ST2004 & ST2352 Assessment 2: League

Submission date **strictly before 3pm** Friday 22nd January. Late submissions are not accepted without adequate supporting approval from your tutor.

Submit your work during office hours to the School of Computer Science and Statistics Reception Office on the first floor of the O'Reilly Building. There should be one submission per group and it should clearly state the names of the individuals in the group. **It must be stapled or otherwise bound.** At the same time, submit the peer-assessment form.

This is both a learning and an assessment exercise. The exam may draw on this project. It will be scaled to carry 10% of the course marks. Good students will take the opportunity to go well beyond the lectures. 75% of the marks will be awarded 'for doing a competent job'; an extra 25% will be awarded for flair, imagination, thoroughness – *i.e.* the 'extra mile'.

For example, what can you learn from the internet, or the text, about the theory behind the comparison of simulation results with theory in such a way as to assert that they are "consistent"? In Q2 the objective is for you to learn some more probability and tell me about it, and tell me why it is interesting.

Remember: questions such as these are primarily devices to give you the opportunity to tell me that you have imagination, enthusiasm and attention to detail and that you have learnt some probability theory. Use your writing skills to tell me how you are interlacing theory with experiment.

Q1 (13 marks) The site http://understandinguncertainty.org/node/45 contains several discussions and animations of football – and other – leagues. One aspect discussed is: How much is due to chance?

The file Assessment2Resource available from the course website https://www.scss.tcd.ie/Brett.Houlding/Index/ST2004_ST2352.html contains a very simple Monte Carlo simulation of a league with just four teams. Each team plays the other twice: at home and away. In the simplest case the probability of a draw is zero; the teams have equal skills and the odds of winning are different when playing at home or away. The teams may have different skills, which can be specified; these control the odds of winning. You are invited to experiment with this, to extend the number of replications, to generalise it to a larger number of teams and to study the relationship between the variability in the skills and the variability in the points scored.

(i) For four teams with equal skill:

a. When there is no home advantage, the variation in the random number N of matches won by each team follows the Binomial distribution B(6,0.5). But when there is a home advantage N does not follow a Binomial distribution; on the contrary it can be thought of as that of $N = N_{home}+N_{away}$ where each of these follows different Binomial distributions. Discuss using probability and simulations. (2 marks)

b. In either case the random number of wins by two different teams are *not* independent. Discuss using correlation and scatter plots. (2 marks)

(ii) Generalise the simulation to four teams of *unequal* skill.

a. Validate your simulation empirically by confirming experimentally that the results of simulations vary as you would expect if one or more teams have very much better/worse skills than the others. (2 marks)

b. When all but team A have equal skills compute (by enumerating all the possibilities or otherwise) the probability distribution for the number N_A of matches won by team A. Confirm your answer by contrasting with the results of simulations. (2 marks)

(iii) Generalise to 10 teams of possibly different skills. It is argued that if the variation in the skills of the different teams is large, then the variation across the teams in the number of matches will also be large; and vice versa when the variation is small. One measure of variation of both is the standard deviation.

a. Investigate this argument experimentally, by conducting an extensive set of simulations.

b. Summarise these simulations graphically and explain their implications for the argument. (7 marks)

The report on Q1 must comprise of no more than 6 or 7 pages, including tables and diagrams; good use of diagrams is encouraged and will be rewarded. The use of a large number of significant figures in reported computations should be avoided; it suggests naiveté. The report should be prepared professionally using any appropriate software (*e.g.* Word or Latex), with the tables and diagrams being imported from EXCEL or equivalent. Word's Insert Symbol and Insert Equations facilities are adequate for the necessary Maths that you need.

Q2 (7 marks) Choose ONE of the following topics for detailed discussion.

(i) Extend your simulations of the above to explore other or deeper aspects of chance variation in the discussion of league tables. For example you might wish to include Draws in the simulations.

(ii) The Understanding Uncertainty site discusses chance variation in many situations. Choose one of these and build a simple simulation to illustrate the issues.

(iii) Benford's Distribution – as discussed in the course text and, for example, Wikipedia, supposedly applies to the first digits in any set of numbers that arise naturally and exhibit great ('order of magnitude') variation. Use simulation of random 'data' to illustrate the variation captured by Benford's distribution. Find one Irish data set that is roughly consistent with Benford's distribution, and one that is not.

(iv) The course text discusses the 'Kelly' strategy for betting; there is an entry in Wikpedia under Kelly Criterion. Build a simulation to explain the strategy and to illustrate its claimed advantages.

(v) The course text discusses the use of the so-called geometric Brownian motion model for variation in stock-prices. One application for this is the Black- Scholes formula that underlies theory of option-pricing (derivatives) in finance. Use simulation to illustrate the ideas.

Team composition

Groups of size 3 or 4 under self-allocation. If there is any problem in this regard you should contact me as soon as possible.

Team working

Your report is a group project. I propose this mainly because group learning can be valuable, but also because team-working is a useful skill. The team will receive a mark and by default all members will receive the same mark. The project work does naturally break into sections which can be conducted in parallel before being brought together. However, as problems do arise from time to time, I must insist that each team member individually completes the PEER ASSESSMENT form. This will be used as necessary to re-allocate the team mark to the individual team members. If I do not receive such a completed form from you I will take it that you made no contribution and expect no mark. While learning from each other and from other groups is encouraged, plagiarism is remarkably easy to spot, especially with online resources. The same work submitted by two teams will receive a single mark; this will be divided between the teams involved, following discussion.