A Bayesian Analysis of Racing Tipsters

Paul Parsons

Anyone who's tried to follow the recommendations of a good tipster, especially one whose bets are posted for free online, will know the problem with this strategy. Unless you have nerves of the proverbial steel and fingers faster than Yehudi Menuhin after fifteen espressos, then by the time you get to put your money on, the bookmakers have already shortened their prices.

They know, like every savvy punter, that when it comes to making money from betting, price is paramount. Shorten up by a few ticks and a bettor's long-term winnings won't cover their losses – putting the bookies back in the black. This is the all-important concept of 'value' – a successful gambler needs to get on at a price that's more generous than it should be, given the probability of the bet copping. That's why all good tipsters put out a recommended price alongside each of their bets, and the received wisdom is woe betide anyone who dares get involved for less.

But is this strictly true? After all, the handful of bets per day (if that) offered up by most tipsters are the cream of the day's crop, so you'd expect them to hold more than a few percent of value at the advertised price. Meanwhile, the bookies can only shorten the price on a horse by a limited amount before they begin to appear uncompetitive, and so have to lengthen other selections. So the question is: can you still make money betting a tipster's selections at shorter than the recommended price? And, if so, how short?

If your favourite tipster publishes their historical bets in an easily downloadable form (and if they're any good, there's no reason they wouldn't), then a little data analysis can reveal all. The first thing to establish is a framework for relating the decimal price, D, to the true probability, p, of a bet winning. A natural choice of model is a power law, i.e.:

$$p = \left(\frac{1}{D}\right)^{\gamma};\ \gamma \geq 0, D > 0. \tag{1}$$

This ensures $0 \leq p \leq 1$ for all permissible D, as required since p is a probability. Note that a linear model such as $p = \alpha/D$, which might appear simpler, isn't valid because it can't confine p to the required range for all values of D.

The parameter γ in equation (1) quantifies the ability of the tipster. Because the value of a bet can be calculated as $pD - 1$ (and we need this to be positive if we're going to win money long-term), then the tipsters we should follow are those for which $\gamma < 1$ (while those for whom $\gamma > 1$ will lose us money, and those with $\gamma \approx 1$ will merely break even).

There are many ways to determine γ from a tipster's history but I like to use a Bayesian treatment. In this approach, our uncertainty about the value of γ manifests itself as a probability distribution, which we seek to update using the tipster's historical recommendations and results. With each bet the evidence accumulates, the variance in the distribution decreases, and the analysis converges on an increasingly more precise estimate of γ. The key formula for performing the update is Bayes' theorem, namely:

$$p(\gamma|E) \propto p(E|\gamma) \times p(\gamma). \tag{2}$$

Here, E is the evidence embodied in the data, and $p(\gamma|E)$ is a conditional probability which represents the probability distribution for γ after taking account of (conditional on) the evidential weight of the data (also known as the 'posterior distribution'). On the right-hand side, $p(E|\gamma)$ is the conditional probability of obtaining the dataset given a particular value of γ (also known as the 'likelihood function'), and $p(\gamma)$ is our initial belief about the probability distribution for γ before taking account of the tipster's results (also known as the 'prior distribution', because it encapsulates our beliefs about γ prior to taking account of the data).

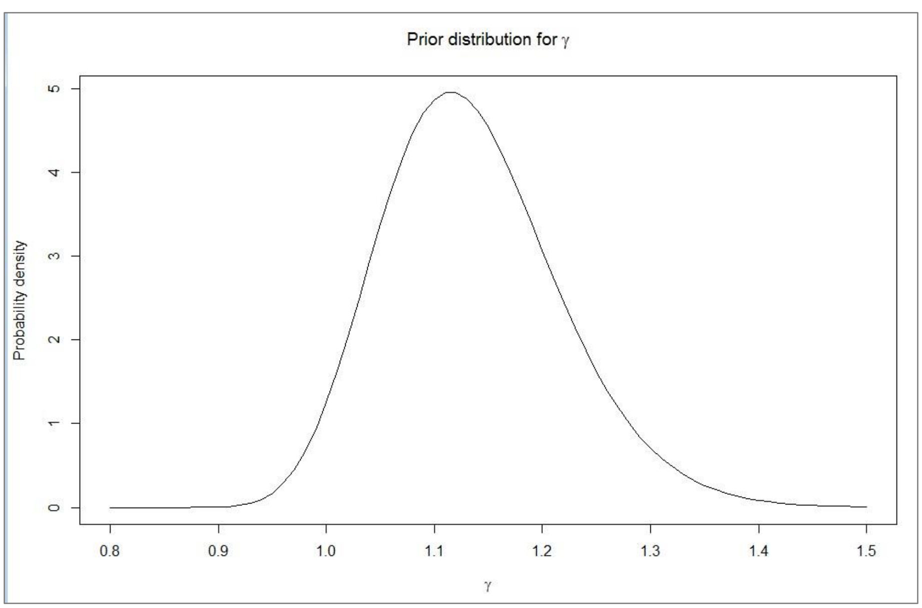
Let the dataset E consist of n bets each at price, D_i, for bet i, and with outcome w_i ($= 1$ for a win; $= 0$ for a loss), where $i = 1 \ldots n$. Each bet is a Bernoulli trial with a win probability inferred from the price by equation (1). However, because the bets are not all at the same price, we can't simply apply a binomial form for the likelihood function. Given that all the bets in the tipster's recommendations are straight singles then it's fair to assume independence. In this case, the likelihood function (the probability of obtaining all the results in the dataset E given a particular value of γ) is given by the product of the probabilities of obtaining each result, namely:

