Homework 3 - Applied Stochastic Processes

kipngeno koech - bkoech

October 18, 2024

Question 1: Random Vectors and Principal Component Analysis

Reading: Random vectors are fundamental constructs in probability and statistics, allowing researchers and practitioners to analyze relationships among multiple variables simultaneously. Each component of a random vector can represent a different feature or measurement, and the joint distribution encapsulates the uncertainty inherent in those variables.

For instance, consider a random vector X - $X_1, X_2, ..., X_n$ where each X_i is a random variable. The covariance matrix of X plays a crucial role in understanding the linear relationships among the components, guiding decisions in fields such as finance, machine learning, and signal processing. Sampling from random vectors introduces excitement in multivariate analyses, where one can explore properties like independence, marginal distributions, and conditional relationships. Moreover, techniques such as principal component analysis (PCA) leverage the variance structure of these vectors to reduce dimensionality while preserving essential information.

1. **5 points** let $X = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}$, and $Y = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix}$ are related by $Y - AX$ where

$$A = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}$$

The joint PMF of X is given by:

$$P_X(X) = \begin{cases} (1-p)p^{x_3} & \text{if } X_1 < X_2 < X_3 \\ 0 & \text{otherwise} \end{cases}$$

where $x_1, x_2, x_3 \in \{0, 1, 2, \ldots\}$ and $0 < p < 1$.
Find the joint PMF $P_Y(y)$ of the transformed random vector Y.

$$Y = AX = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} = \begin{pmatrix} X_1 \\ X_2 - X_1 \\ X_3 - X_2 \end{pmatrix}$$

$$X_1 = Y_1, X_2 = Y_1 + Y_2, X_3 = Y_1 + Y_2 + Y_3$$

$$P_Y(y) = P_X(A^{-1}y) = P_X\left(\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}\right) = P_X\left(\begin{pmatrix} y_1 \\ y_1 + y_2 \\ y_1 + y_2 + y_3 \end{pmatrix}\right)$$

conditions for $X_1 < X_2 < X_3$ to hold:

$$y_1 < y_1 + y_2 < y_1 + y_2 + y_3$$

$$0 < y_2 < y_3$$

$$P_Y(y) = (1-p)p^{y_1+y_2+y_3} = (1-p)p^{y_1}p^{y_2}p^{y_3}$$

$$P_Y(y) = (1-p)p^{y_1}p^{y_2}p^{y_3}$$

2. You are working as a data analyst for a startup that collects various statistics from users' activities on its platform. The startup wants to reduce the dimensionality of its collected data without losing significant information. Your goal is to apply Principal Component Analysis (PCA) to the dataset to retain as much variance (information) as possible while reducing the dimensionality. This exercise will take you from the conceptual understanding of random vectors and covariance matrices to the practical application of PCA using Python.

Part 1: understanding the covariance Matrix of Random Vectors (12 points)
You are given a random vector $X = [X_1, X_2, X_3, X_4]^T$, representing four features of platforms users. The covariance matrix of this random vector is:

$$\Sigma_x = \begin{bmatrix} 5 & 1.2 & 0.8 & 0.6 \\ 1.2 & 4 & 0.5 & 0.3 \\ 0.8 & 0.5 & 3 & 0.2 \\ 0.6 & 0.3 & 0.2 & 2 \end{bmatrix}$$

Intepretation of the covariance matrix

(a) **(2 points)** What do the diagonal elements of the covariance matrix represent?

The diagonal elements of the covariance matrix represent the variance of the individual features.

(b) **(2 points)** What do the off-diagonal elements signify in terms of the relationship between the features?

The off-diagonal elements signify the covariance between the features.

