1. Each of the 150 ENGG2430 students shows up to class independently with probability 0.9 and asks Poisson(0.05) questions in there. Let S be the number of students in class and Q the total number of questions asked. Find (a) $\mathrm{E}[S]$, (b) $\mathrm{E}[Q|S]$, (c) $\mathrm{E}[Q]$, (d) $\mathrm{Var}[\mathrm{E}[Q|S]]$, (e) $\mathrm{Var}[Q|S]$, (f) $\mathrm{E}[\mathrm{Var}[Q|S]]$, (g) $\mathrm{Var}[Q]$.

   Solution: Let Q_i be the number of question asked by the i-th student present in class; $Q = Q_1 + \cdots + Q_S$.

   (a) $\mathrm{E}[S] = 150 \cdot 0.9 = 135$.
   (b) $\mathrm{E}[Q|S] = \sum_{i=1}^{S} \mathrm{E}[Q_i] = S \cdot 0.05 = 0.05S$ by linearity of expectation.
   (c) $\mathrm{E}[Q] = \mathrm{E}[\mathrm{E}[Q|S]] = \mathrm{E}[0.05S] = 0.05 \cdot 135 = 6.75$ by (b).
   (d) $\mathrm{Var}[\mathrm{E}[Q|S]] = \mathrm{Var}[0.05S] = 0.05^2\,\mathrm{Var}[S] = 0.05^2 \cdot (150 \cdot 0.9 \cdot 0.1) = 0.03375$ by (b).
   (e) $\mathrm{Var}[Q|S] = \sum_{i=1}^{S} \mathrm{Var}[Q_i] = S \cdot 0.05 = 0.05S$ by independence of Q_i's.
   (f) $\mathrm{E}[\mathrm{Var}[Q|S]] = \mathrm{E}[0.05S] = 6.75$ by (e).
   (g) $\mathrm{Var}[Q] = \mathrm{Var}[\mathrm{E}[Q|S]] + \mathrm{E}[\mathrm{Var}[Q|S]] = 6.78375$ by (d) and (f).

2. 100 people put their hats in a box and each one pulls a random hat out.

   (a) Let G be any 10-person group. What is the probability that everyone in G pulls their own hat?
   (b) What is the expected *number* of 10-person groups in which everyone pulls their own hat?
   (c) Show that the probability that 10 or more people pull their own hat is less than 10^{-6}.

   Solution:

   (a) The probability that the first person in the group pulls their own hat is $1/100$. Given this happened, the probability that the second person in the group does so is $1/99$, and so on. So the probability that everyone in the group succeeds is $1/(100 \cdot 99 \cdots 91)$.
   (b) Let X_S be the random variable indicating that everyone in group S pulled their own hat. Then the number of people who pulled their own hat X is the sum of the random variables X_S. By linearity of expectation, $\mathrm{E}[X]$ is the sum of $\mathrm{E}[X_S] = 1/(100 \cdot 99 \cdots 91)$ over all 10-person groups S. As there are $\binom{100}{10}$ ways to choose a 10-person group,

   $$\mathrm{E}[X] = \binom{100}{10} \cdot \frac{1}{100 \cdot 99 \cdots 91} = \frac{1}{10!}.$$

   (c) By Markov's inequality, the probability that at least one group succeeded in pulling all of their own hats is at most

   $$\mathrm{P}(X \geq 1) \leq \frac{\mathrm{E}[X]}{1} = \frac{1}{10!} \approx 2.7557 \times 10^{-7} < 10^{-6}$$

3. In a school fair, you put up a game stall. In each game, the participant pays you \$10, he or she then draws a ball from a box of 9 white balls and 1 red ball, if the ball drawn is red, you pay \$40 back, otherwise the participant gains nothing. Estimate the probability that you have gained \$300 after 100 games.

