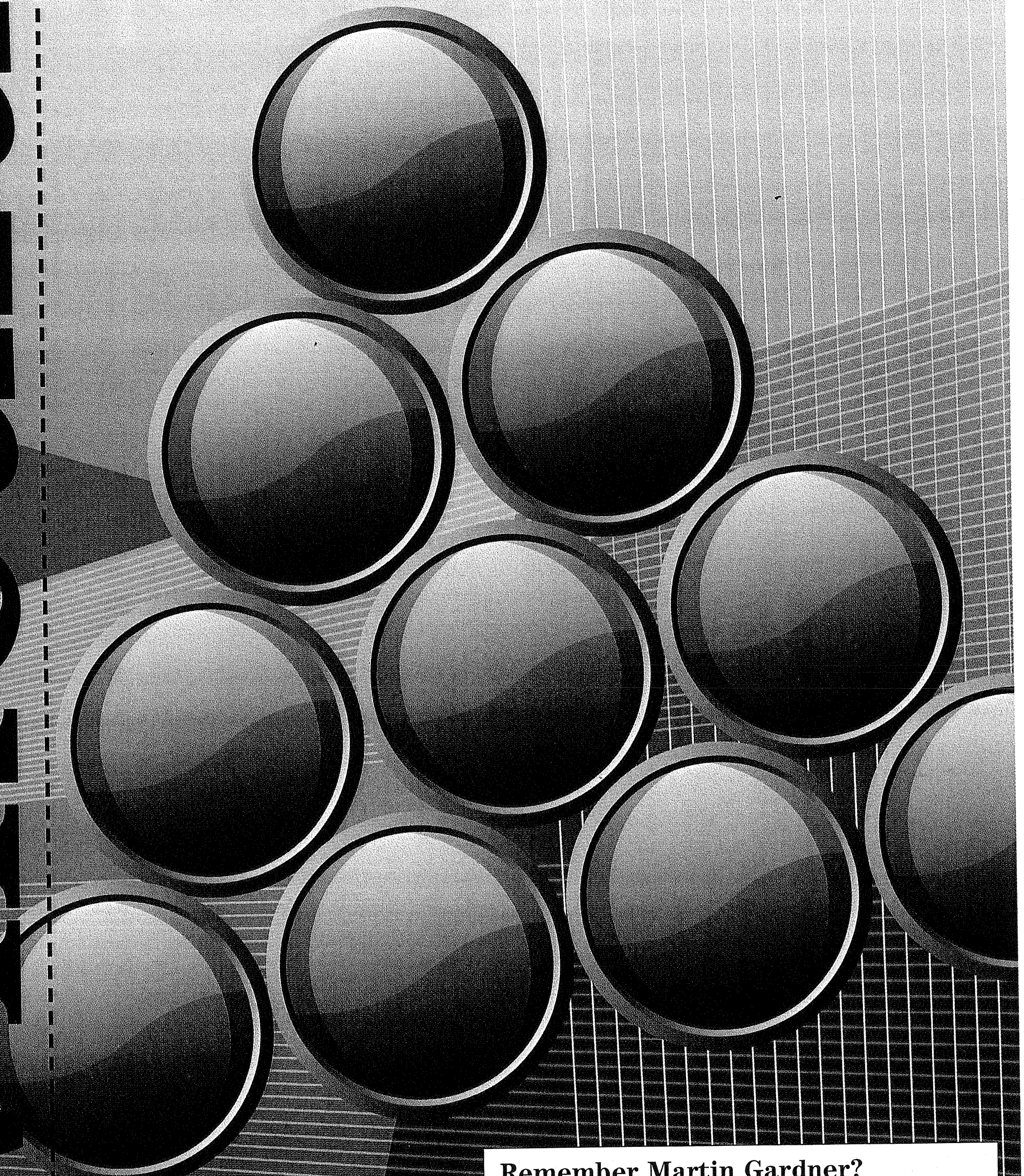
vinculum

Volume 48 Term 2 2011

THE MATHEMATICAL
ASSOCIATION OF VICTORIA



Remember Martin Gardner?
Five Interesting Puzzles

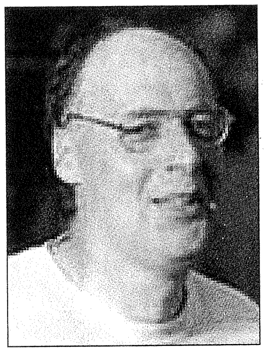
Your Secret Footy Tipping Weapon

My Web Site: CensusAtSchool

Mathematical Modelling

ISSN 0157 - 759X Registered by Australia Post - Publication No. VGB 0254

Probabilistic and Gaussian Football Tipping



David Dowe —
David is an Associate Professor in the IT faculty in Computer Science at Monash University. His research focuses on machine learning, statistics, philosophy and other areas of the information-theoretic Minimum Message Length principle.