Random Vector and Variance

(a) **(2 points)** Calculate the total variance of the random vector X.

$$\text{Total Variance} = \text{Trace}(\Sigma_x) = 5 + 4 + 3 + 2 = \mathbf{14}$$

(b) **(2 points)** How would you compute the variance captured by a single feature (e.g, the first feature X_1)?

$$\text{Variance of } X_1 = \Sigma_{11} = \mathbf{5}$$

Eigenvalues and Eigenvectors of the Covariance Matrix

(a) **(2 points)** Calculate the eigenvalues and eigenvectors of the covariance matrix Σ_x by hand method used to calculate eigenvalues and eigenvectors is eigen decomposition:
The characteristic equation is given by:

$$\det(\Sigma_x - \lambda I) = 0$$

$$\begin{vmatrix} 5-\lambda & 1.2 & 0.8 & 0.6 \\ 1.2 & 4-\lambda & 0.5 & 0.3 \\ 0.8 & 0.5 & 3-\lambda & 0.2 \\ 0.6 & 0.3 & 0.2 & 2-\lambda \end{vmatrix} = 0$$

$$\lambda^4 - 14\lambda^3 + 68.18\lambda^2 - 139.254\lambda + 101.356 = 0$$

$$\lambda_1 = \mathbf{6.20306}, \lambda_2 = \mathbf{3.20619}, \lambda_3 = \mathbf{2.71066}, \lambda_4 = \mathbf{1.88009}$$

To calculate the eigen vectors, we substitute the eigen values into the equation:
The eigen vectors are:
for eigen value $\lambda_1 = 6.20306$

$$\begin{bmatrix} -1.20306 & 1.2 & 0.8 & 0.6 \\ 1.2 & -2.20306 & 0.5 & 0.3 \\ 0.8 & 0.5 & -3.20306 & 0.2 \\ 0.6 & 0.3 & 0.2 & -4.20306 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

The augmented matrix is:

$$\begin{bmatrix} -1.20306 & 1.2 & 0.8 & 0.6 & \mid & 0 \\ 1.2 & -2.20306 & 0.5 & 0.3 & \mid & 0 \\ 0.8 & 0.5 & -3.20306 & 0.2 & \mid & 0 \\ 0.6 & 0.3 & 0.2 & -4.20306 & \mid & 0 \end{bmatrix}$$

$$R_1 = \frac{1}{-1.20306} R_1 = \begin{bmatrix} 1 & -0.997 & -0.66497 & -0.4987 & \mid & 0 \\ 1.2 & -2.20306 & 0.5 & 0.3 & \mid & 0 \\ 0.8 & 0.5 & -3.20306 & 0.2 & \mid & 0 \\ 0.6 & 0.3 & 0.2 & -4.20306 & \mid & 0 \end{bmatrix}$$

$$R_2 = R_2 - 1.2R_1 = \begin{bmatrix} 1 & -0.997 & -0.66497 & -0.4987 & \mid & 0 \\ 0 & -1.00611 & 1.297965 & 0.89847 & \mid & 0 \\ 0.8 & 0.5 & -3.20306 & 0.2 & \mid & 0 \\ 0.6 & 0.3 & 0.2 & -4.20306 & \mid & 0 \end{bmatrix}$$

(b) **(2 points)** List the eigenvalues in descending order and explain what they represent in terms of variance

$$\lambda_1 = \mathbf{6.20306}, \lambda_2 = \mathbf{3.20619}, \lambda_3 = \mathbf{2.71066}, \lambda_4 = \mathbf{1.88009}$$

The eigen values represent the variance of the data along the principal components. The first eigen value $\lambda_1 = 6.20306$ represents the variance of the data along the first principal component, the second eigen value $\lambda_2 = 3.20619$ represents the variance of the data along the second principal component, the third eigen value $\lambda_3 = 2.71066$ represents the variance of the data along the third principal component, and the fourth eigen value $\lambda_4 = 1.88009$ represents the variance of the data along the fourth principal component.

Part 2: Principal Component Analysis (PCA) (8 points)
Now that you have a grasp of the covariance matrix and its eigenvalues, you will apply PCA to a random vector

Principal Component Directions

(a) **(2 points)** Using the eigenvectors, describe the principal component directions. What do these directions represent in terms of variance in the data?

(b) **(2 points)** Explain the concept of orthogonality in PCA and why is it important?

Orthogonality in PCA means that the principal components are perpendicular to each other. This is important because it ensures that the principal components are independent of each other. This means that the variance of the data is maximized along the principal components.

