant, hence we have a sequence of independent and identical Bernoulli trials (and/or that we are using sampling *with* replacement). Therefore, we can assume that, if $X$ is the random variable which counts the number of $S$s out of forty people surveyed, then $X \sim \text{Bin}(40, 0.4)$. Hence:

$$P(X = 20) = \binom{40}{20} \times (0.4)^{20} \times (0.6)^{20} = 0.0554.$$

Alternatively, we could use a normal approximation to the binomial distribution where $Y \sim N(16, 9.6)$, since $\text{E}(X) = 16$ and $\text{Var}(X) = 9.6$. Hence (using the continuity correction):

$$\begin{aligned} P(X = 20) = P(19.5 \leq Y \leq 20.5) &= P\left(\frac{19.5 - 16}{\sqrt{9.6}} \leq Z \leq \frac{20.5 - 16}{\sqrt{9.6}}\right) \\ &= P(1.13 \leq Z \leq 1.45) \\ &= 1 - P(Z > 1.45) - (1 - P(Z > 1.13)) \\ &= 1 - 0.0735 - (1 - 0.1292) \\ &= 0.0557. \end{aligned}$$

(c) With $n = 100$ (and a 'large' population) we know that we can use the normal approximation to $X \sim \text{Bin}(100, 0.4)$. Since $\text{E}(X) = 40$ and $\text{Var}(X) = 24$, we can approximate $X$ with $Y$, where $Y \sim N(40, 24)$. Hence (using the continuity correction):

$$\begin{aligned} P(X \geq 30) = P(Y \geq 29.5) &= P\left(Z \geq \frac{29.5 - 40}{\sqrt{24}}\right) \\ &= P(Z \geq -2.14) \\ &= 1 - P(Z \geq 2.14) \\ &= 0.98382. \end{aligned}$$

(d) The only approximation used here is the normal approximation to the binomial distribution, used in (b) and (c), and it is justified because:

- $n$ is 'large' (although whether $n = 40$ in (b) can be considered large is debatable)
- the population is large enough to justify using the binomial in the first place
- $n\pi > 5$ and $n(1 - \pi) > 5$ in each case.

(e) Three possible comments are the following.

- For (a), if we had (wrongly) used $\text{Bin}(4, 0.4)$ then we would have obtained $P(X = 3) = 0.1536$, which is quite a long way from the true value (roughly 11%, proportionally) – we might have reached the wrong conclusion.
- For (b) and (c) it was important, to justify using the binomial distribution, that the population was 'large'.
- The true value (to 4 decimal places) for (c) is 0.9852, so the approximation obtained is pretty good – it is very unlikely that we might have reached the wrong conclusion.

3. (a) The marginal distributions are found by adding across rows and columns:

| $X = x$ | $-1$ | $0$ | $1$ |
|---|---|---|---|
| $p_X(x)$ | 0.45 | 0.30 | 0.25 |

and:

| $Y = y$ | $0$ | $1$ |
|---|---|---|
| $p_Y(y)$ | 0.65 | 0.35 |

(b) We have:

$$\mathrm{E}(X) = -1 \times 0.45 + 0 \times 0.30 + 1 \times 0.25 = -0.20$$

and:

$$\mathrm{E}(X^2) = (-1)^2 \times 0.45 + 0^2 \times 0.30 + 1^2 \times 0.25 = 0.70$$

so $\mathrm{Var}(X) = 0.70 - (-0.20)^2 = 0.66$. Also:

$$\mathrm{E}(Y) = 0 \times 0.65 + 1 \times 0.35 = 0.35$$

and:

$$\mathrm{E}(Y^2) = 0^2 \times 0.65 + 1^2 \times 0.35 = 0.35$$

so $\mathrm{Var}(Y) = 0.35 - (0.35)^2 = 0.2275$.

