How to Win Football Prediction Leagues

By Michael Gibson
mikegibson2010@gmail.com

Background

- I don’t know much about football.
- I don’t have a huge interest in football.
- This is just as well, given that I support Newcastle United.

- I do know a little bit about Maths.
- I do know a little bit about computers and programming.
- This is just as well, given that I work as a teacher of Maths and Computing.

A Bit More Background

- Between around 2010 and 2012 I had the opportunity to take part in a couple of “football prediction leagues”.
- The first one was run by a colleague of mine.
- Another was run by my wife’s friend’s dad’s former colleague.
- Since I know very little about football, I decided to take a mathematical/computational approach to this challenge.
- This proved to be very successful!

A Typical Prediction League

- Participants pay £5 per month to take part.
- Each week, participants submit their predictions for the scorelines for a selection of 10 matches.
- Points are accumulated according to the following system:

Correct scoreline (draw)	**12 points**
Correct scoreline (non-draw)	**10 points**
Correct goal difference (draw)	**5 points**
Correct goal difference (non-draw)	**3 points**
Correct outcome	**1 point**

- Small prizes of around £40 are awarded each week.
- A large prize of around £300 and second prize of around £100 are awarded at the end of the season.

Expectation Values

- Manchester Rovers are playing Liverpool United next weekend.
- I predict that the score will be 3-1 to Rovers.
- How many points do I expect to get?

10 x Probability of "3-1"

\+ $3 \times$ P("2-0") + $3 \times$ P("4-2") + $3 \times$ P("5-3") + $\cdots$

\+ $1 \times$ P("1-0") + $1 \times$ P("3-0") + $\cdots$

- If I do this calculation for all possible predictions, I can pick the one I expect to give me the most points.

Probabilities

- But what are the probabilities for each of the possible scorelines?
- Surely this requires some knowledge football?
- Not if you consult someone who *really* knows…

Probabilities

- Bookmakers quote odds for individual scorelines.
- These are probably the closest thing you can get to actual probabilities ahead of a sporting event.
- Obviously bookmakers “tighten” their odds to make a profit.
- I assumed that they would make an equal profit margin on all bets. This allowed me to infer working “probabilities”.

Practicalities

- Doing all of this manually proved to be somewhat time-consuming! Here are some ways of speeding this up…
- Microsoft Excel allows imports of data from web pages.
- It also has a built-in programming language (VBA) that allows quite straightforward processing of the data after importing.
- This speeds up the process and amounts to a “black box” – import the odd from the bookies, press “go”, output a list of recommended predictions.

Outcomes

	A	B	C	D	E	F	G	H	I	J	K	L	M
4		**LILLEY & GILLIE PREDICTION LEAGUE**											
5		**FINAL RESULT**											
6	POS	TEAM	TOTAL	EXACT									
7	1	THE UNKNOWN UNKNOWNS	1532	58	1194	15	53	45	39	43	50	22	41
8	2	X	1528	59	1184	16	43	42	47	32	44	37	55
9	3	OLYMPIC HEROES	1441	47	1127	19	27	41	60	43	44	25	21
10	4	MIGHTY MAGS	1438	47	1146	8	40	41	46	22	36	36	42
11	5	WATTS WONG	1436	54	1162	8	47	21	33	26	63	18	27
12	6	BANANARAMA	1426	52	1170	12	39	29	33	24	35	24	35
13	7	PARTY WITH MARTY	1421	55	1116	18	23	30	43	44	48	16	50
14	8	SPIDERS FROM MARS	1391	52	1089	30	30	23	60	14	67	38	16

- 2nd place overall in the first season, prize ~£100.
- 1st place overall in the second season, prize ~£300
- After that the company split in two and the league no longer operates.
- Awkward moments at parties when my wife's friend's dad tries to engage me in intelligent conversation about football!

Some interesting maths…

- Optimal predictions tend to be “low-scoring”
- For examples, the total number of goals “predicted” across the 10 matches is usually somewhat less than the actual total.
- In the league at my own workplace, my (non-mathematician) colleague thought that my system was simply to give my predictions in “binary”!
- At the root of this is a very common statistical phenomenon…

Simple Demonstration…

- Suppose I roll an ordinary dice until I get a 6.
- You need to predict how many times I will roll the dice.
- The mean number of times will be 6.
- The median number of times will be 4.
- The modal number of times will be…

1!

- This is the most sensible prediction.

Back to Football

- The situation with football predictions is a little more complicated but the underlying principle is the same.
- My hunch is that people who know lots about football but less about probability will tend to make predictions that are more "plausible" as opposed to more "probable".
- In the 2010 World Cup the most likely scoreline, according to bookmakers, for Brazil vs North Korea was 3-0. Most people's gut reaction was "Brazil will win by more goals than that!" – but of course "more than 3-0" is not a single option!
- Does anyone remember what the actual scoreline was?

BRAZIL 2-1 NORTH KOREA

Well that's football for you!

Finally: What troubles me…

- Since doing this I have sometimes asked myself…
- If I had simply looked up the scoreline with the shortest bookmaker odds and used that for all my predictions, would it have made a difference…
- I know it would sometimes have given different predictions, but might I have still won?
- After all the, expectation values for the top few options differed by very little.
- I have all the data I need should I choose to find this out. I *really* would prefer not to know!
- But my advice would be, if you're in a football prediction league, unless you have a lot of time on your hands, just go for the scoreline with the shortest bookmaker odds – you'll save a lot of time and maybe make a lot of money!