Transformation of Random Vector
Let the eigenvector matrix be P and defined the transformed random vector Y by $Y = P^T X$

(a) **(2 points)** What is the covariance matrix of the vector Y?

$$\Sigma_Y = P^T \Sigma_x P$$

(b) **(2 points)** How does this Transformation affect the correlation between the transformed features?
The transformation affects the correlation between the transformed features by making them uncorrelated. The covariance matrix of the transformed random vector Y is a diagonal matrix, which means that the transformed features are uncorrelated.c

Part 3: Performing PCA by Hand on a Simple Dataset (8 points) consider a simple dataset represented by the following 2-dimensional random vector $Y = [Y_1, Y_2]^T$:

$$Y = \begin{bmatrix} 1.2 & 2.8 \\ 0.8 & 2.4 \\ 1.6 & 3.2 \\ 1.4 & 2.9 \end{bmatrix}$$

Mean Centering

(a) **(2 points)** Calculate the mean of the dataset for each feature Y_1 and Y_2

$$\text{Mean of } Y_1 = \frac{1.2 + 0.8 + 1.6 + 1.4}{4} = \mathbf{1.25}$$

$$\text{Mean of } Y_2 = \frac{2.8 + 2.4 + 3.2 + 2.9}{4} = \mathbf{2.825}$$

(b) **(2 points)** Subtract the mean from each feature to center the data

$$\text{Centered Data} = \begin{bmatrix} 1.2 - 1.25 & 2.8 - 2.825 \\ 0.8 - 1.25 & 2.4 - 2.825 \\ 1.6 - 1.25 & 3.2 - 2.825 \\ 1.4 - 1.25 & 2.9 - 2.825 \end{bmatrix} = \begin{bmatrix} -0.05 & -0.025 \\ -0.45 & -0.425 \\ 0.35 & 0.375 \\ 0.15 & 0.075 \end{bmatrix}$$

Covariance Matrix (2 points)

(a) Calculate the covariance matrix of the centered data

$$\text{Covariance Matrix} = \frac{1}{n-1} \text{Centered Data}^T \text{Centered Data}$$

$$\text{Covariance Matrix} = \frac{1}{4-1} \begin{bmatrix} -0.05 & -0.025 \\ -0.45 & -0.425 \\ 0.35 & 0.375 \\ 0.15 & 0.075 \end{bmatrix}^T \begin{bmatrix} -0.05 & -0.025 \\ -0.45 & -0.425 \\ 0.35 & 0.375 \\ 0.15 & 0.075 \end{bmatrix}$$

$$\text{Covariance Matrix} = \frac{1}{3} \begin{bmatrix} -0.05 & -0.45 & 0.35 & 0.15 \\ -0.025 & -0.425 & 0.375 & 0.075 \end{bmatrix} \begin{bmatrix} -0.05 & -0.025 \\ -0.45 & -0.425 \\ 0.35 & 0.375 \\ 0.15 & 0.075 \end{bmatrix}$$

$$\text{Covariance Matrix} = \frac{1}{3} \begin{bmatrix} 0.35 & 0.335 \\ 0.335 & 0.328 \end{bmatrix}$$

$$\text{Covariance Matrix} = \begin{bmatrix} 0.1167 & 0.1117 \\ 0.1117 & 0.1093 \end{bmatrix}$$

Eigenvalue decomposition (2 points)

(a) Manually compute the eigenvalues and eigenvectors of the covariance matrix
The Eigen values are obtained by solving the characteristic equation:

$$\text{Characteristic Equation: } \det(\Sigma - \lambda I) = 0$$

$$\begin{vmatrix} 0.1167 - \lambda & 0.1117 \\ 0.1117 & 0.1093 - \lambda \end{vmatrix} = 0$$

$$\lambda^2 - 0.226\lambda + 0.0003 = 0$$

$$\lambda_1 = 0.225, \lambda_2 = 0.0013$$

The Eigen vectors are obtained by solving the equation:
for eigen value $\lambda_1 = 0.225$