(c) The conditional probability functions $p_{Y|X=-1}(y \mid x = -1)$ and $p_{X|Y=0}(x \mid y = 0)$ are given by, respectively:

| $Y = y \mid X = -1$ | $0$ | $1$ |
|---|---|---|
| $p_{Y\|X=-1}(y \mid x = -1)$ | $0.30/0.45 = 0.\dot{6}$ | $0.15/0.45 = 0.\dot{3}$ |

and:

| $X = x \mid Y = 0$ | $-1$ | $0$ | $1$ |
|---|---|---|---|
| $p_{X\|Y=0}(x \mid y = 0)$ | $0.30/0.65 = 0.4615$ | $0.25/0.65 = 0.3846$ | $0.10/0.65 = 0.1538$ |

(d) We have $\mathrm{E}_{Y|X}(Y \mid X = -1) = 0 \times 0.\dot{6} + 1 \times 0.\dot{3} = 0.\dot{3}$.
Also, $\mathrm{E}_{X|Y}(X \mid Y = 0) = -1 \times 0.4615 + 0 \times 0.3846 + 1 \times 0.1538 = -0.3077$.

(e) We have $\mathrm{E}(XY) = \sum_x \sum_y xy\, p(x, y) = -1 \times 0.15 + 0 \times 0.70 + 1 \times 0.15 = 0$.
Also, $\mathrm{Cov}(X, Y) = \mathrm{E}(XY) - \mathrm{E}(X)\,\mathrm{E}(Y) = 0 - (-0.20)(0.35) = 0.07$ and:

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}} = \frac{0.07}{\sqrt{0.66 \times 0.2275}} = 0.1807.$$

(f) We have $P(X > Y) = P(X = 1, Y = 0) = 0.10$.
$P(X^2 > Y^2) = P(X = -1, Y = 0) + P(X = 1, Y = 0) = 0.30 + 0.10 = 0.40$.

(g) Since $X$ and $Y$ are (weakly) positively correlated (as determined in (e)), they cannot be independent.
While the non-zero correlation is a sufficient explanation in this case, for other such bivariate distributions which are uncorrelated, i.e. when $\mathrm{Corr}(X, Y) = 0$, it becomes necessary to check whether $p_{X,Y}(x, y) = p_X(x)\, p_Y(y)$ for all pairs of values of $(x, y)$. Here, for example, $p_{X,Y}(0, 0) = 0.25$, $p_X(0) = 0.30$ and $p_Y(0) = 0.65$. We then have that $p_X(0)\, p_Y(0) = 0.195$, which is not equal to $p_{X,Y}(0, 0) = 0.25$. Hence $X$ and $Y$ cannot be independent.

4.* (a) We have:

$$P(X=0,Y=0)=\frac{3}{10}\times\frac{2}{9}=\frac{6}{90}=\frac{1}{15}$$

$$P(X=0,Y=1)=2\times\frac{3}{10}\times\frac{3}{9}=\frac{18}{90}=\frac{3}{15}$$

$$P(X=0,Y=2)=\frac{3}{10}\times\frac{2}{9}=\frac{6}{90}=\frac{1}{15}$$

$$P(X=1,Y=0)=2\times\frac{4}{10}\times\frac{3}{9}=\frac{24}{90}=\frac{4}{15}$$

$$P(X=1,Y=1)=2\times\frac{4}{10}\times\frac{3}{9}=\frac{24}{90}=\frac{4}{15}$$

$$P(X=2,Y=0)=\frac{4}{10}\times\frac{3}{9}=\frac{12}{90}=\frac{2}{15}.$$

All other values have probability 0. We then construct the table of joint probabilities:

| | $Y=0$ | $Y=1$ | $Y=2$ |
|---|---|---|---|
| $X=0$ | $1/15$ | $3/15$ | $1/15$ |
| $X=1$ | $4/15$ | $4/15$ | $0$ |
| $X=2$ | $2/15$ | $0$ | $0$ |

(b) The number of blue balls in the sample.

(c) We have:

$$\mathrm{E}(X)=1\times\left(\frac{4}{15}+\frac{4}{15}\right)+2\times\frac{2}{15}=\frac{4}{5}$$

$$\mathrm{E}(Y)=1\times\left(\frac{3}{15}+\frac{4}{15}\right)+2\times\frac{1}{15}=\frac{3}{5}$$

and:

$$\mathrm{E}(XY)=1\times 1\times\frac{4}{15}=\frac{4}{15}.$$

So:

$$\mathrm{Cov}(X,Y)=\frac{4}{15}-\frac{4}{5}\times\frac{3}{5}=-\frac{16}{75}.$$

(d) We have:

$$P(X=1\,|\,|X-Y|<2)=\frac{4/15+4/15}{1/15+3/15+4/15+4/15}=\frac{2}{3}.$$