$$p(E|\gamma) = \prod_{i=1}^{n}\left\{w_i\left(\frac{1}{D_i}\right)^{\gamma} + (1-w_i)\left[1-\left(\frac{1}{D_i}\right)^{\gamma}\right]\right\}. \quad (3)$$

The choice of the prior distribution, $p(\gamma)$, is often cited as a weakness of Bayesian inference – for the very reason that the choice itself is subjective. However, in this case this freedom can work in our favour by allowing us to pick a form for the prior that rules out obviously implausible values of γ (simply by setting the probability for these values to zero). For example, we know (Leicester City notwithstanding) that bookmakers can't be overestimating their odds by *that* much – and this rules out extremely low values of γ. There isn't a tipster in the world who consistently finds 100/1 shots that come in 95% of the time – which is what an extremely small γ (=0.01) would imply.

I adopt an empirical form for the prior distribution that looks similar to that shown in Figure 1.

Figure 1



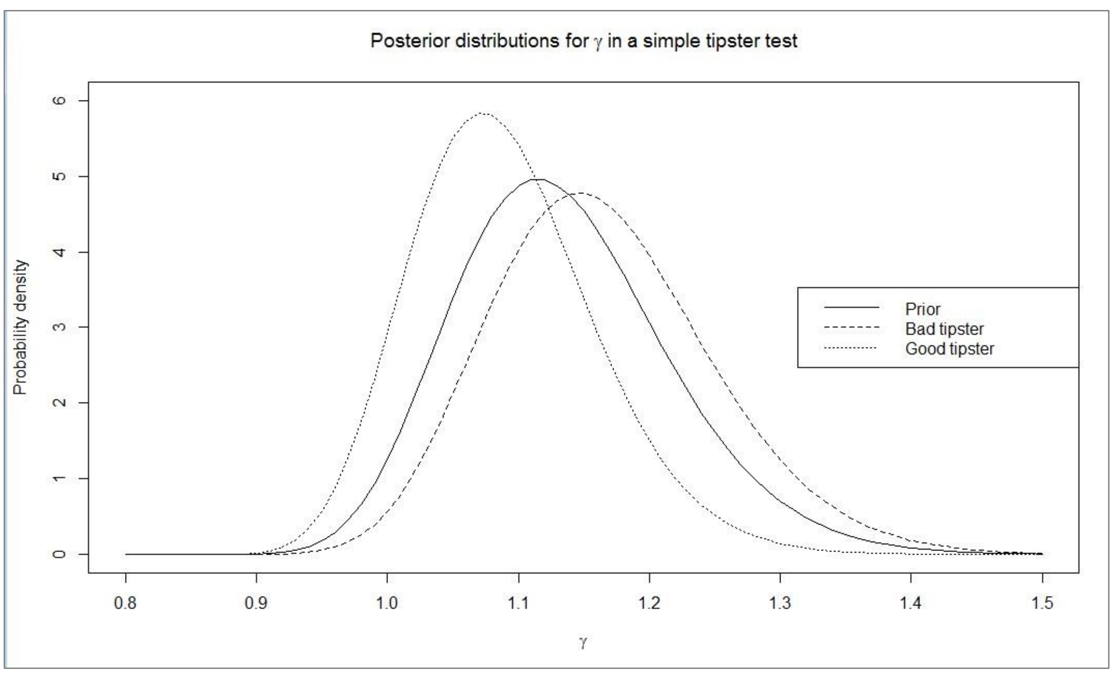
This density function has a lower bound at $\gamma = 0.8$. It has 96% of its probability mass in the range $\gamma > 1$, and a single peak located at $\gamma = 1.12$ – which together embody our principle expectation, namely that most tipsters are rubbish. The distribution is skewed, with a longer tail to the right, reinforcing our view on the quality of the typical tipster, and also allowing for the possibility of bookmakers heavily over-rounding their markets so that prices are stacked against the bettor.