Solution: Let \$$X$ be the total amount of money collected. We want to estimate $\mathrm{P}(X \geq 300)$. X is the sum of 100 independent random variables with the same PMFs so we can use the Central Limit Theorem. We have

$$\mu = \mathrm{E}[X] = 100 \times (10 \times 0.9 + (-30) \times 0.1) = 600$$
$$\sigma = \sqrt{\mathrm{Var}[X]} = \sqrt{100 \times ((10-6)^2 \times 0.9 + (-30-6)^2 \times 0.1)} = \sqrt{100 \cdot 144} = 120$$

Therefore,

$$\mathrm{P}(X \geq 300) \approx \mathrm{P}(X \geq \mu - 2.5\sigma) \approx \mathrm{P}(N \geq -2.5) \approx 0.9938,$$

where N is a Normal$(0, 1)$ random variable.

4. 100 balls are tossed at random into 100 bins. Each ball is equally likely to land in any of the bins, independently of the other balls.

   (a) Find the expected number and variance of the number of non-empty bins.

   (b) Show that there are fewer than 80 non-empty bins with a probability at least 90%.

Solution:

(a) It is a bit easier to count the number E of empty bins. The number N of non-empty bins is then $100 - E$. We can write E as $E_1 + \cdots + E_{100}$, where E_i indicates that bin i is empty. By linearity of expectation,

$$\mathrm{E}[E] = \mathrm{E}[E_1] + \mathrm{E}[E_2] + \cdots + \mathrm{E}[E_{100}] = \sum_{i=1}^{100} \mathrm{P}(X_i = 1) = 100 \cdot p,$$

where $p = 0.99^{100}$, so $\mathrm{E}[N] = 100 - 100 \cdot 0.99^{100} \approx 63.3968$. To calcluate the variance we apply the sum of covariances formula:

$$\mathrm{Var}[E] = \sum_{i=1}^{100} \mathrm{Var}[E_i] + \sum_{i \neq j} \mathrm{Cov}[E_i, E_j].$$

Each of the variances $\mathrm{Var}[E_i]$ equals $p(1-p) = 0.99^{100}(1 - 0.99^{100})$. As for the covariances,

$$\mathrm{Cov}[\mathrm{E_i}, \mathrm{E_j}] = \mathrm{E}[E_i E_j] - \mathrm{E}[E_i]\,\mathrm{E}[E_j] = \mathrm{P}(E_i = 1 \text{ and } E_j = 1) - \mathrm{P}(E_i = 1)\,\mathrm{P}(E_j = 1).$$

The probability that both bins i and j are empty is 0.98^{100} as all the balls must go into the other 98 bins, so each covariance term equals $0.98^{100} - (0.99^{100})^2$. Putting everything together we get

$$\mathrm{Var}[E] = 100 \cdot (0.99)^{100} \cdot (1 - 0.99^{100}) + 100 \cdot 99 \cdot \left(0.98^{100} - (0.99^{100})^2\right) \approx 9.7401.$$

Since $N = 100 - E$, N has the same variance as E.

(b) The expectation of N is $\mu \approx 63.3968$ and its standard deviation is $\sigma \approx \sqrt{9.7401} \approx 3.1209$. By Chebyshev's inequality,

$$\Pr(N \geq 80) \leq \Pr(N \geq \mu + 5.3200\sigma) \leq \Pr(|N - \mu| \geq 5.3200\sigma) \leq 1/5.3200^2 \approx 0.0353,$$

so $\mathrm{P}(N < 80) \geq 1 - 0.0353 = 0.9647 > 95\%$ as required.
For comparison, Markov's inequality gives a much looser bound of

$$\mathrm{P}(N < 80) = 1 - \mathrm{P}(N \geq 80) \geq 1 - 63.3968/80 \approx 0.2075.$$

The Central Limit Theorem does not apply because the E_i are not independent.