> Right/wrong is a very obvious and sensible criterion used in many life situations. However, there is also a time and a place for knowing how sure we are of things, in terms of probabilities.

The Probabilistic Football Tipping Competition

Most football tipping competitions reward people for correctly tipping the winning team in AFL (or whatever code) football matches. Right/wrong is a very obvious and sensible criterion used by multiple choice tests and applicable to many life situations.

However, there is also a time and a place for knowing how sure we are of things. It may be more relevant to tell a person diagnosed with cancer the probability of his continued survival, and how this probability may change with various treatments. Probabilities are also very relevant to bookmakers, insurers and many other occupations.

So, why not create a football tipping competition where, instead of tipping the winner of each match, you exercised your skill in estimating the probability of each team of winning? The competition would need to be designed so that the highest expected (long term) score can be achieved by tipping the *true* probability, even though this is never known. It can be shown that this is achieved by giving a reward of $\log p$ to a tipper who has assigned a probability p to a team which wins, and $\log(1-p)$ as a reward for picking the losing team. Proving this could be an interesting exercise. We use logarithms to the base 2, so tippers in our competition get a score in bits.

Note that if $p > ½$, then $p > 1-p$, and $\log_2(p) > \log_2(1-p)$, so the tipper is better off selecting the winning team, and gains (and loses) more the higher the probability assigned to the selected team. We have assigned a reward of $1 + \log_2(p)$ if the team wins and $1 + \log_2(1-p)$ for a loss.

Now there will be games that you think are really, really close ones. In your typical tipping competition, you've got to pick a team and you might not like that because you might feel it's almost a coin toss and you're ripped off by having to choose one team. A probabilistic competition enables you to say, "Oh, I'm only 50% sure, or 51% sure," and you enter that. If the tipper chooses a probability of 50%, then $1 + \log_2(p) = 1 + \log_2(1-p) = 0$. This means that they are certain to score zero points, whatever the outcome of the match.

If there is a draw, the average of the winning and losing rewards are given.

Reward $= (1 + \log_2(p) + 1 + \log_2(1-p)) \div 2 = 1 + \frac{\log_2 p(1-p)}{2}$

This is an upside-down "U" shaped function, which is negative for every value of p except for $p = 0.5$. If you think a draw will occur you should tip 0.5. If correct, your reward is to lose no bits; everyone who tipped otherwise will lose bits.

For matches where one team is a lot stronger than the other, more bits are gained by allocating a high probability. However, be warned. The scoring is **not symmetrical** and can be very non-intuitive for the beginner. Choosing a value close to 1 is *very* risky as $\log_2(1-p)$ is extremely small when p is close to one. For example, a loss using a probability of 0.95 will score you $1 + \log_2(1-p) = 1 + \log_2(0.05) \approx 1 - 4.32 = -3.32$ bits compared to a win which would earn $1 + \log_2(0.95) \approx 0.93$ bits. However, if the real probability is 0.95, this is your best option in the long term. For this reason, only probabilities between 0.001 and 0.999 are accepted, so that people won't get negative scores from which they could not recover.

p	Score if win	Score if lost	Score if draw
1.00	+1.000	-Infinity	-Infinity
0.95	+0.926	-3.322	-1.198
0.90	+0.848	-2.322	-0.737
0.85	+0.766	-1.737	-0.486
0.80	+0.678	-1.322	-0.322
0.75	+0.585	-1.000	-0.208
0.70	+0.485	-0.737	-0.126
0.65	+0.379	-0.515	-0.068
0.60	+0.263	-0.322	-0.029
0.55	+0.138	-0.152	-0.007
0.50	+0.000	+0.000	+0.000

Table 1. Effect of p *on scoring.*

Since the competition has been running, we have also developed mathematical measures for boldness, calibration and trust. You can find the mathematics behind this by logging on to

Why not create a football tipping competition where, instead of tipping the winner of each match, you exercised your skill in estimating the probability of each team of winning?

www.csse.monash.edu.au/~footy/about.shtml and following the appropriate links.

The probabilistic footy tipping competition is a fun way to engage young people at all levels of the curriculum in mathematical thinking. To the best of our knowledge, we have achieved a world first in producing the first real-world tipping competition which is based not on right/wrong but on probabilities and information-theoretic scores. You can join the competition on **www.csse.monash.edu.au/~footy/index.shtml.** It is no great disadvantage to join the probabilistic competition late, as you begin with a score of zero and many (over confident) tippers already have a negative score after the first few rounds. Unfortunately we can offer no prizes, just the honour and glory of winning the world's only football tipping competition of this kind.