This function has been created by dividing the γ (horizontal) axis into bins that are each 0.01 units wide, manually assigning a value to each bin, and then scaling these values en masse to ensure that the numerical integral of the resulting function totals to 1.

Computing $p(\gamma|E)$ requires integration of the right-hand side of equation (2) in order to normalize the resulting probability distribution. Given a numerically defined prior distribution and the form for the likelihood specified in equation (3), it's not possible to do this analytically. However, it's a fairly simple matter to apply a basic numerical integration technique such as the trapezium rule or Simpson's method across each of the bins (of width 0.01) for γ.

As a simple test, consider a series of 30 recommended bets, all priced at evens. Now consider two scenarios – one where 10 of the bets win, leading to a net loss, and another where 20 of the bets win, producing an overall gain. The form of the posterior distribution in both these cases is plotted in Figure 2 below. The losing case leads to a posterior distribution for γ that's shifted, relative to the prior, to the right slightly (pushing the peak of the distribution out from 1.12 to 1.15) - remember, the bigger γ is, the worse the quality of the tipster. Meanwhile the winning case produces the opposite effect, shifting the distribution to the left (pulling the peak in to 1.07).

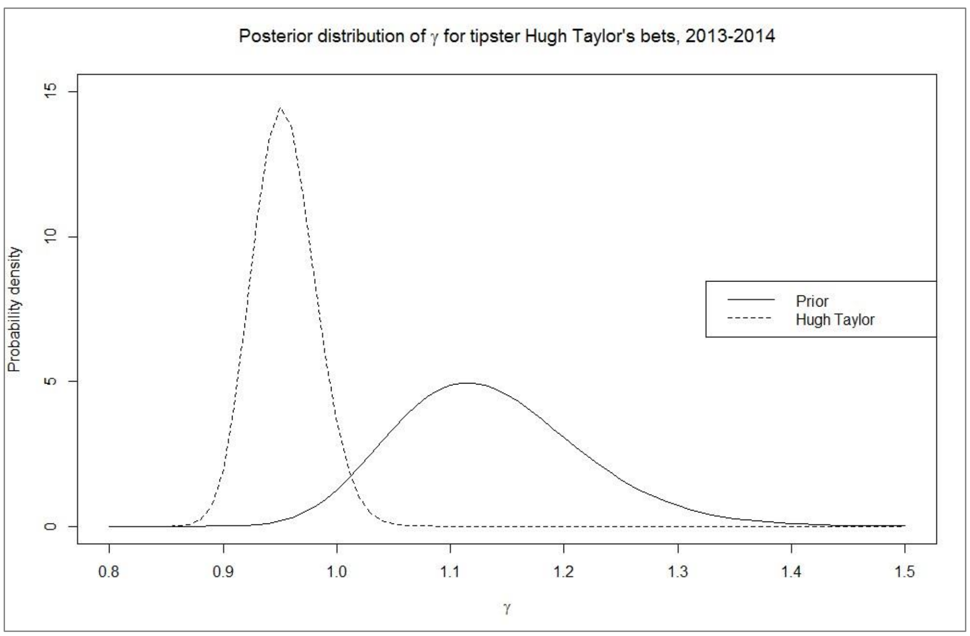
Figure 2



But what does this approach yield when given some real data? One tipster followed by many is Hugh Taylor, of horse racing website *At The Races*. His bets typically yield annual returns of around 30%, which is why most bookmakers shorten their prices on his selections often within seconds of them appearing online. So is it possible to make money taking a shorter price than Hugh recommends?

I looked at two years' worth of Hugh Taylor's tips, covering 2013 and 2014. I removed all the each way and multiple bets (he posts the occasional double), since our treatment here is not yet able to accommodate these. Most of his tips are straight singles so this isn't a major issue. In this data set which we use to fit our model (the fitting data set), Hugh recommended a total of 1026 bets, chalking up 176 wins (a strike rate of 17.15%) at a mean decimal price of around 10 (9/1 in old money). The bets yielded a level profit of +£292.03, which corresponds to an impressive return on investment of 28.46%. Applying our Bayesian treatment to these bets yields the following posterior distribution for Hugh Taylor's γ-parameter shown in Figure 3.