$$\begin{bmatrix} 0.1167 - 0.225 & 0.1117 \\ 0.1117 & 0.1093 - 0.225 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} -0.1083 & 0.1117 \\ 0.1117 & -0.1157 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$R_1 = \frac{1}{-0.1083} R_1 = \left[\begin{array}{cc|c} 1 & -1.031 & 0 \\ 0.1117 & -0.1157 & 0 \end{array}\right]$$

$$R_2 = R_2 - 0.1117R_1 = \begin{bmatrix} 1 & -1.031 & | & 0 \\ 0 & -0.0005 & | & 0 \end{bmatrix}$$

$$R_2 = \frac{1}{-0.0005}R_2 = \begin{bmatrix} 1 & -1.031 & | & 0 \\ 0 & 1 & | & 0 \end{bmatrix}$$

$$v_1 = 1.031v_2$$

$$v_1 = \begin{bmatrix} 1.031 \\ 1 \end{bmatrix}$$

for eigen value $\lambda_2 = 0.0013$

$$\begin{bmatrix} 0.1167 - 0.0013 & 0.1117 \\ 0.1117 & 0.1093 - 0.0013 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 0.1154 & 0.1117 \\ 0.1117 & 0.1080 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$R_1 = \frac{1}{0.1154}R_1 = \begin{bmatrix} 1 & 0.968 & | & 0 \\ 0.1117 & 0.1080 & | & 0 \end{bmatrix}$$

$$R_2 = R_2 - 0.1117R_1 = \begin{bmatrix} 1 & 0.968 & | & 0 \\ 0 & -0.0001256 & | & 0 \end{bmatrix}$$

$$R_2 = \frac{1}{-0.0001256}R_2 = \begin{bmatrix} 1 & 0.968 & | & 0 \\ 0 & 1 & | & 0 \end{bmatrix}$$

$$v_1 = -0.968v_2$$

$$v_1 = \begin{bmatrix} -0.968 \\ 1 \end{bmatrix}$$

so the eigen vectors are as follows:

$$\lambda_1 = 0.225, v_1 = \begin{bmatrix} 1.031 \\ 1 \end{bmatrix}$$

$$\lambda_2 = 0.0013, v_2 = \begin{bmatrix} -0.968 \\ 1 \end{bmatrix}$$

Project the data (2 points)

(a) Using the principal component corresponding to the largest eigenvalue, project the original data onto the principal component axis

$$\text{Projection} = \text{Centered Data} \times v_1$$

$$\text{Projection} = \begin{bmatrix} -0.05 & -0.025 \\ -0.45 & -0.425 \\ 0.35 & 0.375 \\ 0.15 & 0.075 \end{bmatrix} \begin{bmatrix} 1.031 \\ 1 \end{bmatrix}$$

$$\text{Projection} = \begin{bmatrix} -0.05 \times 1.031 + -0.025 \times 1 \\ -0.45 \times 1.031 + -0.425 \times 1 \\ 0.35 \times 1.031 + 0.375 \times 1 \\ 0.15 \times 1.031 + 0.075 \times 1 \end{bmatrix}$$

$$\text{Projection} = \begin{bmatrix} -0.07655 \\ -0.88895 \\ 0.73585 \\ 0.22965 \end{bmatrix}$$

(b) Show the final transformed data in 1D (along the principal component axis)

$$\text{Final Transformed Data} = \begin{bmatrix} -0.07655 \\ -0.88895 \\ 0.73585 \\ 0.22965 \end{bmatrix}$$

PCA in Practice with a Large Dataset (10 points)

You are now provided with a dataset consisting of 500 users, where each user has four features: Usage time, Interactions, Activity type 1, Activity type 2. You will apply PCA using python to reduce the dimensionality

[**in the notebook**]

Intepretation & Business Insights (20 points)

Part 1: Feature Intepretation in PCA:

(a) **(5 points)** Based on the principal components directions, explain which features (original dimensions) contribute most to the first and second principal components.

(b) **(5 points)** How would you explain the reduced features to a non-technical team in terms of the user behavior patterns?

