1. Each of the 200 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) $E[S]$, (b) $E[Q|S]$, (c) $E[Q]$, (d) $\text{Var}[E[Q|S]]$, (e) $\text{Var}[Q|S]$, (f) $E[\text{Var}[Q|S]]$, (g) $\text{Var}[Q]$.

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) $E[S] = 200 \cdot 0.9 = 180$.

(b) $E[Q|S] = \sum_{i=1}^{S} E[Q_i] = S \cdot 0.05 = 0.05S$ by linearity of expectation.

(c) $E[Q] = E[E[Q|S]] = E[0.05S] = 0.05 \cdot 180 = 9$ by (b).

(d) $\text{Var}[E[Q|S]] = \text{Var}[0.05S] = 0.05^2 \cdot 0.05S = 0.0025 \cdot 200 \cdot 0.9 \cdot 0.1 = 0.045$ by (b).

(e) $\text{Var}[Q|S] = \sum_{i=1}^{S} \text{Var}[Q_i] = S \cdot 0.05 = 0.05S$ by independence of $Q_i$’s.

(f) $E[\text{Var}[Q|S]] = E[0.05S] = 9$ by (e).

(g) $\text{Var}[Q] = \text{Var}[E[Q|S]] + E[\text{Var}[Q|S]] = 9.00045$ by (d) and (f).

2. You flip a coin with unknown probability of heads $p$. You want to learn the value of $p$.

(a) Alice suggests the following estimator $\hat{P}_A$: Keep flipping the coin until you see the first head in the $N$-th flip. Set $\hat{P}_A = 1/N$.

(b) Bob suggests another estimator $\hat{P}_B$: Flip the coin 10 times, count the number of heads $Y$ and set $\hat{P}_B = Y/10$.

What is the expectation of each estimator in terms of $p$? Which one is better?

Solution: $\hat{P}_A$ has expectation

$$E[\hat{P}_A] = \sum_{n=1}^{\infty} \frac{1}{n} \cdot (1-p)^{n-1}p = \frac{p}{1-p} \sum_{n=1}^{\infty} \frac{(1-p)^n}{n} \approx \frac{p}{1-p} \cdot (-\log p),$$

since the infinite series is the Taylor expansion of $-\log p$. $\hat{P}_B$ has expectation

$$E[\hat{P}_B] = E\left[\frac{Y}{10}\right] = \frac{1}{10} E[Y] = \frac{10p}{10} = p.$$

The expectation of $\hat{P}_A \neq p$ in general while $\hat{P}_B$ does; $\hat{P}_A$ is said to be biased and $\hat{P}_B$ unbiased.

In the next two questions, estimate the quantity of your interest using the method of your choice: Markov’s inequality, Chebyshev’s inequality, or the Central Limit Theorem. Justify why the method is applicable and discuss the quality of the estimate.

Summary of the assumption and result of the three methods:

- **Markov’s inequality:** $P(X \geq a) \leq E[X]/a$.
  - Applies for any non-negative random variable $X$, and any $a > 0$ ($a > E[X]$ for a meaningful bound).
  - Requires only $E[X]$, useful when it is small.
  - Is an upper bound to a one-sided (right tail) probability.
• Chebyshev’s inequality: \( P(|X - \mu| \geq t\sigma) \leq \frac{1}{t^2} \), where \( \mu = E[X] \), \( \sigma = \sqrt{\text{Var}[X]} \).
  
  – Applies for any random variable \( X \) (with finite \( \mu, \sigma \)), and any \( t \) (\( t > 1 \) for a meaningful bound).
  
  – Requires both expectation and variance of \( X \).
  
  – Is an upper bound to a two-sided probability.

• Central Limit Theorem: \( \frac{X - \mu_X}{\sigma_X} \approx \text{Normal}(0, 1) \), where \( X = X_1 + \cdots + X_n \) are independent and have the same PDF/PMF, \( \mu_X = E[X] = n E[X_i], \sigma_X = \sqrt{\text{Var}[X]} = \sqrt{n \text{Var}[X_i]} \).
  
  – Applies for \( X \) being sum of many (usually \( n \geq 30 \)) independent random variables; no restriction on distribution of \( X_i \)’s.
  
  – Requires both \( E[X_i] \) and \( \text{Var}[X_i] \).
  
  – Approximates the CDF of \( X \). Using the axioms of probability, we can use it to approximate other events of interest (e.g. \( P(X < -5 \) or \( X > 7) \)).

Roughly speaking, in terms of generality, Markov’s inequality > Chebyshev’s inequality > CLT; and in terms of tightness, Markov’s inequality < Chebyshev’s inequality < CLT (if \( n \) is large enough).

3. The following exam statistics are posted on the course website:

<table>
<thead>
<tr>
<th>section</th>
<th>no. students</th>
<th>average grade</th>
<th>std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>30</td>
<td>65</td>
<td>5</td>
</tr>
<tr>
<td>B</td>
<td>20</td>
<td>70</td>
<td>10</td>
</tr>
</tbody>
</table>

what can you say about the number of students whose exam grade was 30 or below?