Figure 3

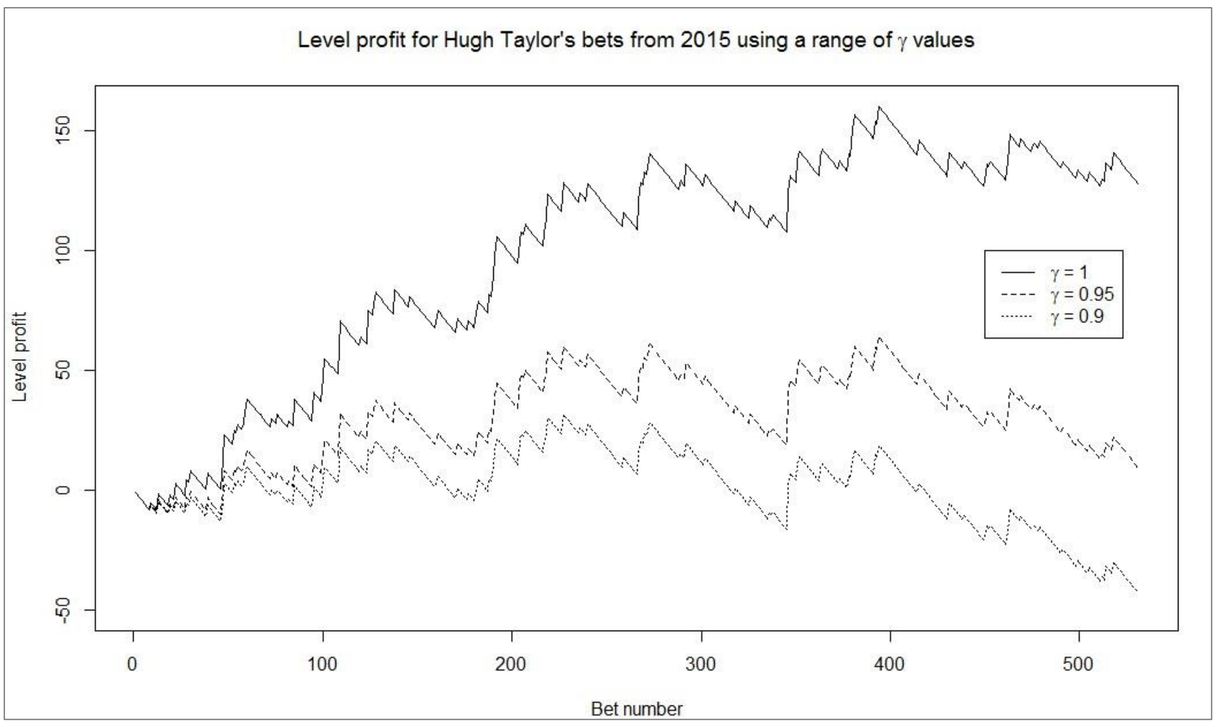


The evidential weight of the data has moved the peak of the distribution to $\gamma = 0.95$, well into the 'profitable' range. The posterior distribution is also noticeably narrower than the prior, reflecting the improved precision in our new estimate of γ – the 95% confidence interval for γ is approximately (0.9, 1), and only 3.8% of the probability mass is in the range $\gamma > 1$ (corresponding to the probability that the results so far have been a fluke).

Given a new tip at price D, the full Bayesian approach is to substitute the above posterior distribution into equation (1) (for p in terms of D and γ) and then integrate over all γ to obtain a prediction of the true probability of the bet winning. In practice, simply substituting the *max aposteriori* estimate – namely, the mode of the posterior distribution (in this case $\gamma_{mode} = 0.95$) – into equation (1) is typically good enough. And so, given a tip at price D, then any price bigger than $D^{\gamma_{mode}}$ should represent value.

We can validate these findings by applying them to some out-of-sample data – in the form of Hugh Taylor's tips for the year 2015 (remember, our fitting dataset comprised 2013 and 2014). The results are shown in Figure 4. This shows the cumulative level profit from various implementations of Hugh Taylor's betting tips for the year 2015. The solid line shows the case where we bet without fail at the recommended price, leading to a healthy level return of +127.55 units from level staking 531 bets or a 24% Return on Investment (RoI). The dashed line shows what happens if we bet at the recommended price raised to the power 0.95. As we've shown above, this price should closely approximate the inverse of the true probability, leading to a bet with zero value. So we expect to break even in this case – which we more or less do, recording a profit of +9.25 units (or just +1.7% RoI). Finally, the dotted line is a salutary tale showing what happens if you push your luck too far. Here, betting the prices to the power 0.9 leads to a clear loss of 43 units (-8% RoI). The message to take away from this graph is that, during 2015 at least, getting on at any price bigger than $D^{0.95}$ would have made you money – exactly as shown in the previous analysis.

Figure 4



One topic for future investigation might be staking plans. If level staking isn't your preferred method, then the calculations outlined here will allow you to employ Kelly staking – as well as any other staking plan that requires knowledge of each bet's true probability. Another topic for research might be to investigate the effect of a more generous prior distribution. The prior we've used makes quite pessimistic initial assumptions about the quality of the tipster under study – which is why it takes a full two years' worth of evidence to the contrary to convincingly turn these initial beliefs around. But when dealing with a tipster who we already suspect to be profitable, a prior distribution with a peak close to $\gamma = 1$ may be more appropriate.