Question 2: SUM OF RANDOM VARIABLES, CENTRAL lIMIT THEOREM & PROBABILITY BOUNDS (5O marks)

Reading: The sum of random variables is a fundamental concept in probability and statistics, shedding light on the behavior of combined outcomes under uncertainty. When adding two random variables, X and Y ($i.e., Z = X + Y$), we analyze the distribution of Z. For independent variables, the distribution of Z can be derived by convolving the individual distributions.
For example, if both X and Y are normally distributed with means μ_X and μ_Y and variances σ_X^2 and σ_Y^2, then Z will also be normally distributed, with mean $\mu_Z = \mu_X + \mu_Y$ and variance $\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2$ which is beneficial for statistical modeling.
When X and Y are not independent, we must include covariance, which accounts for how the variables change together, in the variance of Z: $\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2 + 2\text{Cov}(X, Y)$ + 2Cov(X, Y).
This concept applies to any number of random variables. For independent variables $X_1, X_2, ..., X_n$, the sum $Z = X_1 + X_2 + ... + X_n$ is analyzed similarly. The Central Limit Theorem indicates that as the number of independent variables increases, their standardized sum approaches a normal distribution. Understanding the sum of random variables is essential in finance, insurance, and natural sciences for risk assessment, forecasts, and decision-making, revealing insights into complex systems. Probability bounds are crucial in statistics, quantifying the likelihood of events within specified limits, particularly in finance and engineering.
Markov's inequality estimates the probability that a non-negative random variable X exceeds a certain value a, stating that for any $a > 0$, the probability that $X \geq a$ is at most the expected value of X divided by a. This demonstrates that limited information about X's distribution can provide useful probability estimates.
Chebyshev's inequality extends Markov's by considering variance. It states that for any random variable X with mean μ and finite variance σ^2, the probability that X deviates from its mean by more than k standard deviations is at most 1 k2 . This finding is fundamental for statistical inference. Hoeffding's inequality offers bounds for sums of independent random variables, ensuring exponential decay in tail probabilities, which is especially valuable in large sample scenarios to keep observed averages close to expected values.
Overall, these probability bounds enhance decision-making and deepen our understanding of stochastic processes, facilitating robust conclusions across various fields.

Part 1: Mobile Network Data Analysis (10 points)
In a study conducted by a telecommunication company in Rwanda, mobile network Clls are classified as either voice (V) when someone is speaking or data (D) when there is a modem or fax transmission. Based on observed data, the probabilities are:

$$P(V) = 0.6 \text{ (60\% voice calls)}$$

$$P(D) = 0.4\text{(40\% data calls)}$$

Assume data calls and voice calls occur independently of each other, and let the random variable K_n represent the number of data calls in a collection of n calls.

1. **(2 points)** What is the $E[K_{100}]$, the expected number of voice calls in a set of 100 calls?

$$E[K_{100}] = n \times P[D] = 100 \times 0.4 = \mathbf{40}$$

2. **(2 points)** What is $\sigma_{k_{100}}$, the standard deviation of the number of voice calls in a set of 100 calls?

$$\sigma_{k_{100}} = \sqrt{n \times P[D] \times P[V]} = \sqrt{100 \times 0.4 \times 0.6} = \mathbf{4.9}$$

3. **(2 points)** Using Central Limit Theorem, estimate $P[K_{100} \geq 18]$, the probability of at least 18 data calls in a set of 100 calls

$$P[K_{100} \geq 18] = 1 - P[K_{100} < 18] = 1 - P[K_{100} \leq 17]$$

$$P[K_{100} \leq 17] = P\left[\frac{K_{100} - E[K_{100}]}{\sigma_{K_{100}}} \leq \frac{17 - 40}{4.9}\right] = P[Z \leq -4.69]$$

$$P[Z \leq -4.69] = 0$$

$$P[K_{100} \geq 18] = 1 - 0 = \mathbf{1}$$

4. **(2 points)** Using the CLT, estimate $P[16 \leq K_{100} \leq 24]$, the probability of between 16 and 24 data calls in a set of 100 calls

$$P[16 \leq K_{100} \leq 24] = P[K_{100} \leq 24] - P[K_{100} \leq 16]$$

$$P[K_{100} \leq 24] = P\left[\frac{K_{100} - E[K_{100}]}{\sigma_{K_{100}}} \leq \frac{24 - 40}{4.9}\right] = P[Z \leq -3.27]$$

