STU22004 Applied Probability I

Group assignment

See instructions at the end of this document.

Question 1 (30 marks)

Consider a scenario where the temperature $X(t)$ varies randomly over a continuous time interval t, where t is in the range from 0 to 1. We begin with the assumption that $X(0) = 0$, which means that the temperature at time 0 is 0. Now, if we choose a small time increment represented by Δt, we can make the assumption that the change in temperature from time t to $t + \Delta t$, denoted as $X(t + \Delta t) - X(t)$, follows a normal distribution. This normal distribution is characterized by a mean of 0 and a variance of Δt.

1. Let P be the random variable denoting the proportion of time in $[0, 1]$ such that the temperature is positive. Estimate the distribution of P by Monte Carlo simulation and experimenting with various values of Δt (e.g. $\Delta t = 0.01, 0.001, 0.0001, \cdots$.)

2. Let $T_{\max}$ be the random variable denoting the time in $[0, 1]$ such that the temperature is at its maximum. Estimate the distribution of $T_{\max}$ by Monte Carlo simulation and experimenting with various values of Δt (e.g. $\Delta t = 0.01, 0.001, 0.0001, \cdots$.)

Question 2 (70 marks)

Create a probabilistic model and perform Monte Carlo simulations to forecast the final points for Premier League teams in the 2024-2025 season. This model will include some unknown parameters that you will determine based on the data you gather. In Premier League matches, teams earn three points for a win, one point for a draw, and no points for a loss. For your predictions, you could use statistics from the beginning of the season up to a specific date to estimate the parameters of your model, and then run Monte Carlo simulations to project the outcomes of the remaining matches, ultimately predicting the final points for each team at the season's end.

Here's a brief outline of a relatively straightforward way to model this scenario. For each match, you can treat the number of shots attempted by the home and away teams as random variables, such as Poisson random variables. The rate parameters of the Poisson distribution will be influenced by the strengths of both teams. Each shot taken will have an associated probability of scoring, which also varies depending on the teams involved.

You are encouraged to create your own models, but it's essential to explain and justify your choices. Discuss how you determine the parameters and outline the advantages and limitations of the models you select.

You can collect data from various online sources, such as this example: `https://www.premierleague.com/stats/top/clubs/total_scoring_att`.

Instructions

Submission

Submission date: **Monday 2nd of December 2024**. Late submissions will not be accepted. You can submit your project by uploading to Blackboard.

Assessment

The project will carry 20% of the course marks.

Project and report

To complete this project you can any software or resource that you think might help. The report should comprise no more than 10 pages[1], including tables, plots and diagrams, a good use of which is encouraged and will be rewarded, as long as they provide useful information. Details on the computations should be provided concisely. The report should be prepared **professionally** using any appropriate software (e.g. Word or Latex).

Teamwork

A group project is proposed mainly because group learning can be valuable, but also because team-working is a useful skill. **All members of each team must make significant contribution** to the group project. Clearly the project work does naturally break into sections which can be conducted in parallel before being brought together. The team will receive a mark and all members will receive by default the same mark.

Plagiarism and sources

While learning from each other and from other groups is encouraged, plagiarism is remarkably easy to spot, especially with online resources. This includes projects that were submitted for the same course in the last years. The same work submitted by two teams will receive a single mark which will be then divided between the teams involved, following discussion. Sources must be cited.

[1] this page limit is only indicative but reports that are too long will be penalised