5. Consider the following simplified model of infection spread. On any given day, any carrier independently infects one additional person with probability p and is cured with probability $1-p$. The number X_d of virus carriers on day d is given by $X_d = 2 \cdot \text{Binomial}(X_{d-1}, p)$.

   (a) Let $e_d = \text{E}[X_d]$. Express e_d in terms of e_{d-1}. What is e_d in terms of X_0, p, and d?

   (b) Show that when $X_0 = 100$ and $p = 0.4$, the probability 100 or more people are carriers on day 21 is less than 1%.

   (c) Let $v_d = \text{Var}[X_d]$. Express v_d in terms of v_{d-1}.

   (d) **(Optional)** Show that when $X_0 = 100$ and $p = 0.6$, the probability that 100 or more people are carriers on day 21 is more than 95%.

Solution:

(a) Since X_d is $2 \cdot \text{Binomial}(X_{d-1}, p)$, $\text{E}[X_d|X_{d-1}] = 2X_{d-1}p$ and

$$e_d = \text{E}[\text{E}[X_d|X_{d-1}]] = \text{E}[2X_{d-1}p] = 2pe_{d-1}$$

Applying the relation recursively and using $e_0 = X_0$, we have

$$e_d = 2pe_{d-1} = (2p)^2 e_{d-2} = \cdots = (2p)^d e_0 = (2p)^d X_0$$

(b) When $X_0 = 100$, $p = 0.4$, $e_{21} = 100(0.8)^{21}$. By Markov's inequality,

$$\text{P}(X_{21} \geq 100) \leq \frac{\text{E}[X_{21}]}{100} = (0.8)^{21} \approx 0.0092 < 0.01$$

(c) By the total variance theroem,

$$\begin{aligned} v_d &= \text{E}\big[\text{Var}[X_d|X_{d-1}]\big] + \text{Var}\big[\text{E}[X_d|X_{d-1}]\big] \\ &= \text{E}[2^2 \cdot X_{d-1}p(1-p)] + \text{Var}[2X_{d-1}p] \\ &= 4p(1-p) \cdot (2p)^{d-1} X_0 + (2p)^2 v_{d-1} \\ &= 2(1-p)X_0 \cdot (2p)^d + 4p^2 v_{d-1} \end{aligned}$$

(d) Set $C = 2(1-p)X_0$, then we can write $v_d = C(2p)^d + (2p)^2 v_{d-1}$.
Applying the relation recursively and using $v_0 = 0$, we have

$$\begin{aligned} v_d &= C(2p)^d + (2p)^2 v_{d-1} \\ &= C(2p)^d + (2p)^2 \cdot (C(2p)^{d-1} + (2p)^2 v_{d-2}) \\ &= C(2p)^d + C(2p)^{d+1} + (2p)^4 v_{d-2} \\ &= C(2p)^d + C(2p)^{d+1} + (2p)^4 \cdot (C(2p)^{d-2} + (2p)^2 v_{d-3}) \\ &= C(2p)^d + C(2p)^{d+1} + C(2p)^{d+2} + (2p)^6 v_{d-3} \\ &= \cdots \\ &= C(2p)^d + C(2p)^{d+1} + C(2p)^{d+2} + \cdots + C(2p)^{2d-1} + (2p)^{2d} v_0 \\ &= C(2p)^d \cdot \frac{(2p)^d - 1}{2p - 1} \end{aligned}$$

For $X_0 = 100$, $p = 0.6$ and $d = 21$, $C = 80$ and $v_d = 400(1.2)^{21}(1.2^{21} - 1)$.
Then $\mu = \text{E}[X] \approx 4600.51$, $\sigma = \sqrt{\text{Var}[X]} \approx 910.05$.
By Chebyshev's inequality, we have:

$$\text{P}(X \geq 100) \approx \text{P}(X \geq \mu - 4.9453\sigma) \geq \text{P}(|X - \mu| \leq 4.94\sigma) \geq 1 - \frac{1}{4.94^2} \approx 0.9590.$$