$$P[Z \leq -3.27] = 0.0005$$

$$P[K_{100} \leq 16] = P\left[\frac{K_{100} - E[K_{100}]}{\sigma_{K_{100}}} \leq \frac{16 - 40}{4.9}\right] = P[Z \leq -4.9]$$

$$P[Z \leq -4.9] = 0$$

$$P[16 \leq K_{100} \leq 24] = 0.0005 - 0 = \mathbf{0.0005}$$

5. **(2 points)** Based on your calculations, what can you infer about the likelyhood of high data traffic during a given period? How might this information help a telecom optimize their resources for voice and data services?

The likelihood of high data traffic during a given period is very low. This information can help a telecom optimize their resources for voice and data services by ensuring that they have enough capacity to handle voice calls, which are more likely to occur. They can also allocate resources to handle data calls, but they do not need to allocate as many resources since data calls are less likely to occur.

Part 2: Chernoff Bound & Gaussian Random Variables (4 points)
Use the Chernoff bound to show that for a Gaussian (Normal) random variable X with mean μ and standard deviation σ, the probability that X exceeds a certain threshold c can be bounded by:

$$P[X \geq c] \leq e^{-\frac{(c-\mu)^2}{2\sigma^2}}$$

proof:

$$P[X \geq c] = P[e^{tX} \geq e^{tc}]$$
$$P[X \geq c] = P[e^{tX} \geq e^{tc}] \leq \frac{E[e^{tX}]}{e^{tc}}$$
$$E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$
$$E[e^{tX}] = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{tx-\frac{(x-\mu)^2}{2\sigma^2}} dx$$
$$E[e^{tX}] = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{tx-\frac{x^2-2x\mu+\mu^2}{2\sigma^2}} dx$$

but the integral is the Gaussian integral:

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{tx-\frac{x^2-2x\mu+\mu^2}{2\sigma^2}} dx = e^{\frac{t^2\sigma^2}{2}+t\mu}$$

solving for t:

$$e^{\frac{t^2\sigma^2}{2}+t\mu} = e^{tc}$$
$$\frac{t^2\sigma^2}{2} + t\mu = tc$$
$$\frac{t\sigma^2}{2} + \mu = c$$
$$t = \frac{2(c-\mu)}{\sigma^2}$$

substituting back into the Chernoff bound:

$$P[X \geq c] \leq \frac{e^{\frac{2(c-\mu)}{\sigma^2}X}}{e^{tc}}$$
$$P[X \geq c] \leq e^{-\frac{(c-\mu)^2}{2\sigma^2}}$$

Given this result, how would you use it to provide a worst-case scenario estimate in a real-world context, such as predicting an extreme event like an abnormally high network traffic spike or stock price surge?

Part 3: Soccer Tournament Performance (11 points)
Manchester United is competing in a knockout-style tournament, where each game can result in a win, loss, or tie. For every win, they earn 3 points, for every tie 1 point, and for a loss 0 points. The outcome of each game is independent of the others, and each game result is equally likely (win, loss, or tie). Let X_i be the number of points earned in game i, and Y represent the total number of points earned over the course of the tournament.

1. **(3 points)** Derive the moment generating function (MGF) of ϕ_Y(s)
random variables X_i are independent and identically distributed, so the MGF of the sum of the random variables is the product of the MGFs of the individual random variables
A random variable X_i can take on the values 0, 1, 3 with probabilities $\frac{1}{3}$ each

$$\phi_{X_i}(s) = E[e^{sX_i}] = \frac{1}{3}e^0 + \frac{1}{3}e^s + \frac{1}{3}e^{3s}$$
$$\phi_{X_i}(s) = \frac{1}{3} + \frac{1}{3}e^s + \frac{1}{3}e^{3s}$$
$$\phi_{X_i}(s) = \frac{1}{3}(1 + e^s + e^{3s})$$
$$\phi_Y(s) = \phi_{X_1}(s) \times \phi_{X_2}(s) \times \phi_{X_3}(s) \times \phi_{X_4}(s) \cdots \times \phi_{X_n}(s)$$
$$\phi_Y(s) = \left(\frac{1}{3}(1 + e^s + e^{3s})\right)^n$$