You may wish to use dominance matrices to predict probabilities, or find the bookmakers' odds. These of course can be converted to probabilities, but it should be remembered that bookmakers' odds reflect the relative amounts of money wagered and not necessarily the true probabilities.

The Gaussian Football Tipping Competition

This could be called the probability distribution competition. In this case, the tipper must specify a winning margin, and a standard deviation.

The tipper is rewarded for getting close to the actual winning margin. We can begin by using the tipped winning margin and the selected standard deviation to generate a Normal or Gaussian probability distribution.

$$p_m = \frac{1}{\sigma\sqrt{2\pi}} \int_a^b e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

μ, the mean, is the tipped winning margin and σ is the tipped standard deviation. Suppose m is the actual winning margin. The reward will be related to the probability, using this normal distribution, of getting the actual winning margin m. As the normal (Gaussian) distribution is the well known bell shaped curve, the more standard deviations this is away from the mean, the lower the tipper's score. Since the distribution is continuous and m must be an integer, we must find the area under the curve from $x = m - 0.5$ to $x = m + 0.5$, as below.

$$p_m = \frac{1}{\sigma\sqrt{2\pi}} \int_{m-0.5}^{m+0.5} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

To find the probability for all a tipper's selections over the entire competition, we must multiply the probabilities of these independent events. However, in tipping competitions, scores are generally added. We have therefore calculated the logarithm of this probability, again using a base of 2. Now adding the logarithms relates to calculating a probability for the complete set of predictions. There is the added advantage of making the scores a reasonable size, as the given probabilities will all be very small. As the logarithms will all be negative, we have added 10 to the scores, so everyone will usually get a positive score, unless they are really way off. The formula for the game score, in bits, is

$$10 + \log_2 \frac{1}{\sigma\sqrt{2\pi}} \int_{m-0.5}^{m+0.5} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

where m is the actual winning margin, μ and σ are the tipped mean and standard deviation respectively. We use the Z score, the number of standard deviations between the tipped margin and the actual margin.

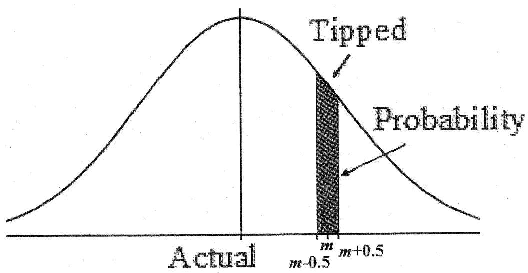


Figure 1. The probability is the grey strip.

In this competition the accuracy of prediction of winning margin is important. The standard deviation proves to be difficult; the larger you make it, the more spread out and flatter the probability distribution function will be and the smaller the probability of your score, and hence

It can be shown that the highest expected (long term) score can be achieved by tipping the *true* probability, even though this is never known.

Since the competition has been running, we have also developed mathematical measures for boldness, calibration and trust.

the smaller your actual score. Too small a standard deviation means your prediction is more likely to be three or more standard deviations from the mean, and therefore extremely small; but if you 'nail it' with a low standard deviation, you will get a high score. Experience has shown that a value of sigma around 40 is a good place to start. It is also possible to join this competition after the start without being too disadvantaged.

I established the probabilistic and Gaussian competitions in 1995 and 1996 respectively, with the probabilistic competition being a spin-off from joint research which led me to suggest the Gaussian competition the following year in 1996. I have proved uniqueness results about the logarithmic scoring methods for these competitions. The probabilistic competition began in Round 3 of the 1995 AFL season and the Gaussian competition began with the 1996 AFL season. These are world record longest-running compression-based competitions. The scoring using logarithms of probabilities amounts to the same field, namely information theory, used to compress computer files.

We also run a normal footy tipping competition in the usual way where you select winners and winning margins. You can get bonus points if your winning margin is sufficiently close to the actual winning margin. This applies even if you pick a losing team.

So we hope you and your students will enjoy our unique football tipping competitions, and appreciate the mathematics behind them.

It is no great disadvantage to join the probabilistic competition late, as you begin with a score of zero and many (over confident) tippers already have a negative score after the first few rounds.

Further reading

Dowe, D. L. (2011). MML, hybrid Bayesian network graphical models, statistical consistency, invariance and uniqueness. In Bandyopadhyay, P.S. & Forster, M.R. (Eds.), *Handbook of the Philosophy of Science* (Vol. 7). Elsevier, *Philosophy of Statistics*. (Section 3.5, pp. 901–982)