**Solution:** Let \( X_A \) and \( X_B \) be the grade of a random student in section A and section B, respectively. The table tells us that \( \mu_A = E[X_A] = 65, \sigma_A = \sqrt{\text{Var}[X_A]} = 5, \mu_B = E[X_B] = 70, \sigma_B = \sqrt{\text{Var}[X_B]} = 10 \). By Chebyshev’s inequality, for a random student in section A,

\[
P(X_A \leq 30) = P(X_A \leq \mu_A - 7 \cdot \sigma_A) \leq P(|X_A - \mu_A| \geq 7\sigma_A) \leq \frac{1}{1/49} \approx 0.0204.
\]

Although we are only interested in the probability that \( X_A \) is 7 standard deviations smaller than its mean, Chebyshev’s inequality only tells us the probability of the possibly larger event that \( X_A \) is either 7 standard deviations smaller or 7 standard deviations larger than its mean. This is already a tremendously small probability – about 2%.

Similarly, for a random student in section B,

\[
P(X_B \leq 30) = P(X_B \leq \mu_B - 4 \cdot \sigma_B) \leq P(|X_B - \mu_B| \geq 4\sigma_B) \leq \frac{1}{1/16} \approx 0.00625.
\]

Since there are 30 students in section A, at most \( 1/49 \cdot 30 \) students must have received 30 or below, so nobody got that kind of grade. In section B, at most \( 1/16 \cdot 20 \) students got 30 or below, so at most one student in the whole class could have received 30 or below on the exam.

**Alternative Solution:** Alternatively, we can first calculate the statistics for a random student \( X \) in the whole course and then apply Chebyshev’s inequality to \( X \). Let \( X \) be a random student and \( Y \) their section (using 1 and 2 for sections A and B). Then \( E[X|Y] \) takes value 65 with probability \( 3/5 \) and 70 with probability \( 2/5 \).

By total expectation theorem,

\[
\mu = E[X] = E[E[X|Y]] = 65 \cdot 3/5 + 70 \cdot 2/5 = 67.
\]
By total variance theorem $\text{Var}[X] = \text{Var}[\text{E}[X|Y]] + \text{E}[\text{Var}[X|Y]]$, where
\[
\text{Var}[\text{E}[X|Y]] = (65 - 67)^2 \cdot 3/5 + (70 - 67)^2 \cdot 2/5 = 6,
\]
\[
\text{E}[\text{Var}[X|Y]] = 5^2 \cdot 3/5 + 10^2 \cdot 2/5 = 55,
\]

hence standard deviation of $X$ is $\sigma = \sqrt{61} \approx 7.8103$. Chebyshev’s inequality says
\[
P(X \leq 30) = P(X \leq \mu - (37/\sqrt{61}) \cdot \sigma) \leq P(|X - \mu| \geq (37/\sqrt{61}) \cdot \sigma) \leq 61/37^2 \approx 0.0445,
\]
so the number of students who got 30 or below is at most $0.0445 \cdot 50 = 2.2225$, so at most 2.

4. You are collecting donations for a charity. Each donor gives you $10 with probability half and $20 with probability half. Assuming donors are independent, estimate the probability that you have collected at least $1200 after taking in 100 donations.

**Solution:** Let $X$ be the total amount of money collected. We want to estimate $P(X \geq 1200)$. $X$ is the sum of 100 random variables with the same PMF so we can use the Central Limit Theorem. We have
\[
\mu = \text{E}[X] = 100 \cdot (10 \cdot 1/2 + 20 \cdot 1/2) = 1500
\]
\[
\sigma = \sqrt{\text{Var}X} = \sqrt{100 \cdot ((10 - 15)^2 \cdot 1/2 + (20 - 15)^2 \cdot 1/2)} = \sqrt{100 \cdot 25} = 50.
\]

Therefore,
\[
P(X \geq 1200) \approx P(X \geq \mu - 6\sigma) \approx P(N \geq -6) \approx 1 - 9.86 \cdot 10^{-10},
\]
where $N$ is a Normal(0, 1) random variable.

5. You randomly divide 48 boys and 48 girls into teams of equal size. Show that if you divide them into 12 teams of 8 then there are no same-sex teams with probability at least 90%.

**Solution:** For $12 \geq i \geq 1$, let $X_i$ be the indicator variable that the $i$-th team consists of all boys or all girls, then $X = \sum_{i=1}^{12} X_i$. The $X_i$’s are not independent, so the Central Limit Theorem doesn’t apply.

The probability that any given team is all-boys is
\[
p = \frac{48}{96} \cdot \frac{47}{95} \cdots \frac{41}{89}
\]
using the formula for conditional probabilities (the first member is a boy, the second member is a boy given the first one is etc.). As boys and girls are symmetric, the probability the team is same-sex is $2p$. By linearity of expectation,
\[
\text{E}[X] = \text{E}[X_1] + \cdots + \text{E}[X_{12}] = 12 \cdot (2p) \approx 0.068.
\]
At this point we can proceed in two ways. We can use Markov’s inequality to conclude that $P(X \geq 1) \leq \text{E}[X]/1 \approx 0.068$, so the probability of having no same-sex teams is $P(X = 0) \approx 1 - 0.068 = 0.922$. This meets the requirement and we are done.

Alternatively, we can calculate Var$[X]$ and apply Chebyshev’s inequality, which could result in a better bound. This is feasible but a bit difficult since $X_1, \ldots, X_{12}$ are not independent so we need their covariances.