2. **(5 points)** Find E[Y] and Var[Y], the expected total points and variance.
Our moment generating function is:

$$\phi_Y(s) = \left(\frac{1}{3}(1 + e^s + e^{3s})\right)^n$$

to get the expected value, we differentiate the MGF with respect to s and evaluate at s = 0:

$$E[Y] = \phi_Y'(s) = n\left(\frac{1}{3}(1 + e^s + e^{3s})\right)^{n-1} \times \frac{d}{ds}\left(\frac{1}{3}(1 + e^s + e^{3s})\right)$$
$$E[Y] = n\left(\frac{1}{3}(1 + e^s + e^{3s})\right)^{n-1} \times \frac{1}{3}(e^s + 3e^{3s})$$

set s = 0:

$$E[Y] = n\left(\frac{1}{3}(1 + 1 + 1)\right)^{n-1} \times \frac{1}{3}(1 + 3)$$

$$E[Y] = n\left(\frac{3}{3}\right)^{n-1} \times \frac{4}{3}$$

so the expected value of Y is:

$$E[Y] = \mathbf{\frac{4n}{3}}$$

The variance is obtained by differentiating the MGF twice with respect to s and evaluating at s = 0:

$$Var[Y] = \phi_Y''(s) = n(n-1)\left(\frac{1}{3}(1+e^s+e^{3s})\right)^{n-2} \times \frac{d^2}{ds^2}\left(\frac{1}{3}(1+e^s+e^{3s})\right)$$

$$Var[Y] = n(n-1)\left(\frac{1}{3}(1+e^s+e^{3s})\right)^{n-2} \times \frac{d}{ds}\left(\frac{1}{3}(e^s+3e^{3s})\right)$$

$$Var[Y] = n(n-1)\left(\frac{1}{3}(1+e^s+e^{3s})\right)^{n-2} \times \frac{1}{3}(e^s+9e^{3s})$$

set s = 0:

$$Var[Y] = n(n-1)\left(\frac{1}{3}(1+1+1)\right)^{n-2} \times \frac{1}{3}(1+9)$$

$$Var[Y] = n(n-1)\left(\frac{3}{3}\right)^{n-2} \times \frac{10}{3}$$

so the variance of Y is:

$$Var[Y] = \mathbf{\frac{10n(n-1)}{9}}$$

3. **(3 points)** Based on your calculations, what can you infer about the likely performance of Manchester United over the course of multiple tournaments? How might the expected points impact their overall ranking or their chances of advancing in the competition?

   Based on the calculations, we can infer that the expected performance of Manchester United over the course of multiple tournaments is to earn an average of $\frac{4n}{3}$ points with a variance of $\frac{10n(n-1)}{9}$. This information can help predict their overall ranking in the competition and their chances of advancing. Teams with higher expected points are more likely to rank higher and advance in the competition, while teams with lower expected points are less likely to do so.

Part 4: Course Enrollment and Resource Planning (6 points)

The number of students enrolling in a popular data science course is modelled as a poisson random variable with a mean of 100 students. The Professor has decided that if 120 students enroll, he will split the class into two sectioons, otherwise , he will teach all the students in a single Section

1. **(3 points)** What is the probability that the professor will need to teach two sections?
   using Markov's inequality:

$$P[X \geq 120] \leq \frac{E[X]}{120}$$

$$P[X \geq 120] \leq \frac{100}{120}$$

$$P[X \geq 120] \leq 0.833$$

$$P[X \geq 120] = \mathbf{0.833}$$

2. **(3 points)** Based on Probability, what recommendations would you make regarding resource planning for future courses? Should the professor prepare for two sections or allocate resources differently based on expected enrollments?

   Based on the probability that the professor will need to teach two sections, it is recommended that the professor prepare for two sections. Since the probability is 0.833, which is greater than 0.5, it is more likely that 120 students will enroll. By preparing for two sections, the professor can ensure that there are enough resources to accommodate the students and provide a better learning experience.

Part 5: Comparison of Markov, Chebyshev and Chernoff Inequalities (19 points)